



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
"ΔΙΑΧΕΙΡΙΣΗ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΔΕΔΟΜΕΝΩΝ"**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Περιβάλλον Επιμέλειας των αποτελεσμάτων της Αυτόματης  
Θεματικής Κατηγοριοποίησης Κειμένων**

**Αντώνης Χ. Κούλαλης**

**Επιβλέποντες: Ιωάννης Ιωαννίδης, Καθηγητής ΕΚΠΑ  
Δρ. Χάρης Παπαγεωργίου, Δρ. Ιωάννα Γρυπάρη**

**ΑΘΗΝΑ**

**ΟΚΤΩΒΡΙΟΣ 2019**

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Περιβάλλον Επιμέλειας των αποτελεσμάτων της Αυτόματης Θεματικής  
Κατηγοριοποίησης Κειμένων

**Αντώνης Χ. Κούλαλης**

**A.M.: M1405**

**ΕΠΙΒΛΕΠΟΝΤΕΣ:** **Ιωάννης Ιωαννίδης, Καθηγητής ΕΚΠΑ**  
**Δρ. Χάρης Παπαγεωργίου, Δρ. Ιωάννα Γρυπάρη**

Οκτώβριος 2019

## ΠΕΡΙΛΗΨΗ

Το Topic Modeling είναι ένας αλγόριθμος μηχανικής εκμάθησης που χρησιμοποιείται για την ταξινόμηση μιας μεγάλης βάσης δεδομένων με έγγραφα, σε κατηγορίες που χαρακτηρίζονται από ένα σύνολο λέξεων. Για την επιμέλεια αυτών των συνόλων, που ονομάζονται θέματα, χρειάζεται η ανθρώπινη παρέμβαση προκειμένου να γίνουν αναγνώσιμα από το σύνολο των αναλυτών. Απαραίτητη είναι λοιπόν η συμβολή των επιμελητών, οι οποίοι αναλαμβάνουν να δώσουν ευανάγνωστη μορφή στην έξοδο των διαφόρων μεθόδων του Topic Modeling. Η διαδικασία αυτή ονομάζεται επιμέλεια δεδομένων και είναι μια αρκετά επίπονη και χρονοβόρα διαδικασία. Ένας επιμελητής καλείται να μελετήσει τεράστιους όγκους από δεδομένα, συνήθως γραμμένα μέσα σε πίνακες λογιστικών φύλλων, προτού προχωρήσει στην αξιολόγηση, τη διόρθωση και τη γενικότερη επιμέλειά τους. Ως εκ τούτου η διαδικασία αυτή μπορεί να διαρκέσει αρκετά μεγάλο χρονικό διάστημα για έναν επιμελητή. Επιπλέον η έλλειψη αρκετών επιμελητών, αυξάνει τον όγκο εργασίας που καλείται καθένας από αυτούς να αναλάβει. Η απουσία ενός εργαλείου που θα μειώνει τον χρόνο εργασίας τους και θα τους παρέχει τη δυνατότητα πρόσβασης σε άλλες επιμέλειες δεδομένων που θα τους ήταν χρήσιμες, επιβαρύνει ακόμα περισσότερο το έργο τους. Η πολύωρη ενασχόληση προκαλεί την κούραση του επιμελητή, αυξάνει την πιθανότητα λαθών, μειώνοντας έτσι την ποιότητα των αποτελεσμάτων του. Στην παρούσα διπλωματική εργασία, παρουσιάζουμε μια φιλική προς τον χρήστη δικτυακή εφαρμογή που θα μπορέσει να αντιμετωπίσει τα παραπάνω προβλήματα και να διευκολύνει το έργο των επιμελητών. Ο χρήστης πέρα από την εξερεύνηση των δεδομένων, έχει τη δυνατότητα να τα αξιολογήσει, να τα διορθώσει και γενικότερα να τα επιμεληθεί. Ενέργειες όπως κατηγοριοποίηση, συγχώνευση, διαχωρισμό, ονοματοδοσία δεδομένων και πολλές άλλες θα μπορεί πλέον να τις πραγματοποιήσει μέσω της εφαρμογής μας εύκολα και γρήγορα. Πλούσια γραφήματα έρχονται να συμπληρώσουν τις παραπάνω λειτουργικότητες και όλα μαζί να συνθέσουν ένα εργαλείο απαραίτητο για τους επιμελητές δεδομένων.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Επιμέλεια Δεδομένων

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** αξιολόγηση δεδομένων, εξερεύνηση δεδομένων, αξιολόγηση δεδομένων στατιστικών μοντέλων, επιστημονικές δικτυακές εφαρμογές, διαδραστικές οπτικοποιήσεις

## **ABSTRACT**

Topic Modelling is a machine learning algorithm used in order to classify a large database of documents into categories each characterised by a set of words. These, so-called topics need to be manually curated in order to be more easily understandable. The contribution of curators, who undertake to make readable the output of the various Topic Modelling methods, are therefore required. This process is called data curation and is quite a laborious and time consuming process. Curator is required to study huge volumes of data, usually written in spreadsheets, before proceeding with their evaluation, correction and general curation. Therefore this process can take quite a long time for a curator. In addition, the lack of several curators increases the amount of work, that each of them is required to undertake. The absence of a tool that will reduce their working time and give them access to other data curations that would be useful to them, makes their work even more difficult. Long time work causes curator's fatigue, increases the likelihood of errors, thereby reducing the quality of his results. In this thesis, we present a user-friendly web application that can address the above problems and facilitate the work of curators. In addition to exploring the data, the user is able to evaluate, edit and generally curate them. Actions such as categorization, merging, splitting, data labeling and many more will now be able to execute them easily and quickly through our application. Rich visualizations come to complement the above functionalities and together make up a tool essential for data curators.

**SUBJECT AREA:** Data Curation

**KEYWORDS:** data evaluation, data exploration, evaluation of statistical model data, scientific web applications, interactive visualizations, topic modeling curation

*Στην οικογένειά μου.*

## ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω κατά αρχήν τον Καθ. Ιωάννη Ιωαννίδη, που με εμπιστεύτηκε δίνοντάς μου τη συγκεκριμένη Διπλωματική Εργασία και επιτρέποντάς μου να υλοποιήσω την ιδέα μου.

Τους επιβλέποντες Δρ. Ιωάννα Γρυπάρη και τον Δρ. Χάρη Παπαγεωργίου που μου προσέφεραν την καλύτερη υποστήριξη που θα μπορούσα να έχω για την διεκπεραίωση της παρούσας Διπλωματικής Εργασίας.

Τους φίλους μου που πάντα με ενθαρρύνουν και μου συμπαραστέκονται σε ότι κάνω.

Τέλος, φυσικά την οικογένειά μου για την κατανόηση και την υποστήριξη που μου προσέφεραν σε όλη τη διάρκεια των σπουδών μου και γιατί είναι πάντα δίπλα μου και με στηρίζουν να πραγματοποιήσω όσα ονειρεύομαι.

# ΠΕΡΙΕΧΟΜΕΝΑ

|   |           |
|---|-----------|
| <b>ΠΡΟΛΟΓΟΣ</b> .....                               | <b>11</b> |
| <b>1. ΕΙΣΑΓΩΓΗ</b> .....                            | <b>12</b> |
| 1.1 Αντικείμενο Εργασίας .....                      | 12        |
| 1.2 Στόχος Εργασίας .....                           | 13        |
| 1.3 Κίνητρα .....                                   | 13        |
| 1.4 Διάρθρωση της εργασίας.....                     | 14        |
| <b>2. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ</b> .....            | <b>15</b> |
| 2.1 Εισαγωγή.....                                   | 15        |
| 2.2 Υλοποίηση Εξυπηρετητή .....                     | 15        |
| 2.2.1 Περιβάλλον Ανάπτυξης .....                    | 15        |
| 2.2.2 Γλώσσα Προγραμματισμού και πρόσθετα .....     | 17        |
| 2.3 Υλοποίηση Πελάτη .....                          | 19        |
| 2.3.1 Περιβάλλον Ανάπτυξης .....                    | 19        |
| 2.3.2 Γλώσσες Προγραμματισμού και βιβλιοθήκες ..... | 20        |
| 2.4 Συμπεράσματα .....                              | 24        |
| <b>3. ΠΑΡΟΥΣΙΑΣΗ ΕΡΓΑΛΕΙΟΥ</b> .....                | <b>25</b> |
| 3.1 Εισαγωγή.....                                   | 25        |
| 3.2 Παρουσίαση πλατφόρμας .....                     | 25        |
| 3.2.1 Οθόνη των Experiments.....                    | 25        |
| 3.2.2 Οθόνη των Topics .....                        | 27        |
| 3.2.3 Οθόνη των Documents .....                     | 38        |
| 3.3 Συμπεράσματα .....                              | 42        |
| <b>4. ΣΥΓΚΡΙΣΗ ΜΕ ΑΛΛΑ ΕΡΓΑΛΕΙΑ</b> .....           | <b>43</b> |
| 4.1 Παρουσίαση άλλων εργαλείων .....                | 43        |
| 4.1.1 Interactive Topic Modeling (ITM) .....        | 43        |
| 4.1.2 UTOPIAN .....                                 | 44        |

|       |  |           |
|-------|--|-----------|
| 4.1.3 | iVisClustering .....   | 45        |
| 4.1.4 | Interactive Visualization for Topic Model Curation .....                             | 46        |
| 4.2   | Σύνοψη σύγκρισης με τη δική μας πλατφόρμα.....                                       | 47        |
| 4.3   | Συμπεράσματα .....   | 48        |
| 5.    | <b>ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ .....</b>                                 | <b>50</b> |
| 5.1   | Ανασκόπηση.....  | 50        |
| 5.2   | Συμπεράσματα .....   | 50        |
| 5.3   | Μελλοντικές επεκτάσεις.....  | 51        |
| 5.4   | Επίλογος.....  | 53        |
|       | <b>ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ .....</b>                                    | <b>56</b> |
|       | <b>ΠΑΡΑΡΤΗΜΑ Ι: ΚΩΔΙΚΑΣ .....</b>  | <b>57</b> |
|       | <b>ΠΑΡΑΡΤΗΜΑ ΙΙ: ΠΛΗΡΟΦΟΡΙΕΣ ΕΚΤΕΛΕΣΗΣ ΚΑΙ ΛΗΨΗΣ ΤΟΥ ΚΩΔΙΚΑ ΤΗΣ ΠΛΑΤΦΟΡΜΑΣ .....</b> | <b>60</b> |
|       | <b>ΑΝΑΦΟΡΕΣ .....</b>  | <b>61</b> |



## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

|  |    |
|--|----|
| Εικόνα 1: IntelliJ IDEA .....  | 16 |
| Εικόνα 2: WebStorm.....  | 20 |
| Εικόνα 3: Bootstrap Grid System για responsiveness .....                         | 22 |
| Εικόνα 4: Αρχική οθόνη εργαλείου .....   | 25 |
| Εικόνα 5: Φιλτράρισμα των experiments εντός συγκεκριμένης χρονικής περιόδου..... | 26 |
| Εικόνα 6: Φιλτράρισμα των experiments βάσει κάποιου keyword .....                | 27 |
| Εικόνα 7: Οθόνη των topics (Κύρια οθόνη του εργαλείου).....                      | 28 |
| Εικόνα 8: Χαρακτηρισμός topic ως “good” .....                                    | 29 |
| Εικόνα 9: Αναδυόμενο παράθυρο καταχώρησης ετικέτας στο topic .....               | 29 |
| Εικόνα 10: Αναδυόμενο παράθυρο διαχείρισης κατηγοριών του topic .....            | 30 |
| Εικόνα 11: Φιλτράρισμα των topics βάσει κατηγορίας.....                          | 30 |
| Εικόνα 12: Φιλτράρισμα των topics βάσει κατηγορίας - αποτελέσματα .....          | 31 |
| Εικόνα 13: Επεξεργασία κατηγοριών experiment .....                               | 31 |
| Εικόνα 14: Μετονομασία κατηγορίας .....  | 32 |
| Εικόνα 15: Μαζική κατηγοριοποίηση των topics .....                               | 32 |
| Εικόνα 16: Διάγραμμα εξέλιξης στον χρόνο των topic .....                         | 33 |
| Εικόνα 17: Αναδυόμενο παράθυρο Commit .....                                      | 34 |
| Εικόνα 18: Παράδειγμα commit τις τρέχουσες αλλαγές .....                         | 34 |
| Εικόνα 19: Σελιδοποίηση πίνακα δεδομένων .....                                   | 35 |
| Εικόνα 20: Γράφημα κατανομής των topic στις κατηγορίες του experiment.....       | 35 |
| Εικόνα 21: Επιλογή μιας κατηγορίας από το γράφημα .....                          | 36 |
| Εικόνα 22: Επιλογή περισσότερων κατηγοριών από το γράφημα .....                  | 36 |
| Εικόνα 23: Οδηγίες για τη χρήση του εργαλείου .....                              | 37 |
| Εικόνα 24: Οθόνη των documents.....  | 38 |
| Εικόνα 25: Αναδυόμενο παράθυρο για curation του topic .....                      | 39 |
| Εικόνα 26: Αναδυόμενο παράθυρο για curation του topic - αλλαγές .....            | 39 |

|   |    |
|---|----|
| Εικόνα 27: Πίνακας των documents .....  | 40 |
| Εικόνα 28: Παράθυρο με εκτενείς πληροφορίες για το document.....                      | 40 |
| Εικόνα 29: Ιστόγραμμα journal-πλήθος documents.....                                   | 41 |
| Εικόνα 30: Φιλτράρισμα των documents συγκεκριμένου journal .....                      | 41 |
| Εικόνα 31: Επιφάνεια διεπαφής ITM .....   | 44 |
| Εικόνα 32: Επιφάνεια διεπαφής UTOPIAN.....  | 45 |
| Εικόνα 33: Επιφάνεια διεπαφής iVisClustering.....                                     | 46 |
| Εικόνα 34: Επιφάνεια διεπαφής Interactive Visualization for Topic Model Curation..... | 47 |

## **ΠΡΟΛΟΓΟΣ**

Η παρούσα εργασία εκπονήθηκε στο τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών ως διπλωματική εργασία του Μεταπτυχιακού Προγράμματος “Διαχείριση Πληροφορίας και Δεδομένων” του Τμήματος. Η διάρκεια της διεξαγωγής της ήταν ένας χρόνος. Υπεύθυνος Καθηγητής ήταν ο κ. Ιωάννης Ιωαννίδης και επιβλέποντες η Δρ. Ιωάννα Γρυπάρη και ο Δρ. Χάρης Παπαγεωργίου.

## 1. ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο γίνεται μια εισαγωγή στο αντικείμενο της εργασίας. Αρχικά παρουσιάζεται το θέμα της εργασίας, δηλαδή ποιο θα είναι το περιεχόμενο της. Στη συνέχεια αναφέρεται ο στόχος και τα κίνητρα για τη συγκεκριμένη διπλωματική, με άλλα λόγια τι ήταν αυτό που με ώθησε στην ενασχόλησή μου με το συγκεκριμένο αντικείμενο και τέλος επισημαίνεται ο τρόπος με τον οποίο διαρθρώνεται η εργασία.

### 1.1 Αντικείμενο Εργασίας

Το αντικείμενο της παρούσας εργασίας είναι η διαδικασία του Topic Modeling curation και η δημιουργία μιας διαδραστικής πλατφόρμας που θα δίνει την δυνατότητα στον curator να υλοποιήσει τη διαδικασία αυτή γρηγορότερα και ποιοτικότερα. Γιατί όμως να χρειάζεται ειδική πλατφόρμα για αυτήν την διαδικασία;

Από την στιγμή που ανακαλύφθηκαν οι υπολογιστές, υπήρχε η αντίληψη ότι μπορούμε να τους χρησιμοποιήσουμε για περίπλοκους υπολογισμούς αλλά και για την διαχείριση πολλών δεδομένων τα οποία είναι δύσκολο να τα συγκρατήσει και να τα διαχειριστεί ο ανθρώπινος νους. Με την πάροδο των χρόνων λοιπόν, δημιουργήθηκαν διάφορα μοντέλα και αλγόριθμοι μηχανικής μάθησης, οι οποίοι μπορούν να εκτελεστούν πάνω σε υπολογιστές και να τους εκπαιδεύσουν με σκοπό να εξάγουν σχετικά γρήγορα και με μεγάλη ακρίβεια δεδομένα που για να τα παράξει ο άνθρωπος μόνος του χωρίς τη βοήθεια των μηχανών πιθανόν να χρειαζόταν και μέρες. Ένα τέτοιο μοντέλο είναι το Topic Modeling. Το μοντέλο αυτό ανακαλύφθηκε για να διευκολύνει την ποιοτική και ποσοτική ανάλυση μεγάλων συλλογών κειμένων, με την ταξινόμησή τους σε κατηγορίες, που ονομάζονται topics και εκπροσωπούνται από σύνολα λέξεων. Αυτή είναι και η έξοδος του Topic Modeling, σύνολα λέξεων που καθένα από αυτά χαρακτηρίζει ένα θέμα. Με την ανάλυση λοιπόν κειμένων προερχόμενων από το διαδίκτυο, τα social media και τα blogs είναι δυνατόν να βγάλεις συμπεράσματα για το ποια θέματα είναι στην επικαιρότητα κάθε χρονική περίοδο, αλλά και να δεις την εξέλιξή τους με την πάροδο των χρόνων. Επίσης μπορείς να βρεις ποια κείμενα είναι πιο αντιπροσωπευτικά για ένα θέμα, ποιες λέξεις εμφανίζονται με μεγαλύτερη συχνότητα σε αυτά τα κείμενα και πολλά άλλα. Προκειμένου όμως ο αναλυτής να μπορέσει να διαβάσει όλη την πληροφορία που εξάγεται από το Topic Modeling και να διαμορφώσει τα συμπεράσματά του, θα πρέπει να μπορεί να κατανοήσει την παραγόμενη αυτή πληροφορία. Τη διαδικασία αυτή, πιο συγκεκριμένα τη μετατροπή των διαφόρων συνόλων από λέξεις σε μορφή αναγνώσιμη και κατανοητή από τους αναλυτές, την αναλαμβάνουν οι επιμελητές. Η διαδικασία αυτή λέγεται επιμέλεια δεδομένων (data curation).

Για να μπορέσει ο curator να επιμεληθεί τα δεδομένα που εξάγονται από το Topic Modeling, να τους προσθέσει ετικέτες και να τους δώσει μια μορφή κατανοητή στο σύνολο των αναλυτών που δεν είναι ειδικοί του Topic Modeling και της μηχανικής μάθησης, είναι αναγκασμένος μέχρι και σήμερα να διαβάζει για πολλές ώρες άπειρα δεδομένα, συνήθως γραμμένα σε ατελείωτους πίνακες excel. Και αυτή η διαδικασία πρέπει να επαναληφθεί κάθε φορά που η εκτέλεση του Topic Modeling επιστρέφει νέα δεδομένα. Όπως εύκολα μπορεί να γίνει κατανοητό, κάτι τέτοιο είναι επίπονο για έναν curator και πολλές φορές η κούραση μπορεί να τον οδηγήσει σε λάθη κατά την αξιολόγηση. Επίσης, δεδομένου ότι οι curators είναι πολύ λίγοι στο σύνολό τους, αν το curation των εξαγόμενων δεδομένων από μια μόνο εκτέλεση Topic Modeling διαρκεί τόσο πολύ ώρα, θα είναι πολύ δύσκολο να βρούμε ελεύθερο curator να δεσμεύσουμε να επιμεληθεί τα εξαγόμενα δεδομένα από μια άλλη εκτέλεση Topic Modeling. Συνεπώς δεν είναι και σίγουρο ότι ο ίδιος curator θα αναλάβει την επιμέλεια των δεδομένων όλων

των διαδοχικών εκτελέσεων του ίδιου Topic Modeling, αφού μπορεί να είναι απασχολημένος με το curation των δεδομένων κάποιας άλλης εκτέλεσης. Στη συγκεκριμένη περίπτωση, με τα μέχρι τώρα δεδομένα, δεν υπάρχει η δυνατότητα ένας επιμελητής να έχει πρόσβαση σε παλαιότερα curation τα οποία θα του ήταν χρήσιμα κατά την επιμέλεια του τρέχοντος run του Topic Modeling. Όλα τα παραπάνω λοιπόν οδηγούν στην ιδέα για μια διαδραστική πλατφόρμα που θα μειώσει τον χρόνο εργασίας του curator, θα αυξήσει την ποιότητα του curation και θα κάνει εφικτή την πρόσβαση και την επιπλέον επεξεργασία όλων των υπάρχοντων curations από όλους τους curators.

Η συγκεκριμένη εργασία λοιπόν επικεντρώνεται στη δημιουργία αυτής της πλατφόρμας.

## 1.2 Στόχος Εργασίας

Μέσω του Topic Modeling, όπως αναφέραμε και προηγουμένως, μπορούμε να μετατρέψουμε ένα σύνολο από κείμενα βάσει του περιεχομένου τους σε clusters από λέξεις όπου κάθε ένα από αυτά εκφράζει ένα συγκεκριμένο θέμα. Στόχος της εργασίας αυτής είναι η δημιουργία μιας πλατφόρμας μέσω της οποίας ο curator θα μπορέσει να διαχειριστεί ευκολότερα, γρηγορότερα και αποτελεσματικότερα τα σύνολα αυτά με το μεγάλο πλήθος λέξεων που του παρέχει το Topic Modeling. Είναι γεγονός ότι όταν ένας curator έχει να αναλύσει τέτοια μεγάλα σύνολα δεδομένων, συνήθως γραμμένα σε πίνακες λογιστικών φύλλων με χιλιάδες γραμμές χρειάζεται ένα εργαλείο που θα τον διευκολύνει σε αυτό το έργο και θα του μειώσει τον χρόνο εργασίας. Ένα τέτοιο εργαλείο θα πρέπει να παρουσιάζει συνοπτικά και δομημένα τις πληροφορίες που εξάγει το Topic Modeling, και παράλληλα θα του δίνει τη δυνατότητα να εκτελεί τη διαδικασία του curation με τον απλούστερο δυνατό τρόπο.

Έτσι σκοπός και ιδέα μας είναι να δημιουργηθεί μια δικτυακή εφαρμογή που θα δίνει στο χρήστη (curator) τη δυνατότητα να επεξεργαστεί τα παραγόμενα από το Topic Modeling δεδομένα. Μεταξύ άλλων θα μπορεί να τους προσθέσει ετικέτες, να τα κατηγοριοποιήσει, να τα αξιολογήσει και γενικά θα μπορεί να εκτελέσει όλες τις απαραίτητες ενέργειες για να μετατρέψει τα δεδομένα αυτά σε μορφή κατανοητή προς το σύνολο των αναλυτών που δεν είναι ειδικοί να μπορούν να τα διαβάσουν στη μορφή των clusters από λέξεις που παράγονται από το Topic Modeling.

Επομένως, στο τέλος αυτής της εργασίας θα υπάρχει μια πλατφόρμα που θα απευθύνεται κυρίως σε curators και θα δίνει λύση στα τρία προβλήματα που αναφέραμε και στην προηγούμενη ενότητα. Θα τους επιτρέψει λοιπόν να υλοποιήσουν τη διαδικασία του curation σε μικρότερο χρονικό διάστημα και με λιγότερες πιθανότητες λάθους από πριν, ενώ ταυτόχρονα θα μπορούν να έχουν όλοι πρόσβαση στα διαθέσιμα run του Topic Modeling και στο curation τους με αποτέλεσμα να είναι πιο εύκολη η επεξεργασία υπάρχοντων curation αλλά και η ολοκλήρωση νέων από όλους.

## 1.3 Κίνητρα

Το κίνητρό μας για να ασχοληθούμε με το συγκεκριμένο θέμα ήταν η επιθυμία μας να βοηθήσουμε τους curators να υλοποιήσουν ευκολότερα το έργο τους, ελαχιστοποιώντας το χρόνο εργασίας τους και αυξάνοντας την ποιότητα του curation.

Οι επιμελητές δεδομένων είναι περιορισμένοι στο σύνολό τους, συνεπώς όταν υποχρεώνονται να αφιερώσουν αρκετό χρόνο για ένα curation είναι δύσκολο να βρούμε ελεύθερο επιμελητή να δεσμεύσουμε για το curation των δεδομένων που παρήχθησαν από ένα run του Topic Modeling. Αν υπήρχε όμως μια εφαρμογή που θα μείωνε τον χρόνο υλοποίησης του curation, οι επιμελητές θα μπορούσαν να αναλάβουν και να υλοποιήσουν γρηγορότερα την επιμέλεια περισσότερων συλλογών δεδομένων.

Επιπροσθέτως, δεν μπορεί να μας εγγυηθεί κανείς ότι θα είναι ο ίδιος επιμελητής που θα αναλάβει το curation ενός run του Topic Modeling και των διαδοχικών του. Μπορεί ο συγκεκριμένος curator να είναι δεσμευμένος με το curation κάποιου άλλου run και να μην μπορεί να αναλάβει το curation του νέου run του Topic Modeling που αυτός είχε επιμεληθεί προηγουμένως, με αποτέλεσμα να πρέπει να το αναλάβει κάποιος άλλος. Με τα μέχρι τώρα δεδομένα αυτό είναι δύσκολο να γίνει, διότι δεν έχουν πρόσβαση οι curators να εξερευνήσουν και να επεξεργαστούν όλα τα υπάρχοντα curations. Συνεπώς δεν υπάρχει η δυνατότητα, ένας επιμελητής να αντλήσει πληροφορία από ένα παλαιότερο curation ώστε να υλοποιήσει ένα νέο, αλλά και να μπορέσει να επεξεργαστεί περαιτέρω το παλιό αν δεν είναι ο ίδιος που το είχε αναλάβει στο παρελθόν.

Όλα τα παραπάνω μας οδήγησαν να σκεφτούμε ότι αν υπήρχε μια εφαρμογή που με έναν εύκολο χειρισμό θα επέτρεπε στους επιμελητές να υλοποιήσουν το curation μεγάλων συνόλων από λέξεις, παραγόμενων από την εκτέλεση του Topic Modeling σε μεγάλες βάσεις δεδομένων με έγγραφα, και η διαδικασία αυτή να γίνεται εύκολα, γρήγορα και τα αποτελέσματά της να είναι προσβάσιμα και επεξεργάσιμα από όλους τους επιμελητές, θα αποτελούσε ένα εργαλείο απαραίτητο για όλους τους curators δεδομένου ότι αντίστοιχη εφαρμογή που να επιτυγχάνει όλα τα παραπάνω με τον πιο ολοκληρωμένο τρόπο δεν υπάρχει υλοποιημένη στην αγορά. Αυτός λοιπόν ήταν ένας καλός λόγος για να ασχοληθούμε με κάτι τόσο ενδιαφέρον και καινοτόμο.

#### 1.4 Διάρθρωση της εργασίας

Ολοκληρώνοντας το **κεφάλαιο 1** αναφέρουμε περιληπτικά τι περιέχουν τα επόμενα κεφάλαια:

Στο **κεφάλαιο 2** θα αναλύσουμε τον κώδικα που αναπτύχθηκε για την υλοποίηση της εφαρμογής και τα εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη του συγκεκριμένου κώδικα.

Στο **κεφάλαιο 3** παρουσιάζεται η πλατφόρμα που αναπτύξαμε και δίνεται και ένα παράδειγμα από τις διάφορες δραστηριότητες που μπορεί να εκτελέσει ο χρήστης εντός της πλατφόρμας, συνοδευόμενο από μια καλή επεξήγηση.

Στο **κεφάλαιο 4** θα παρουσιάσουμε εργαλεία συναφή με το δικό μας, τα οποία ήδη κυκλοφορούν στην αγορά ή απλά έχουν προταθεί αλλά δεν έχουν υλοποιηθεί. Θα επισημάνουμε τις διαφορές τους και θα καταλήξουμε στο που υπερτερεί η δικιά μας εφαρμογή έναντι αυτών.

Στο τελευταίο **κεφάλαιο 5** κάνουμε μια τελική ανασκόπηση και αναφέρουμε κάποιες πιθανές μελλοντικές επεκτάσεις.

## 2. ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΣΥΣΤΗΜΑΤΟΣ

### 2.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με τον κώδικα της πλατφόρμας. Θα αναλύσουμε τις τεχνικές που χρησιμοποιήθηκαν για τη συγγραφή του. Θα αναφερθούμε στις γλώσσες προγραμματισμού που χρησιμοποιήσαμε και το περιβάλλον στο οποίο εργαστήκαμε τόσο στην πλευρά του εξυπηρετητή όσο και στην πλευρά του πελάτη. Στο τέλος θα συνοψίσουμε όλα όσα είδαμε στο παρόν κεφάλαιο.

Η ανάπτυξη της πλατφόρμας βασίζεται σε δύο κομμάτια, τον εξυπηρετητή και τον πελάτη ακολουθώντας την αρχιτεκτονική του μοντέλου πελάτης-εξυπηρετητής [22]. Ο εξυπηρετητής είναι ένα πρόγραμμα, το οποίο προσφέρει λειτουργικότητα στον πελάτη. Τη λειτουργικότητα αυτή συνηθίζουμε να την ονομάζουμε “υπηρεσίες” του εξυπηρετητή και συνήθως περιλαμβάνει την αλληλεπίδραση του με την βάση δεδομένων και την επιστροφή συγκεκριμένης πληροφορίας στον πελάτη, η οποία προκύπτει είτε κατόπιν επεξεργασίας στα δεδομένα της βάσης δεδομένων είτε απλής ανάκτησής τους από αυτήν. Όταν κληθήκαμε λοιπόν να αναπτύξουμε την πλατφόρμα μας, έπρεπε να αποφασίσουμε σε τι τεχνολογίες θα το κάνουμε. Βασιζόμενοι στην παραπάνω αρχιτεκτονική αποφασίσαμε να είναι μια Web εφαρμογή και κατά συνέπεια επιλέξαμε να χρησιμοποιήσουμε την αντίστοιχη τεχνολογία. Τα εργαλεία και οι γλώσσες προγραμματισμού που χρησιμοποιήσαμε, είναι τα πιο σύγχρονα και αυτά που χρησιμοποιούνται ευρέως πλέον στην υλοποίηση web εφαρμογών.

Αυτά τα εργαλεία αλλά και το περιβάλλον στο οποίο εργαστήκαμε θα σας παρουσιάσουμε και θα αναλύσουμε στο υπόλοιπο του κεφαλαίου.

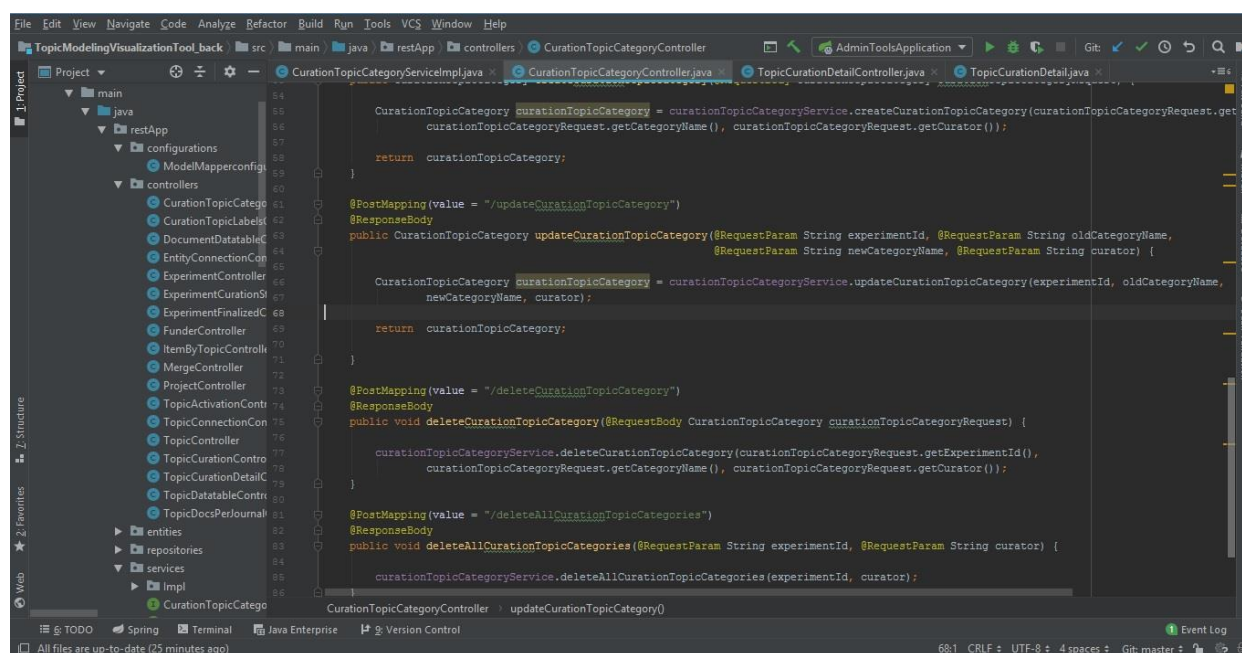
### 2.2 Υλοποίηση Εξυπηρετητή

Σε αυτό το υποκεφάλαιο θα περιγράψουμε αναλυτικά το περιβάλλον, τις γλώσσες προγραμματισμού και οτιδήποτε χρειάστηκε για να υλοποιήσουμε το πρόγραμμα του εξυπηρετητή.

#### 2.2.1 Περιβάλλον Ανάπτυξης

Στην συγκεκριμένη ενότητα θα αναφερθούμε στο σύστημα και τα εργαλεία που χρησιμοποιήσαμε για να γράψουμε τον κώδικα. Το σύστημα λοιπόν στο οποίο κατά βάση αναπτύχθηκε η εφαρμογή είναι ένας υπολογιστής HP Pavilion 15 – e005sv με λειτουργικό σύστημα Windows 8.1, 8GB φυσικής μνήμης και σκληρό δίσκο 500GB HDD. Ωστόσο έχει ελεγχθεί στο σύνολό του αλλά και αναπτυχθεί ως προς ένα κομμάτι του σε υπολογιστή Dell Latitude E5550 με λειτουργικό σύστημα Windows 10, 8GB φυσικής μνήμης και σκληρό δίσκο 500GB SSD.

Για τη συγγραφή του προγράμματος εξυπηρετητή χρησιμοποιήσαμε το **IntelliJ IDEA**.



Εικόνα 1: IntelliJ IDEA

Πρόκειται για ένα ολοκληρωμένο περιβάλλον ανάπτυξης προγραμμάτων κυρίως Java, αλλά και άλλων γλωσσών προγραμματισμού, το οποίο έχει αναπτυχθεί από την JetBrains. Έχει δύο εκδόσεις, μία ανοιχτού λογισμικού (IntelliJ IDEA Community) και μία εμπορική (IntelliJ IDEA Ultimate). Ο εξυπηρετητής της πλατφόρμας μας αναπτύχθηκε στην Ultimate έκδοση, που υποστηρίζει ανάπτυξη διαδικτυακών εφαρμογών. Επιλέχθηκε το συγκεκριμένο περιβάλλον ανάπτυξης διότι μας διευκόλυνε πολύ στην συγγραφή και εκτέλεση της εφαρμογής μας. Όπως αναφέρθηκε και προηγουμένως, χρησιμοποιείται κυρίως για την ανάπτυξη προγραμμάτων Java και είναι πολύ βολικό γιατί προσφέρει πολλές δυνατότητες, plug-in<sup>1</sup> και frameworks<sup>2</sup> που συμβάλουν στην εύκολη και γρήγορη ανάπτυξη κώδικα για την υλοποίηση ενός εξυπηρετητή. Από αυτές τις δυνατότητες θα αναφερθούμε μόνο σε εκείνες που μας φάνηκαν χρήσιμες στην εργασία μας γιατί πραγματικά είναι πάρα πολλές να τις αναφέρουμε όλες. Πρώτη και βασικότερη είναι ότι υποστηρίζει το Spring Boot το οποίο είναι μια συλλογή από framework που χρησιμοποιείται για την ανάπτυξη αυτόνομων πλατφόρμων. Μια δεύτερη πολύ σημαντική δυνατότητα είναι ότι υποστηρίζει το Apache Maven. Πρόκειται για ένα build automation<sup>3</sup> εργαλείο, το οποίο διευκολύνει και αυτοματοποιεί την ανάπτυξη, τον έλεγχο και την εκτέλεση προγραμμάτων γραμμένων σε Java. Θα αναφερθούμε παρακάτω με περισσότερη λεπτομέρεια στο Maven αλλά και στα frameworks, γλώσσες προγραμματισμού και βιβλιοθήκες που χρησιμοποιήσαμε για την συγγραφή του κώδικα.

<sup>1</sup> Είναι ένα σύστημα συστατικών κάποιου λογισμικού που προσθέτει ιδιαίτερες δυνατότητες σε ένα μεγαλύτερο λογισμικό.

<sup>2</sup> Είναι λογισμικά τα οποία προσφέρουν στον χρήστη πληθώρα από λειτουργικότητες που τον διευκολύνουν στην συγγραφή, ανάπτυξη, εγκατάσταση και εκτέλεση εφαρμογών.

<sup>3</sup> Είναι η διαδικασία αυτοματοποίησης της δημιουργίας και ανάπτυξης λογισμικού και περιλαμβάνει την μεταγλώπτιση του κώδικα, τη συσκευασία του εκτελέσιμου κώδικα και την εκτέλεση αυτοματοποιημένων τεστ πάνω σε αυτό.



Μια επιπλέον δυνατότητα του IntelliJ που εκμεταλλευτήκαμε ήταν η έξυπνη και η αλυσιδωτή συμπλήρωση που μας προσφέρει. Μέσω αυτών των τεχνικών σου δίνει τη δυνατότητα να γράφεις πιο γρήγορα κώδικα εμφανίζοντας σου μια λίστα από πιθανούς συμβολισμούς (ονόματα συναρτήσεων, μεταβλητών κλπ) ενώ πληκτρολογείς. Οι προτεινόμενοι αυτοί συμβολισμοί βασίζονται στην ανάλυση της συχνότητας που τους χρησιμοποιείς σε ένα συγκεκριμένο κομμάτι του κώδικα σε συνδυασμό με το ποιους συμβολισμούς μπορείς να χρησιμοποιήσεις σε αυτό το κομμάτι βάση περιορισμών της γλώσσας προγραμματισμού. Διαθέτει δηλαδή ένα αρκετά έξυπνο σύστημα που μαντεύει τι πιθανόν να θες να γράψεις και σου δίνει και συμβουλές για αλλαγές στον κώδικα βάσει ασυμβατότητας τύπων ή μη υποστηριζόμενων από την έκδοση της γλώσσας προγραμματισμού που χρησιμοποιείς. Επίσης περιλαμβάνει ένα πολύ καλό σύστημα στατικού και δυναμικού ελέγχου του κώδικα που γράφεις, συμβάλλοντας έτσι στην αποφυγή λαθών. Επιπροσθέτως, διαθέτει έναν εργονομικό σχεδιασμό και σε συνδυασμό με τις συντομεύσεις που σε βοηθούν να εκτελείς διάφορες ενέργειες εντός της πλατφόρμας, η αναζήτηση αλλά και όλες οι υπόλοιπες δραστηριότητες γίνονται πολύ εύκολα και γρήγορα.

Ακόμη, διαθέτει πολλά παράθυρα για ανάπτυξη, μεταγλώττιση, έλεγχο, εκτέλεση του προγράμματος σου, όπως επίσης και παράθυρα για διαχείριση της βάσης δεδομένων και του εξυπηρετητή στον οποίο τρέχεις το πρόγραμμά σου.

Τέλος διαθέτει και ένα πολύ καλό σύστημα που σε βοηθάει να γράφεις καλύτερα και πιο δομημένα τον κώδικα σου είτε υποδεικνύοντας σου τα σημεία στα οποία επαναλαμβάνεις αυτούσιο κώδικα είτε και με διάφορα plug-in που σου δίνουν συμβουλές για την μορφοποίηση του κώδικα.

Αν ενδιαφέρεστε να μάθετε περισσότερα για το IntelliJ IDEA μπορείτε να επισκεφθείτε την επίσημη ιστοσελίδα του: <https://www.jetbrains.com/idea/>.

## 2.2.2 Γλώσσα Προγραμματισμού και πρόσθετα

Στην ενότητα αυτή θα αναφερθούμε στη γλώσσα προγραμματισμού και το βασικό framework που χρησιμοποιήσαμε για τη συγγραφή του εξυπηρετητή, όπως επίσης και στις βιβλιοθήκες που χρειαστήκαμε.

### 2.2.2.1 Java (Spring Framework & Spring Boot)

Η γλώσσα λοιπόν στην οποία δομήσαμε τον εξυπηρετητή της πλατφόρμας μας είναι η Java. Χρησιμοποιήσαμε την Java 8 και τη συνδυάσαμε με το Spring Boot framework. Πρόκειται για ένα ανοιχτού λογισμικού framework της Java το οποίο χρησιμοποιείται κυρίως για τη δημιουργία micro Services<sup>4</sup>. Δίνει τη δυνατότητα στους προγραμματιστές να αναπτύσσουν μια ανεξάρτητη και ποιοτική Spring<sup>5</sup> εφαρμογή η οποία μπορεί να εγκατασταθεί και να τρέξει σε οποιοδήποτε σύστημα χωρίς την χρήση επιπλέον εξωτερικών βιβλιοθηκών και κώδικα. Ο λόγος που επιλέξαμε το συγκεκριμένο

---

<sup>4</sup> Είναι μια αρχιτεκτονική η οποία επιτρέπει στους προγραμματιστές να αναπτύσσουν ανεξάρτητες υπηρεσίες (προγράμματα), που μπορούν να εγκατασταθούν και να τρέξουν με πολύ μικρή κατανάλωση πόρων σε οποιοδήποτε σύστημα.

<sup>5</sup> Το Spring framework είναι ένα λογισμικό ανοιχτού κώδικα που σκοπό έχει να διευκολύνει την ανάπτυξη J2EE λογισμικού σε μεγάλη έκταση και βασίζεται στη γλώσσα προγραμματισμού Java.

framework είναι ότι βοηθάει στην εύκολη ανάπτυξη τέτοιου τύπου προγραμμάτων, μειώνει τον χρόνο συγγραφής τους και αυξάνει την παραγωγικότητα του προγραμματιστή. Επίσης διαθέτει σύστημα αυτόματων ρυθμίσεων, και ως εκ τούτου δε χρειάζεται να ασχοληθεί ο ίδιος ο προγραμματιστής με την ρύθμιση των παραμέτρων του συστήματος. Τέλος διαθέτει ενσωματωμένο σύστημα, όπου μπορούμε να τρέχουμε τα προγράμματα που αναπτύσσουμε χωρίς να χρειάζεται η εγκατάσταση επιπλέον προγραμμάτων.

Αν επιθυμείτε να μάθετε περισσότερα πράγματα για το Spring Boot μπορείτε να επισκεφθείτε κάποιο tutorial όπως το ακόλουθο:

[https://www.tutorialspoint.com/spring\\_boot/spring\\_boot\\_introduction.htm](https://www.tutorialspoint.com/spring_boot/spring_boot_introduction.htm)

### 2.2.2.2 Maven

Το συγκεκριμένο εργαλείο το χρησιμοποιήσαμε σε συνδυασμό με το IntelliJ και το Spring Boot που αναφέραμε προηγουμένως. Το Maven είναι ένα build automation<sup>6</sup> εργαλείο, το οποίο διευκολύνει και αυτοματοποιεί την ανάπτυξη, τον έλεγχο και την εκτέλεση προγραμμάτων γραμμένων σε Java. Μέσω ενός αρχείου XML που ορίζετε μες τον κεντρικό φάκελο του προγράμματός μας περιγράφουμε τον κώδικα που θέλουμε να κάνουμε build, τις εξαρτήσεις από εξωτερικές βιβλιοθήκες Java, τις οποίες αναλαμβάνει να κατεβάσει και να εγκαταστήσει στο πρόγραμμά μας, οδηγίες για το πώς θέλουμε να γίνει το build, όπως επίσης και τα απαιτούμενα plug-in που πρέπει να εγκατασταθούν στο πρόγραμμά μας. Όλα αυτά τα διαβάζει το Maven και εκτελεί αυτόματα τις απαιτούμενες ενέργειες, απαλλάσσοντάς μας από αυτήν τη διαδικασία. Επιπλέον διαθέτει ένα σύνολο από εντολές μέσω των οποίων μπορούμε να εκτελέσουμε διάφορες ενέργειες όπως τη μεταγλώττιση και τη μετατροπή του κώδικά μας σε εκτελέσιμη μορφή, αλλά και πολλά άλλα.

Για περισσότερες πληροφορίες όσον αφορά το συγκεκριμένο εργαλείο, μπορείτε να επισκεφθείτε το επίσημο site του: <https://maven.apache.org/>

### 2.2.2.3 PostgreSQL

Όσον αφορά την υλοποίηση του εξυπηρετητή χρειαστήκαμε μια βάση όπου θέλαμε να αποθηκεύουμε τα δεδομένα μας και να τα ζητάμε μετά για να τα εμφανίσουμε στην πλατφόρμα μας. Αφού ερευνήσαμε όλες τις πιθανές σχεσιακές βάσεις που μπορούσαμε να χρησιμοποιήσουμε, καταλήξαμε στην PostgreSQL. Πρόκειται για μια σχεσιακή βάση δεδομένων ανοιχτού κώδικα με πολλές δυνατότητες. Η ανάπτυξη της συγκεκριμένης βάσης διαρκεί ήδη πάνω από 30 χρόνια και βασίζεται σε μια αποδεδειγμένα καλή αρχιτεκτονική η οποία έχει δημιουργήσει μια ισχυρή αντίληψη των χρηστών της γύρω από την αξιοπιστία, την ακεραιότητα δεδομένων και την ορθή λειτουργία. Επίσης τρέχει σε όλα τα βασικά λειτουργικά συστήματα, είναι συμβατή με ACID<sup>7</sup> και περιλαμβάνει τους περισσότερους τύπους δεδομένων. Τέλος υποστηρίζει αποθήκευση μεγάλων

---

<sup>6</sup> Είναι η διαδικασία αυτοματοποίησης της δημιουργίας και ανάπτυξης λογισμικού και περιλαμβάνει την μεταγλώττιση του κώδικα, τη συσκευασία του εκτελέσιμου κώδικα και την εκτέλεση αυτοματοποιημένων τεστ πάνω σε αυτό.

<sup>7</sup> Στην πληροφορική, το ACID (ατομικότητα, συνέπεια, απομόνωση, μονιμότητα) είναι ένα σύνολο ιδιοτήτων το οποίο εγγυάται ότι οι συναλλαγές στη βάση δεδομένων λειτουργούν αξιόπιστα.

δυναμικών δεδομένων, όπως εικόνες, ήχο και βίντεο και διαθέτει εξαιρετικό εγχειρίδιο χρήσης.

Αν θέλετε να μάθετε περισσότερα για την συγκεκριμένη βάση δεδομένων, μπορείτε να επισκεφθείτε την επίσημη σελίδα της: <https://www.postgresql.org/>

## 2.3 Υλοποίηση Πελάτη

Σε αυτό το υποκεφάλαιο θα περιγράψουμε αναλυτικά το περιβάλλον, τις γλώσσες προγραμματισμού και οτιδήποτε χρειάστηκε για να υλοποιήσουμε το πρόγραμμα του πελάτη. Το πρόγραμμα αυτό είναι το ορατό κομμάτι της πλατφόρμας μας, δηλαδή αυτό που τελικά θα βλέπει ο χρήστης και μέσω του οποίου θα μπορεί να αλληλεπιδράσει με το σύστημα.

### 2.3.1 Περιβάλλον Ανάπτυξης

Στη συγκεκριμένη ενότητα θα αναφερθούμε στο περιβάλλον που χρησιμοποιήσαμε για την ανάπτυξη του κώδικα του προγράμματος πελάτη. Θα μπορούσαμε να χρησιμοποιήσουμε και σε αυτήν την περίπτωση το IntelliJ IDEA που αναφέραμε ως περιβάλλον ανάπτυξης του προγράμματος εξυπηρετητή καθώς υποστηρίζει και την ανάπτυξη διαδικτυακών εφαρμογών και πιο συγκεκριμένα την ανάπτυξη Angular προγραμμάτων και τη συγγραφή κώδικα σε Typescript. Για την ανάπτυξη όμως του “πελάτη” επιλέξαμε να χρησιμοποιήσουμε ένα πιο ελαφρύ περιβάλλον ανάπτυξης που έχει αναπτυχθεί από την JetBrains και αυτό, το **WebStorm**. Ουσιαστικά την δουλειά που κάναμε με το WebStorm θα μπορούσαμε να την είχαμε κάνει και με το IntelliJ IDEA Ultimate, καθώς η συγκεκριμένη έκδοση υποστηρίζει τη συγγραφή προγραμμάτων Angular μέσω εγκατάστασης συγκεκριμένων plug-in. Οι βασικότεροι λόγοι λοιπόν που επιλέξαμε το WebStorm έναντι του IntelliJ IDEA Ultimate ήταν ότι δε χρειάζεται να εγκαταστήσουμε κάποιο επιπλέον plug-in καθώς το συγκεκριμένο IDE προορίζεται ακριβώς για την ανάπτυξη τέτοιων διαδικτυακών προγραμμάτων. Επίσης ένας άλλος λόγος που μας ώθησε στην επιλογή του συγκεκριμένου περιβάλλοντος ανάπτυξης ήταν ότι είναι ελαφρύτερο, ως προς την κατανάλωση πόρων στο σύστημα που τρέχει, σε σύγκριση με το IntelliJ IDEA Ultimate, καθώς υποστηρίζει πολύ λιγότερες γλώσσες προγραμματισμού σε αντίθεση με το δεύτερο που είναι γενικότερο και υποστηρίζει πολύ περισσότερες γλώσσες και τεχνολογίες. Τέλος μας διευκόλυνε πολύ στη δημιουργία και την διαχείριση των προγραμμάτων μας καθώς επιτρέπει το απευθείας άνοιγμα ενός φακέλου που έχει δημιουργηθεί είτε τοπικά είτε έχει γίνει αντιγραφή από ένα άλλο απομακρυσμένο μηχάνημα, σε αντίθεση με το IDEA που δεν προσφέρει αυτήν τη δυνατότητα.

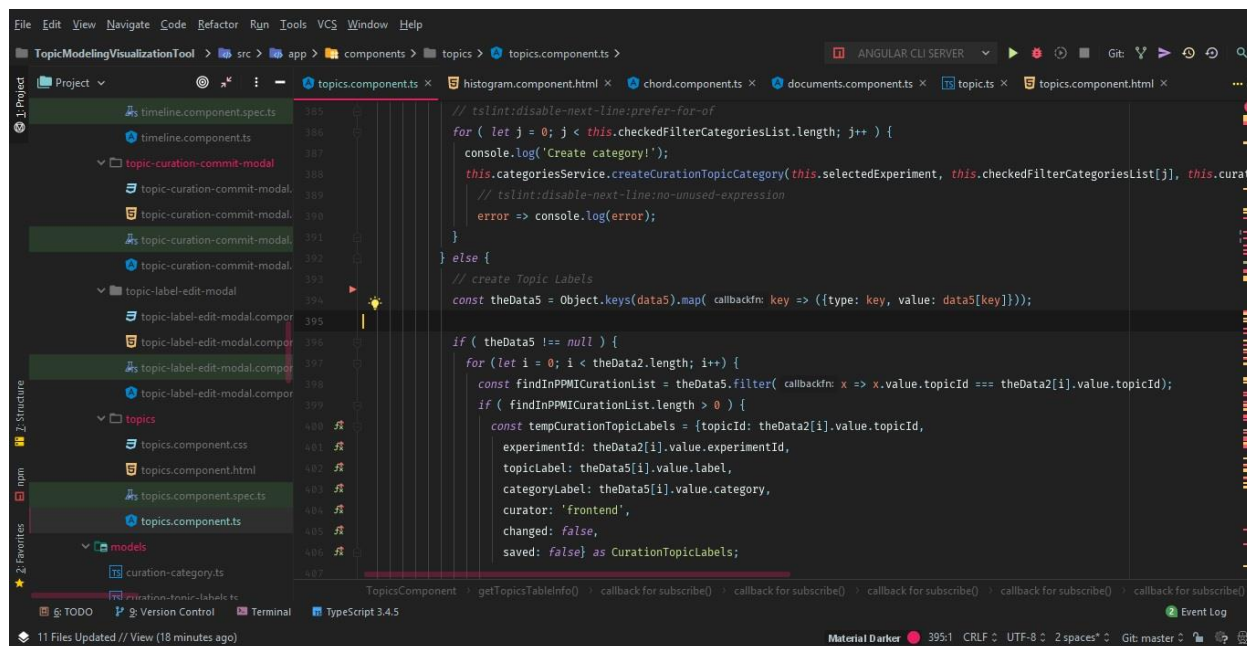
Πρόκειται για μια πλατφόρμα η οποία προορίζεται κυρίως για την ανάπτυξη διαδικτυακών εφαρμογών. Αποτελεί ένα JavaScript IDE και μεταξύ άλλων υποστηρίζει ανάπτυξη κώδικα σε Angular, React, Vue.js, Node.js και Meteor, τη δημιουργία εφαρμογών για κινητές συσκευές μέσω του Cordova, του Ionic και του React Native αλλά και desktop εφαρμογών<sup>8</sup> μέσω της χρήσης του Electron.

Επιπλέον παρέχει αρκετές δυνατότητες που το κάνουν να ξεχωρίζει και να προτιμάται στην ανάπτυξη διαδικτυακών εφαρμογών. Όπως αναφέραμε και προηγουμένως δεν απαιτεί την εγκατάσταση κάποιου plug-in για την ανάπτυξη προγραμμάτων JavaScript. Διαθέτει μεγάλη ποικιλία προεγκατεστημένων εργαλείων και framework που κάνουν

---

<sup>8</sup> Εφαρμογές που προορίζονται για σταθερούς υπολογιστές.

πολύ εύκολη τη δημιουργία τέτοιων εφαρμογών. Επιπλέον συμβάλει στην συγγραφή καλύτερου κώδικα, καθώς διαθέτει σύστημα που σε προειδοποιεί για πιθανά λάθη και ασυμβατότητες με βιβλιοθήκες. Επίσης σώζει πολύ προγραμματιστικό χρόνο διότι διαθέτει αυτόματη συμπλήρωση και σύστημα προτάσεων για μεταβλητές και συναρτήσεις, ανάλογα με τη συχνότητα εμφάνισής τους σε ένα συγκεκριμένο κομμάτι κώδικα. Μια ακόμη σημαντική δυνατότητα που προσφέρει το WebStorm είναι η γρήγορη αναζήτηση κλάσεων, συναρτήσεων, μεταβλητών σε όλα τα αρχεία του προγράμματος με μια ματιά. Τέλος, δίνει τη δυνατότητα για αποσφαλμάτωση του προγράμματος γραμμή γραμμή, τον έλεγχο του με unit testing αλλά και τη χρήση συστημάτων όπως το Git και το GitHub για τον έλεγχο και διαχείριση των διαφόρων εκδόσεων του κώδικά μας.



Εικόνα 2: WebStorm

### 2.3.2 Γλώσσες Προγραμματισμού και βιβλιοθήκες

Στην ενότητα αυτήν θα αναφερθούμε στις βασικές γλώσσες προγραμματισμού και τις βιβλιοθήκες που χρησιμοποιήσαμε για τη συγγραφή του προγράμματος πελάτη της πλατφόρμας. Όπως αναφέραμε και προηγουμένως, βασιστήκαμε σε Web γλώσσες προγραμματισμού και γενικά σε Web τεχνολογία. Για τη δημιουργία λοιπόν της πλατφόρμας ήταν απαραίτητη η χρήση του web framework της Angular σε συνδυασμό με HTML, JavaScript και Typescript και τις βιβλιοθήκες της Bootstrap, της D3.js και της Intro.js.

#### 2.3.2.1 Angular (HTML, JavaScript, Typescript)

Είναι ένα web framework που χρησιμοποιείται για την ανάπτυξη web εφαρμογών γραμμένων σε JavaScript, HTML και Typescript (η οποία είναι μια γλώσσα υπερσύνολο της JavaScript) για φορητούς και επιτραπέζιους υπολογιστές. Είναι ένα ανοιχτού λογισμικού web framework που συντηρείται από την Google. Έχει πολλές προεγκατεστημένες λειτουργίες για κινούμενα σχέδια, υπηρεσίες http και πολλά άλλα. Επίσης παρέχει πολλά επαναχρησιμοποιήσιμα και φιλικά προς τον χρήστη στοιχεία όπως αυτόματη συμπλήρωση, πλοήγηση, toolbar, menus κ.α τα οποία μπορεί να χρησιμοποιήσει ένας προγραμματιστής κατά τη δημιουργία της δικής του εφαρμογής χωρίς να χρειάζεται να τα υλοποιήσει ο ίδιος από την αρχή. Διευκολύνει έτσι πολύ έναν

προγραμματιστή να χτίσει γρήγορα μια κλασική web εφαρμογή γραμμένη σε HTML και JavaScript ή Typescript.

Για τη διαχείριση όλων αυτών των στοιχείων χρησιμοποιείται η Typescript, η οποία είναι μια scripting γλώσσα βασισμένη και γραμμένη πάνω στην JavaScript. Η Typescript αν και μοιάζει πολύ με την JavaScript εισάγει πολλές νέες ιδέες. Μεταξύ άλλων μια από τις σημαντικότερες ιδέες που εισάγει η Typescript είναι το data binding. Με βάση αυτό μπορούμε να δώσουμε τιμή σε κάποιο στοιχείο της HTML σελίδας μας βάση κάποιας έκφρασης η οποία θα υπολογιστεί κατά τη διάρκεια φόρτωσης της σελίδας μας και όχι δημιουργίας του συγκεκριμένου στοιχείου της σελίδας. Με άλλα λόγια, δε χρειάζεται να ορίζουμε ρητά τιμές αλλά να θέτουμε παραμετροποιημένες τιμές στα HTML στοιχεία της σελίδας. Μπορούμε να βάζουμε μεταβλητές ή συναρτήσεις οι οποίες όταν υπολογιστούν θα επιστρέψουν την κατάλληλη τιμή στο πεδίο που επιθυμούμε. Για αυτόν τον λόγο χρησιμοποιείται και ένα scope όπου θα αποθηκεύονται όλες οι (Angular) μεταβλητές και συναρτήσεις της εφαρμογής.

Πέρα από το binding, όμως η Angular εισάγει και άλλες έννοιες, όπως είναι το component, το module, τα models και την επαναχρησιμοποίηση HTML template μέσα στην HTML σελίδα μέσω της χρήσης ενός component εντός άλλου. Γενικά προωθεί την τμηματοποίηση του κώδικα και τη δημιουργία αυτόνομων μικρότερων κομματιών με το καθένα να έχει τη δική του λειτουργικότητα. Με αυτόν τον τρόπο επιτυγχάνεται η δημιουργία δομημένου και επαναχρησιμοποιήσιμου κώδικα, μειώνοντας έτσι το χρόνο συγγραφής μια εφαρμογής.

Έχει κάποια άλλα σημαντικά πλεονεκτήματά της Angular που μας οδήγησαν στην επιλογή της για την ανάπτυξη της πλατφόρμας μας είναι τα ακόλουθα:

- Όπως είπαμε και προηγουμένως έχει αναπτυχθεί και υποστηρίζεται από την Google. Αυτό σημαίνει ότι υπάρχει μια μεγάλη ενεργή κοινότητα που μπορεί να ανταποκριθεί γρήγορα και να σε βοηθήσει σε οποιοδήποτε πρόβλημα αντιμετωπίσεις κατά τη συγγραφή του προγράμματός σου σε Angular.
- Είναι πολύ εύκολη στην εκμάθηση δεδομένου ότι έχει βασιστεί στην JavaScript, την οποία γνωρίζει το σύνολο των προγραμματιστών web εφαρμογών.
- Τέλος, βοηθάει στην ανάπτυξη εφαρμογών εκτελέσιμων σε όλα τα λειτουργικά συστήματα χωρίς να χρειάζεται να αλλάξει πολύ ο κώδικας για την εκτέλεση από σύστημα σε σύστημα.

Για την ανάπτυξη της πλατφόρμας μας χρησιμοποιήσαμε συγκεκριμένα Angular 8, διότι είναι η τελευταία stable έκδοση που έχει κυκλοφορήσει η Google στην αγορά.

Αν επιθυμείτε να μάθετε περισσότερα για το Angular μπορείτε να επισκεφθείτε το επίσημο site του: <https://angular.io/> .

### 2.3.2.2 Bootstrap

Πρόκειται για το δημοφιλέστερο front – end framework που υπάρχει αυτήν τη στιγμή στην αγορά. Είναι μια ανοιχτού λογισμικού συλλογή εργαλείων που χρησιμοποιείται για τη δημιουργία ιστοχώρων και διαδικτυακών εφαρμογών. Περιλαμβάνει ένα σύνολο έτοιμων HTML στοιχείων και CSS εντολών τα οποία μπορεί να χρησιμοποιήσει ένας προγραμματιστής ως έτοιμα template στο site που φτιάχνει. Έτσι μεταξύ άλλων ο προγραμματιστής μπορεί να πάρει έτοιμα buttons, dropdown menus, forms, navbars, breadcrumbs, thumbnails, lists, progress bars, icons και πολλά άλλα. Εκτός όμως από τα HTML στοιχεία και τα CSS χαρακτηριστικά τους, το Bootstrap προσφέρει στον χρήστη και JavaScript υποστήριξη. Κάθε Bootstrap HTML στοιχείο λοιπόν έχει ένα



πακέτο συναρτήσεων και JavaScript κώδικα, που δίνουν στον προγραμματιστή τη δυνατότητα να διαχειριστεί το συγκεκριμένο στοιχείο. Έτσι για παράδειγμα μια Bootstrap φόρμα έχει ενσωματωμένο έλεγχο των στοιχείων που δίνονται στην είσοδο (validation), μια Bootstrap toolbar περιέχει λειτουργία που σου επιτρέπει να διαχειρίζεσαι κάποιο κρυφό menu, μια Bootstrap list μπορεί να ανοιγοκλείνει και να μετατρέπεται σε «ακορντεόν» εξοικονομώντας χώρο, και πολλά άλλα. Όλα αυτά μπορεί να τα χρησιμοποιήσει ένας προγραμματιστής στον κώδικά του και να τα διαχειρίζεται σαν φυσική προέκταση αυτού.

Πολύ σημαντικά και σίγουρα πολύ εύχρηστα όλα αυτά, αλλά άλλο είναι αυτό που ξεχωρίζει το Bootstrap από όλα τα υπόλοιπα front – end frameworks που υπάρχουν στην αγορά. Το Bootstrap διαθέτει ένα πολύ καλό Grid σύστημα, που σου δίνει τη δυνατότητα δημιουργίας responsive ιστοσελίδων και διαδικτυακών εφαρμογών. Τι είναι όμως το responsive που αναφέραμε;; Ένα website ή μια εφαρμογή λέγεται responsive όταν έχει φτιαχτεί με τέτοιο τρόπο ώστε να ανταποκρίνεται και να διαμορφώνεται ανάλογα με την ανάλυση της οθόνης στην οποία εμφανίζεται. Έτσι μέσω του Bootstrap μπορείς να φτιάξεις websites τα οποία θα μπορούν να εμφανίζονται σωστά από οποιαδήποτε κινητή συσκευή και αν ανοιχτούν. Όπως μπορούμε λοιπόν εύκολα να παρατηρήσουμε, η χρήση του Bootstrap για τη δημιουργία μιας διαδικτυακή εφαρμογής κρίνεται απαραίτητη από τη στιγμή που μια τέτοια εφαρμογή τρέχει κυρίως σε κινητές συσκευές των οποίων η ανάλυση ποικίλει.

Πώς όμως επιτυγχάνεται το responsiveness μέσω του Bootstrap;; Για αυτόν το σκοπό το Bootstrap έχει ένα σύστημα το οποίο λέγεται Grid System. Στην παλιά Bootstrap η οθόνη χωριζόταν σε γραμμές και spans και κάθε γραμμή είχε το πολύ 12 spans. Στην Bootstrap 3 που είναι αυτή που χρησιμοποιείται τώρα, η οθόνη χωρίζεται σε γραμμές και στήλες. Έτσι ο προγραμματιστής χρησιμοποιώντας έτοιμες grid κλάσεις μπορεί εύκολα και γρήγορα να φτιάξει responsive οθόνες, που θα παίζουν εξίσου καλά σε κινητά, tablets, laptops και επιτραπέζιους υπολογιστές. Για παράδειγμα μπορείς να χρησιμοποιήσεις την κλάση `.col-xs-*` για να φτιάξεις grid στήλες σε πολύ μικρές αναλύσεις όπως αυτή του κινητού, την κλάση `.col-sm-*` για να φτιάξεις grid στήλες για tablets και τις κλάσεις `.col-md-*` και `.col-lg-*` για να φτιάξεις grid στήλες για laptops και desktops αντίστοιχα.

Με αυτόν τον τρόπο μπορείς να φτιάξεις ένα responsive website και να ελέγξεις πώς θα φαίνεται στις διάφορες συσκευές ανάλογα το μέγεθος της οθόνης τους.



**Εικόνα 3: Bootstrap Grid System για responsiveness**

Δεδομένου λοιπόν ότι στην Bootstrap 3 κάθε γραμμή αποτελείται από 12 «κουτιά περιεχομένου» (στήλες), όπως βλέπουμε και στην **Εικόνα 3** αυτές οι στήλες αναπαρίστανται με διαφορετικό τρόπο ανάλογα το μέγεθός της οθόνης της συσκευής. Έτσι σε έναν desktop έχουμε 3 γραμμές (rows) με 4 στήλες (cols) η καθεμία, σε ένα laptop έχουμε 4 γραμμές (rows) με 3 στήλες (cols) η καθεμία, σε ένα tablet έχουμε 6

rows με 2 cols η καθεμία και σε ένα κινητό έχουμε 12 διαδοχικές γραμμές. Με αυτόν τον τρόπο κατορθώνουμε να αναπαραστήσουμε το site μας σε οποιαδήποτε συσκευή ανεξαρτήτου μεγέθους οθόνης. Ακόμη το Bootstrap σου δίνει τη δυνατότητα ανάλογα το μέγεθος της οθόνης να κρύψεις σε κάποιες οθόνες (τις μικρότερες) κάποια στοιχεία και να εμφανίσεις κάποια άλλα αξιοποιώντας έτσι κατάλληλα των χώρο της οθόνης που σου δίνει κάθε συσκευή.

Αυτή η δυνατότητα της δημιουργίας responsive διαδικτυακών εφαρμογών σε συνδυασμό με τα έτοιμα στοιχεία που σου προσφέρει η Bootstrap την καθιστούν ένα πολύ σημαντικό εργαλείο στο web design και development και αυτός είναι ο λόγος που χρησιμοποιήσαμε κι εμείς το συγκεκριμένο framework στην εργασία.

Για όποιον ενδιαφέρετε και θέλει να μάθει περισσότερα σχετικά με την Bootstrap μπορεί να επισκεφθεί το επίσημο site της: <http://getbootstrap.com/> ή κάποιο tutorial όπως αυτό του w3schools: <http://www.w3schools.com/bootstrap/default.asp> .

### 2.3.2.3 D3.js

Η D3.js είναι μια βιβλιοθήκη οπτικοποίησης δεδομένων. Χρησιμοποιήθηκε στην πλατφόρμα μας για τη δημιουργία γραφικής απεικόνισης ιστογραμμάτων (histograms), ραβδογραμμάτων (barcharts), διαγράμματος χορδών (chord) και διαγράμματος ροής (sankey). Η συγκεκριμένη βιβλιοθήκη δίνει τη δυνατότητα της εύκολης δημιουργίας εντυπωσιακών γραφημάτων από μια είσοδο δεδομένων σε πολύ απλή μορφή. Μπορείς για παράδειγμα χρησιμοποιώντας έναν πίνακα αριθμών, να δημιουργήσεις με την ίδια ευκολία από έναν απλό HTML πίνακα δεδομένων μέχρι ένα ραβδόγραμμα με εφέ κίνησης και δυνατότητα αλληλεπίδρασης του χρήστη με αυτό. Με αυτόν τον τρόπο η D3 επιτυγχάνει την αναπαράσταση των ίδιων δεδομένων με πολλά διαφορετικά γραφήματα εμπλουτίζοντας έτσι την εφαρμογή μας οπτικά και δίνοντας μια ευελιξία δημιουργίας γραφημάτων στον προγραμματιστή. Όπως χαρακτηριστικά αναγράφεται στην ιστοσελίδα της, η D3.js δίνει “ζωή” στα δεδομένα χρησιμοποιώντας HTML, SVG και CSS.

Για μια εκτενέστερη περιγραφή της βιβλιοθήκης D3.js μπορείτε να επισκεφθείτε την επίσημη ιστοσελίδα της: <https://d3js.org/> .

### 2.3.2.4 Introjs

Είναι μια βιβλιοθήκη της JavaScript, πολύ ελαφριά ως προς την χρήση πόρων του συστήματος. Χρησιμοποιείται για τη δημιουργία ξενάγησης σε μια ιστοσελίδα ή μια εφαρμογή. Ένας χρήστης που μπαίνει πρώτη φορά σε μια εφαρμογή δεν μπορεί να γνωρίζει όλες τις λειτουργίες που του παρέχονται για αλληλεπίδραση με την διεπαφή. Έτσι δημιουργώντας μια ξενάγηση στην εφαρμογή, είναι εύκολο να παρουσιάσουμε και να εξηγήσουμε στο χρήστη αυτές τις λειτουργίες.

Στην πλατφόρμα μας ενσωματώσαμε λοιπόν τη συγκεκριμένη βιβλιοθήκη, προκειμένου να παρέχουμε βοηθητικές οδηγίες προς του χρήστες για τον χειρισμό της.

Δε θα αναφερθούμε περισσότερο στη συγκεκριμένη βιβλιοθήκη γιατί μια εκτενέστερη περιγραφή είναι εκτός των ορίων της συγκεκριμένης εργασίας. Για όποιον επιθυμεί να μάθει περισσότερα μπορεί να επισκεφθεί την επίσημη ιστοσελίδα της: <https://introjs.com/> .

## 2.4 Συμπεράσματα

Στο συγκεκριμένο κεφάλαιο αναφερθήκαμε στα πιο τεχνικά ζητήματα της εργασίας, δηλαδή σε ότι σχετίζεται με τον κώδικα της πλατφόρμας και την συγγραφή του. Πιο συγκεκριμένα, μιλήσαμε για το σύστημα στο οποίο αναπτύχθηκε η εφαρμογή και παρουσιάσαμε τα εργαλεία τα οποία χρησιμοποιήσαμε για τη συγγραφή του κώδικα τόσο στην πλευρά του εξυπηρετητή όσο και σε αυτήν του πελάτη. Έπειτα αναφέραμε μία προς μία τις γλώσσες προγραμματισμού και τις βιβλιοθήκες που χρησιμοποιήσαμε και κάναμε μια συνοπτική παρουσίασή τους. Κλείνοντας το παρόν κεφάλαιο θα θέλαμε να επισημάνουμε, ότι οι τεχνολογίες που χρησιμοποιήσαμε για την ανάπτυξη της εφαρμογής είναι στην αιχμή της τεχνολογίας και είναι αυτές που προτιμώνται από το σύνολο σχεδόν των προγραμματιστών για την ανάπτυξη ανάλογων εφαρμογών.



### 3. ΠΑΡΟΥΣΙΑΣΗ ΕΡΓΑΛΕΙΟΥ

#### 3.1 Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζεται το εργαλείο και οι λειτουργικότητές του. Ο χρήστης μπορεί να επιλέξει ένα πείραμα (συλλογή από topics) και να εξερευνήσει τα topic που ανήκουν σε αυτό μέσω πινάκων αλλά και γραφημάτων. Αφού συλλέξει όποιες πληροφορίες χρειάζεται μπορεί να προσθέσει κάποια ετικέτα στα topic, να τα κατηγοριοποιήσει σε συγκεκριμένες κατηγορίες και να τα αξιολογήσει. Όλα αυτά, όπως και άλλες λειτουργικότητες του εργαλείου, περιγράφονται αναλυτικά στην ενότητα αυτή.

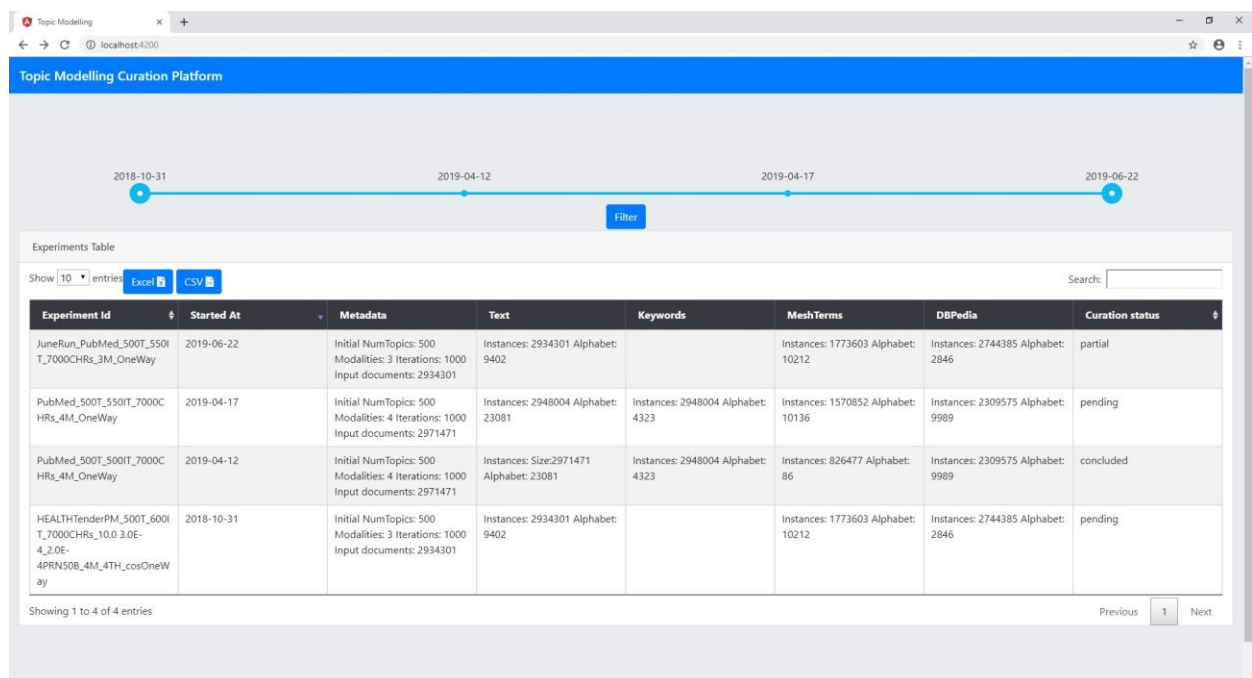
#### 3.2 Παρουσίαση πλατφόρμας

Η πλατφόρμα αποτελείται από τρεις βασικές οθόνες, μία για τα experiments, μία για τα topics και μία για τα documents. Με αυτή τη σειρά που τις αναφέρουμε εμφανίζονται και μες το εργαλείο και εκφράζεται έτσι η λογική αλληλουχία που υπάρχει μεταξύ αυτών των τριών οντοτήτων. Ένα experiment είναι η έξοδος που παίρνουμε από το topic model και αποτελεί μια συλλογή από topics με πληροφορίες για αυτά. Το topic είναι η βασική οντότητα του εργαλείου μας και αποτελεί την είσοδο για τη διαδικασία του curation που μπορεί να υλοποιήσει ένας curator μέσω της εφαρμογής μας. Η τελευταία βασική οντότητα είναι το document. Ένα topic έχει πολλά documents που αναφέρονται σε αυτό. Μέσα από την εξερεύνηση των documents, ο curator μπορεί να συλλέξει πληροφορίες για το topic, οι οποίες θα του φανούν χρήσιμες στην αξιολόγησή του.

Παρακάτω θα παρουσιάσουμε λοιπόν μία προς μία αυτές τις οθόνες και με τη χρήση εικόνων μέσα από το εργαλείο, θα αναλύσουμε τη δομή και τις λειτουργικότητες καθεμιάς από αυτές.

##### 3.2.1 Οθόνη των Experiments

Στην **Εικόνα 4** φαίνεται η αρχική σελίδα του εργαλείου.

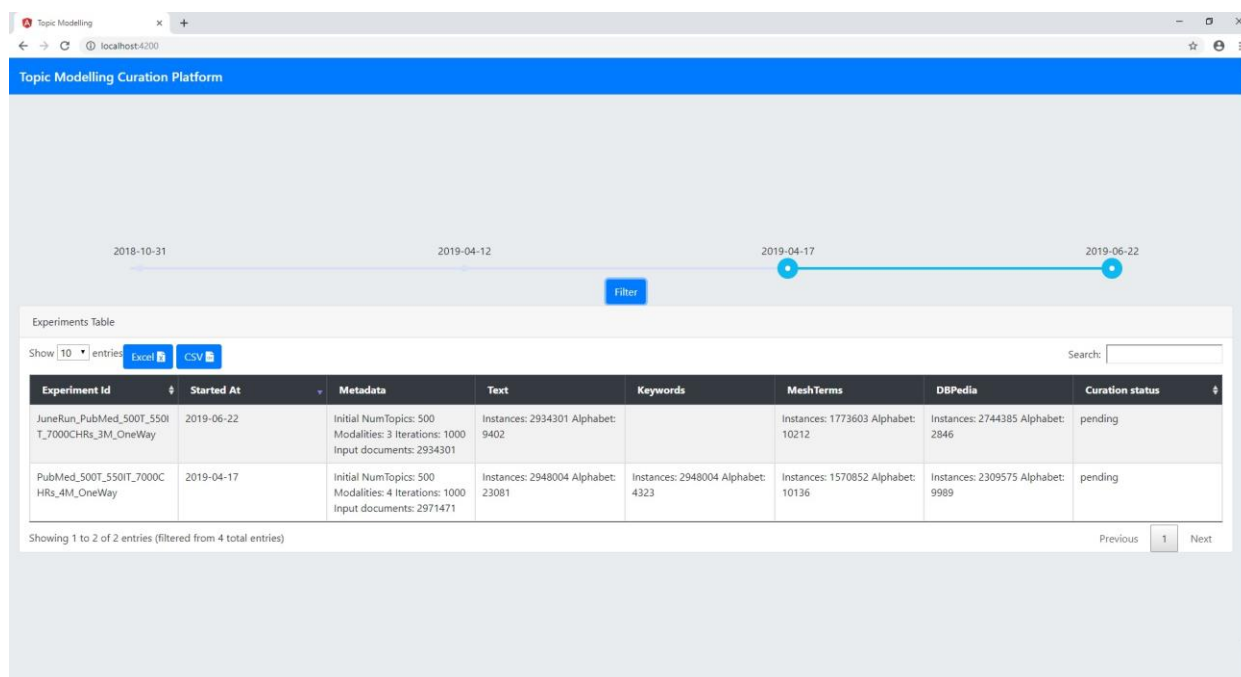


| Experiment Id   | Started At | Metadata   | Text                                    | Keywords                          | MeshTerms                          | DBpedia                           | Curation status |
|---|------------|--|---|-----------------------------------|------------------------------------|-----------------------------------|-----------------|
| JuneRun_PubMed_500T_550I_T_7000CHRS_3M_OneWay                                 | 2019-06-22 | Initial NumTopics: 500<br>Modalities: 3 Iterations: 1000<br>Input documents: 2934301 | Instances: 2934301 Alphabet: 9402       |                                   | Instances: 1773603 Alphabet: 10212 | Instances: 2744385 Alphabet: 2846 | partial         |
| PubMed_500T_550IT_7000C_HRs_4M_OneWay   | 2019-04-17 | Initial NumTopics: 500<br>Modalities: 4 Iterations: 1000<br>Input documents: 2971471 | Instances: 2948004 Alphabet: 23081      | Instances: 2948004 Alphabet: 4323 | Instances: 1570852 Alphabet: 10136 | Instances: 2309575 Alphabet: 9989 | pending         |
| PubMed_500T_500IT_7000C_HRs_4M_OneWay   | 2019-04-12 | Initial NumTopics: 500<br>Modalities: 4 Iterations: 1000<br>Input documents: 2971471 | Instances: Size:2971471 Alphabet: 23081 | Instances: 2948004 Alphabet: 4323 | Instances: 826477 Alphabet: 86     | Instances: 2309575 Alphabet: 9989 | concluded       |
| HEALTHTenderPM_500T_600I_T_7000CHRS_10.0.3.0E-4.2.0E-4PRN50B_4M_4TH_CosOneWay | 2018-10-31 | Initial NumTopics: 500<br>Modalities: 3 Iterations: 1000<br>Input documents: 2934301 | Instances: 2934301 Alphabet: 9402       |                                   | Instances: 1773603 Alphabet: 10212 | Instances: 2744385 Alphabet: 2846 | pending         |

Εικόνα 4: Αρχική οθόνη εργαλείου

Βασικό στοιχείο της συγκεκριμένης οθόνης είναι ο πίνακας με τα experiments. Στο συγκεκριμένο πίνακα ο χρήστης μπορεί να δει ποιες είναι τα διαθέσιμα run του Topic Modeling που μπορεί να κάνει curate, να συλλέξει κάποιες συνοπτικές πληροφορίες για αυτά, να δει ποια experiments έχουν ήδη αξιολογηθεί, ποια αξιολογούνται τώρα και ποια δεν έχουν αξιολογηθεί καθόλου. Επίσης μέσω του πίνακα μπορεί να αναζητήσει συγκεκριμένα experiment, να τα ταξινομήσει βάσει ονόματος ή ακόμη και βάσει της ημερομηνίας κατά την οποία πραγματοποιήθηκε κάθε run του Topic Modeling.

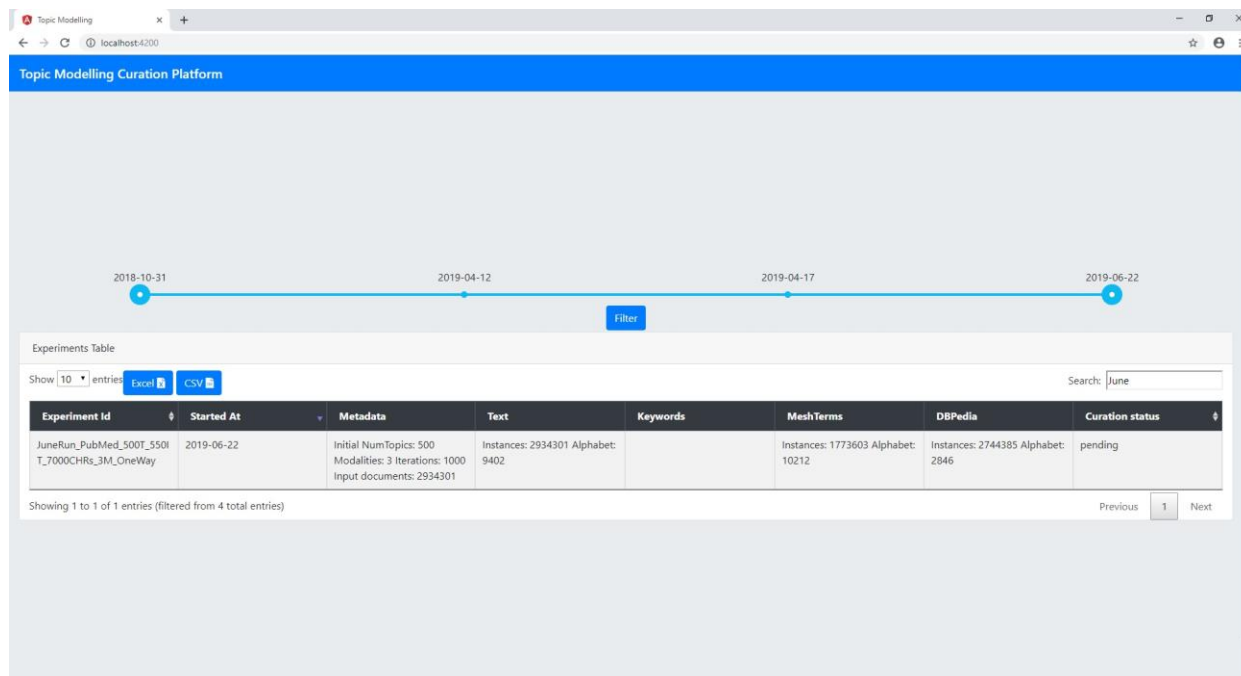
Πιο αναλυτικά λοιπόν, στο επάνω μέρος της οθόνης υπάρχει ένας slider μέσω του οποίου ο χρήστης μπορεί να επιλέξει τη χρονική περίοδο του ενδιαφέρεται να μελετήσει και πατώντας το κουμπί “Filter” να φιλτράρει μόνο τα experiments που ανήκουν εντός αυτής της χρονικής περιόδου. Στην **Εικόνα 5** φαίνεται ένα παράδειγμα όπου από όλα τα experiments που μπορούμε να μελετήσουμε και να αξιολογήσουμε, όπως φαίνεται στην Εικόνα 5, εμείς επιλέξαμε να ασχοληθούμε μόνο με αυτά που ανήκουν εντός της χρονικής περιόδου 17/04/2019 – 22/06/2019.



**Εικόνα 5: Φιλτράρισμα των experiments εντός συγκεκριμένης χρονικής περιόδου**

Κάτω από τον slider ο χρήστης βλέπει τον πίνακα με τα experiments. Κάθε γραμμή του αφορά ένα experiment. Ο χρήστης μπορεί να πάρει πληροφορίες ως προς το όνομα, την ημερομηνία που παρήχθησαν τα δεδομένα του συγκεκριμένου experiment από το topic model, πληροφορίες για τα τεχνικά χαρακτηριστικά, τις παραμέτρους και την είσοδο που χρησιμοποιήθηκε για την παραγωγή αυτών των δεδομένων, όπως επίσης και την κατάσταση στην οποία βρίσκεται το curation του καθενός. Ανάλογα λοιπόν με τον όρο που θα δει στην τελευταία στήλη, pending, partial ή concluded, μπορεί να ξέρει αντίστοιχα αν το curation στο συγκεκριμένο experiment δεν έχει αρχίσει καν, έχει αρχίσει αλλά δεν έχει ολοκληρωθεί ή αν έχει ήδη ολοκληρωθεί. Επιπλέον ο χρήστης μπορεί ανά πάσα στιγμή πατώντας το κουμπί που γράφει Excel ή εκείνο που γράφει CSV να κατεβάσει τα δεδομένα του πίνακα στην αντίστοιχη μορφή αρχείου και να τα μελετήσει για παράδειγμα στο σπίτι του με την ησυχία του. Επίσης επιλέγοντας από το μενού αριστερά από αυτά τα κουμπιά, μπορεί να διαλέξει το πλήθος των γραμμών του πίνακα που εμφανίζονται ανά σελίδα και να πλοηγηθεί στις διάφορες σελίδες του πίνακα με τη σελιδοποίηση που υπάρχει κάτω δεξιά στον πίνακα. Στη συγκεκριμένη περίπτωση έχουμε λίγα experiments, γι'αυτό και βλέπουμε μόνο μία σελίδα. Στις

επόμενες οθόνες που έχουμε περισσότερα αποτελέσματα θα φανεί καλύτερα αυτή η λειτουργικότητα. Μια τελευταία δυνατότητα που έχει ο χρήστης όσον αφορά αυτόν τον πίνακα είναι να αναζητήσει ένα ή περισσότερα experiments μέσω keywords. Χρησιμοποιώντας λοιπόν το πεδίο αναζήτησης που υπάρχει στο δεξί μέρος πάνω από τον πίνακα, όπως φαίνεται και στην **Εικόνα 6**, γράφοντας “June” μπορεί να βρει τα experiments εκείνα που έχουν τη συγκεκριμένη λέξη σε οποιαδήποτε από τις στήλες τους, καθώς η αναζήτηση πραγματοποιείται στο σύνολο των στηλών του πίνακα.



**Εικόνα 6: Φιλτράρισμα των experiments βάσει κάποιου keyword**

Κλείνοντας την ανάλυση της Αρχικής οθόνης του εργαλείου μας, παρατηρούμε ότι η συγκεκριμένη οθόνη παρέχει περισσότερη πληροφορία απ’όση μπορεί να χρειάζεται ένας curator. Θα μπορούσαμε να πούμε ότι δεν απευθύνεται τόσο σε έναν curator που έχει αναλάβει συγκεκριμένα experiment να μελετήσει και να αξιολογήσει. Απευθύνεται περισσότερο σε έναν administrator, που καταλαβαίνει όλες αυτές τις πληροφορίες και τα τεχνικά χαρακτηριστικά που παρέχονται και ο οποίος μπορεί να έχει μια ολοκληρωμένη εικόνα ώστε για παράδειγμα να μοιράσει τα experiments, που δεν έχουν γίνει curate, στους διάφορους curators.

Αφού λοιπόν ο χρήστης πάρει όση πληροφορία χρειάζεται μπορεί να επιλέξει το experiment που θέλει να μελετήσει-αξιολογήσει πατώντας πάνω σε αυτό.

### 3.2.2 Οθόνη των Topics

Πατώντας στο experiment που θέλει να μελετήσει ο χρήστης μεταφέρεται στην οθόνη των topics, που είναι η οθόνη με την περισσότερη λειτουργικότητα στο εργαλείο και ίσως και αυτή που έχει το μεγαλύτερο ενδιαφέρον για τον curator, καθώς σε αυτό το σημείο μπορεί να υλοποιήσει τη διαδικασία του curation στο σύνολό της μέσω διαφόρων επιλογών που του παρέχονται. Μπαίνοντας λοιπόν στη συγκεκριμένη σελίδα όπως φαίνεται και στην **Εικόνα 7** δεσπόζει ο πίνακας των topics. Κάθε γραμμή του συγκεκριμένου πίνακα αναφέρεται σε ένα συγκεκριμένο topic.

Παρατηρώντας τον πίνακα βλέπουμε ότι από τις δέκα στήλες, οι έξι έχουν ένα πιο ανοιχτό γκρι στην επικεφαλίδα τους. Αυτό δείχνει ότι από τις συγκεκριμένες στήλες μπορεί μόνο να αντλήσει πληροφορία για το εκάστοτε topic και δεν έχει τη δυνατότητα

να τις επεξεργαστεί. Μέσα από αυτές στην ουσία μπορεί να έχει μια συνοπτική εικόνα που θα τον βοηθήσει στην αξιολόγηση του topic. Η πληροφορία που του παρέχεται αφορά τα top-5 phrases, words, keywords, meshTerms και τη δυναμικότητα του topic, δηλαδή τη σημαντικότητά του σε σύγκριση με τα υπόλοιπα topics της συλλογής. Μπορεί λοιπόν να ταξινομήσει τα στοιχεία του πίνακα βάσει κάποιας από αυτές τις στήλες και να πάρει την πληροφορία που χρειάζεται. Για παράδειγμα μπορεί να ταξινομήσει τα topics βάσει του topic strength και να βρει το πιο ισχυρό και το πιο αδύναμο topic στο experiment. Στο σημείο αυτό αξίζει να αναφερθεί ότι αυτοί οι τύποι δεδομένων αφορούν το συγκεκριμένο run. Οι τύποι (words, phrases, keywords κλπ) θα προκύπτουν δυναμικά ανάλογα το περιεχόμενο των κειμένων που χρησιμοποιείται στο Topic Modeling. Συνεπώς σε κάποιο άλλο run που τα κείμενα δεν αφορούν ιατρικά δεδομένα, δεν θα έχουμε το meshTerms που εμφανίζεται στο παράδειγμα που χρησιμοποιούμε για την παρουσίαση του εργαλείου μας στο παρόν κείμενο.

| Topic Id | Top-5 Phrases   | Top-5 Words   | Top-5 Keywords   | Top-5 MeshTerms   | Topic Strength | Curation   | Topic Label                          | Category Label                                    |
|----------|---|---|--|---|----------------|--|--------------------------------------|---|
| 0        | <ul style="list-style-type: none"> <li>cardiovascular disease cvd</li> <li>cardiovascular disease</li> <li>cardiovascular diseases</li> <li>cardiovascular events</li> <li>cardiovascular risk</li> </ul> | <ul style="list-style-type: none"> <li>metformin</li> <li>cvd</li> <li>cardiovascular risk</li> <li>disease</li> </ul>          | <ul style="list-style-type: none"> <li>cardiovascular diseases diagnosis</li> <li>metformin pharmacology</li> <li>cardiovascular diseases prevention &amp; control</li> <li>cardiovascular diseases epidemiology</li> <li>hypoglycemic agents therapeutic use</li> </ul> | <ul style="list-style-type: none"> <li>Metformin</li> <li>Cerebrovascular_disease</li> <li>Dipeptidyl_Lpeptidase-4</li> <li>Diabetes_mellitus_type_2</li> <li>Anti-diabetic_medication</li> </ul> | 0.00155        | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | MicroRNAs and gene expression        | • 5. Molecular Cell Biology                       |
| 1        | <ul style="list-style-type: none"> <li>pseudomonas aeruginosa</li> <li>biofilm formation</li> <li>quorum sensing</li> <li>opportunistic pathogen</li> <li>bacterial biofilms</li> </ul>                   | <ul style="list-style-type: none"> <li>biofilm</li> <li>aeruginosa</li> <li>pseudomonas</li> <li>bacterial formation</li> </ul> | <ul style="list-style-type: none"> <li>biofilms growth &amp; development</li> <li>biofilms drug effects</li> <li>pseudomonas aeruginosa genetics</li> <li>pseudomonas aeruginosa drug effects</li> <li>bacterial proteins genetics</li> </ul>                            | <ul style="list-style-type: none"> <li>Biofilm</li> <li>Pseudomonas_aeruginosa</li> <li>Quorum_sensing</li> <li>Pseudomonas</li> <li>Fimbria_(bacteriology)</li> </ul>                            | 0.0018         | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | Prevention of athletic injuries      | • 3. Health Systems, Public Health & Epidemiology |
| 2        | <ul style="list-style-type: none"> <li>sweet taste</li> <li>dioxygenase ido</li> </ul>  | <ul style="list-style-type: none"> <li>athletes</li> <li>taste</li> </ul>   | <ul style="list-style-type: none"> <li>tastelphysiology</li> <li>athletic</li> </ul>   | <ul style="list-style-type: none"> <li>Hydrochloride</li> <li>Dioxygenase</li> </ul>  | 0.00099        | <ul style="list-style-type: none"> <li>👍</li> </ul>                        | Surgical techniques for treatment of | • 4. Diagnostics, Treatment Development & Surgery |

Εικόνα 7: Οθόνη των topics (Κύρια οθόνη του εργαλείου)

Συνεχίζοντας με την ανάλυση των στηλών του πίνακα, παρατηρούμε τις στήλες που έχουν στην επικεφαλίδα τους ένα πιο σκούρο γκρι. Αυτές είναι οι στήλες στις οποίες μπορεί να παρέμβει ο χρήστης και να επεξεργαστεί-αξιολογήσει το topic. Ξεκινώντας την ανάλυσή τους από αριστερά προς τα δεξιά, η πρώτη στήλη (πέρα από αυτήν των checkboxes) που βλέπουμε είναι αυτή της αξιολόγησης του topic (curation). Στη συγκεκριμένη στήλη ο curator μπορεί να χαρακτηρίσει το topic ως “good”, τουτέστιν ενδιαφέρον, πατώντας το πράσινο κουμπί, ως “poor”, δηλαδή κακό, δηλώνοντας έτσι ότι δεν τον ενδιαφέρει, πατώντας το κόκκινο κουμπί, είτε ακόμα και να δηλώσει ότι δεν έχει πάρει την απόφασή του και ότι θα το αξιολογήσει αργότερα, αφού το μελετήσει περαιτέρω, πατώντας το γκρι κουμπί. Πατώντας οποιοδήποτε από τα κουμπιά χρωματίζεται η γραμμή με το αντίστοιχο χρώμα, όπως φαίνεται στην **Εικόνα 8**, ώστε να μπορεί ο curator να ξεχωρίζει γρήγορα, με μια ματιά, ποια topics έκρινε ενδιαφέροντα και ποια όχι. Η επόμενη στήλη στον πίνακα έχει το όνομα “Topic Label” και σε αυτό το σημείο ο χρήστης μπορεί να δώσει μια περιγραφική ετικέτα στο topic ώστε διαβάζοντάς την να καταλαβαίνει κατευθείαν το περιεχόμενο του topic. Πατώντας λοιπόν στην στήλη αυτή για κάποιο topic, ανοίγει ένα αναδυόμενο παράθυρο, όπως φαίνεται στην **Εικόνα 9**, στο οποίο ο χρήστης μπορεί να εισάγει και να αποθηκεύσει καινούρια ετικέτα για το topic. Τελευταία στήλη στον πίνακα είναι το “Category Label”. Στη συγκεκριμένη στήλη

αναγράφεται η κατηγορία στην οποία ανήκει το κάθε topic. Πατώντας πάνω στην στήλη αυτή για κάποιο topic ανοίγει ένα παράθυρο, όπως φαίνεται στην **Εικόνα 10**, μέσω του οποίου ο χρήστης μπορεί να διαχειριστεί την/τις κατηγορίες του. Όπως παρατηρούμε λοιπόν στην Εικόνα 10, ο χρήστης μπορεί να επιλέξει και να απεπιλέξει όποια κατηγορία θέλει από την λίστα, όπως επίσης και να προσθέσει καινούρια κατηγορία πατώντας το “Create New” και εν συνεχεία αφού πληκτρολογήσει το όνομα της κατηγορίας και πατήσει “Add”, η νέα κατηγορία προστίθεται στην λίστα. Αφού ολοκληρώσει τις αλλαγές του, μπορεί να πατήσει “Save” ή “Cancel” για να αποθηκεύσει ή να ακυρώσει αντίστοιχα τις αλλαγές του.

| Topic Id | Top-5 Phrases   | Top-5 Words   | Top-5 Keywords   | Top-5 MeshTerms  | Topic Strength | Curation   | Topic Label                          | Category Label  |
|----------|---|---|--|--|----------------|--|--------------------------------------|---|
| 0        | <ul style="list-style-type: none"> <li>cardiovascular disease cvd</li> <li>cardiovascular disease</li> <li>cardiovascular diseases</li> <li>cardiovascular events</li> <li>cardiovascular risk</li> </ul> | <ul style="list-style-type: none"> <li>metformin</li> <li>cvd</li> <li>cardiovascular risk</li> <li>disease</li> </ul>          | <ul style="list-style-type: none"> <li>cardiovascular diseases diagnosis</li> <li>metformin pharmacology</li> <li>cardiovascular diseases prevention &amp; control</li> <li>cardiovascular diseases epidemiology</li> <li>hypoglycemic agents therapeutic use</li> </ul> | <ul style="list-style-type: none"> <li>Metformin</li> <li>Cerebrovascular_disease</li> <li>Dipeptidyl_peptidase-4</li> <li>Diabetes_mellitus_type_2</li> <li>Anti-diabetic_medication</li> </ul> | 0.00155        | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | MicroRNAs and gene expression        | <ul style="list-style-type: none"> <li>5. Molecular Cell Biology</li> </ul>                           |
| 1        | <ul style="list-style-type: none"> <li>pseudomonas aeruginosa</li> <li>biofilm formation</li> <li>quorum sensing</li> <li>opportunistic pathogen</li> <li>bacterial biofilms</li> </ul>                   | <ul style="list-style-type: none"> <li>biofilm</li> <li>aeruginosa</li> <li>pseudomonas</li> <li>bacterial formation</li> </ul> | <ul style="list-style-type: none"> <li>biofilms growth &amp; development</li> <li>biofilms drug effects</li> <li>pseudomonas aeruginosa genetics</li> <li>pseudomonas aeruginosa drug effects</li> <li>bacterial proteins genetics</li> </ul>                            | <ul style="list-style-type: none"> <li>Biofilm</li> <li>Pseudomonas_aeruginosa</li> <li>Quorum_sensing</li> <li>Pseudomonas</li> <li>Fimbria_(bacteriology)</li> </ul>                           | 0.0018         | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | Prevention of athletic injuries      | <ul style="list-style-type: none"> <li>3. Health Systems, Public Health &amp; Epidemiology</li> </ul> |
| 2        | <ul style="list-style-type: none"> <li>sweet taste</li> <li>dioxygenase ido</li> </ul>  | <ul style="list-style-type: none"> <li>athletes</li> <li>taste</li> </ul>   | <ul style="list-style-type: none"> <li>taste physiology</li> <li>athletic</li> </ul>   | <ul style="list-style-type: none"> <li>Hydrochloride</li> <li>Dioxygenase</li> </ul>   | 0.00099        | <ul style="list-style-type: none"> <li>👍</li> </ul>                        | Surgical techniques for treatment of | <ul style="list-style-type: none"> <li>4. Diagnostics, Treatment Development &amp; Surgery</li> </ul> |

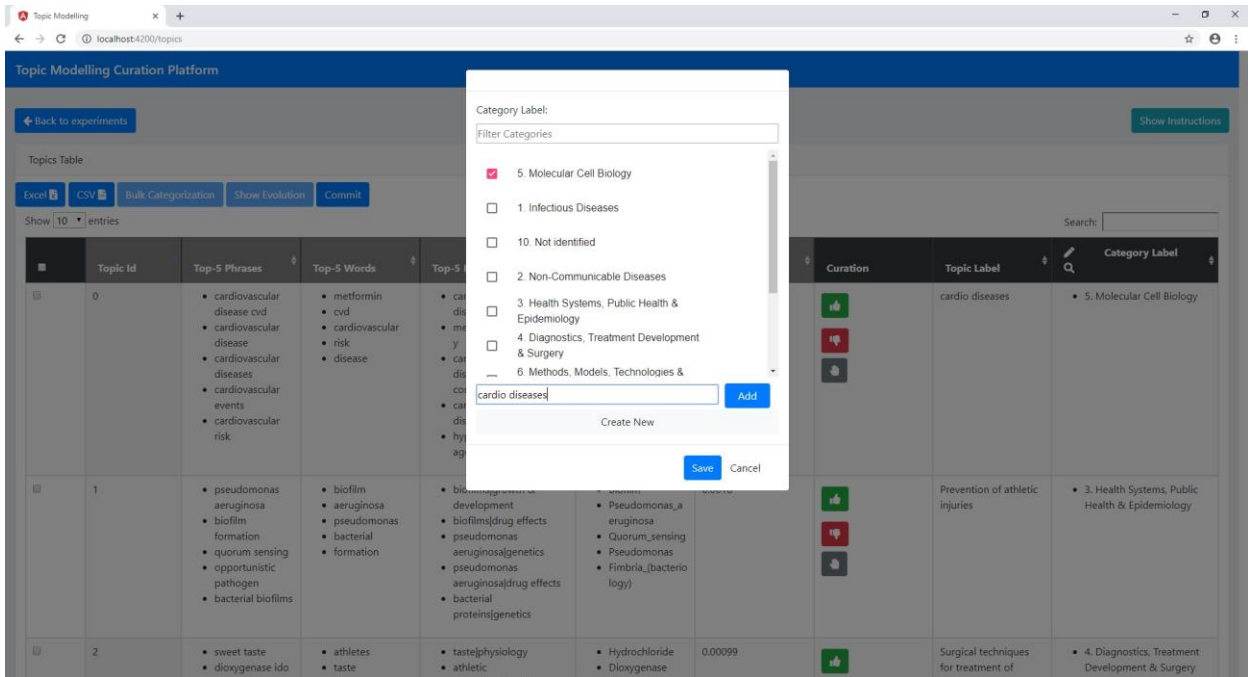
Εικόνα 8: Χαρακτηρισμός topic ως “good”

| Topic Id | Top-5 Phrases   | Top-5 Words   | Top-5 Keywords   | Top-5 MeshTerms  | Topic Strength | Curation   | Topic Label                                    | Category Label  |
|----------|---|---|--|--|----------------|--|--|---|
| 0        | <ul style="list-style-type: none"> <li>cardiovascular disease cvd</li> <li>cardiovascular disease</li> <li>cardiovascular diseases</li> <li>cardiovascular events</li> <li>cardiovascular risk</li> </ul> | <ul style="list-style-type: none"> <li>metformin</li> <li>cvd</li> <li>cardiovascular risk</li> <li>disease</li> </ul>          | <ul style="list-style-type: none"> <li>cardiovascular diseases diagnosis</li> <li>metformin pharmacology</li> <li>cardiovascular diseases prevention &amp; control</li> <li>cardiovascular diseases epidemiology</li> <li>hypoglycemic agents therapeutic use</li> </ul> | <ul style="list-style-type: none"> <li>metformin</li> <li>Cerebrovascular_disease</li> <li>Dipeptidyl_peptidase-4</li> <li>Diabetes_mellitus_type_2</li> <li>Anti-diabetic_medication</li> </ul> | 0.00155        | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | MicroRNAs and gene expression                  | <ul style="list-style-type: none"> <li>5. Molecular Cell Biology</li> </ul>                           |
| 1        | <ul style="list-style-type: none"> <li>pseudomonas aeruginosa</li> <li>biofilm formation</li> <li>quorum sensing</li> <li>opportunistic pathogen</li> <li>bacterial biofilms</li> </ul>                   | <ul style="list-style-type: none"> <li>biofilm</li> <li>aeruginosa</li> <li>pseudomonas</li> <li>bacterial formation</li> </ul> | <ul style="list-style-type: none"> <li>biofilms growth &amp; development</li> <li>biofilms drug effects</li> <li>pseudomonas aeruginosa genetics</li> <li>pseudomonas aeruginosa drug effects</li> <li>bacterial proteins genetics</li> </ul>                            | <ul style="list-style-type: none"> <li>Biofilm</li> <li>Pseudomonas_aeruginosa</li> <li>Quorum_sensing</li> <li>Pseudomonas</li> <li>Fimbria_(bacteriology)</li> </ul>                           | 0.0018         | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | Prevention of athletic injuries                | <ul style="list-style-type: none"> <li>3. Health Systems, Public Health &amp; Epidemiology</li> </ul> |
| 2        | <ul style="list-style-type: none"> <li>sweet taste</li> <li>dioxygenase ido</li> <li>video games</li> <li>video game</li> <li>taste bud</li> </ul>  | <ul style="list-style-type: none"> <li>athletes</li> <li>taste</li> <li>players</li> <li>sports</li> <li>game</li> </ul>        | <ul style="list-style-type: none"> <li>taste physiology</li> <li>athletic injuries epidemiology</li> <li>soccer injuries</li> <li>football injuries</li> <li>taste drug effects</li> </ul>   | <ul style="list-style-type: none"> <li>Hydrochloride</li> <li>Dioxygenase</li> <li>Tryptophan</li> <li>Chemoreceptor</li> <li>Biomechanics</li> </ul>  | 0.00099        | <ul style="list-style-type: none"> <li>👍</li> <li>👎</li> <li>🗑️</li> </ul> | Surgical techniques for treatment of fractures | <ul style="list-style-type: none"> <li>4. Diagnostics, Treatment Development &amp; Surgery</li> </ul> |

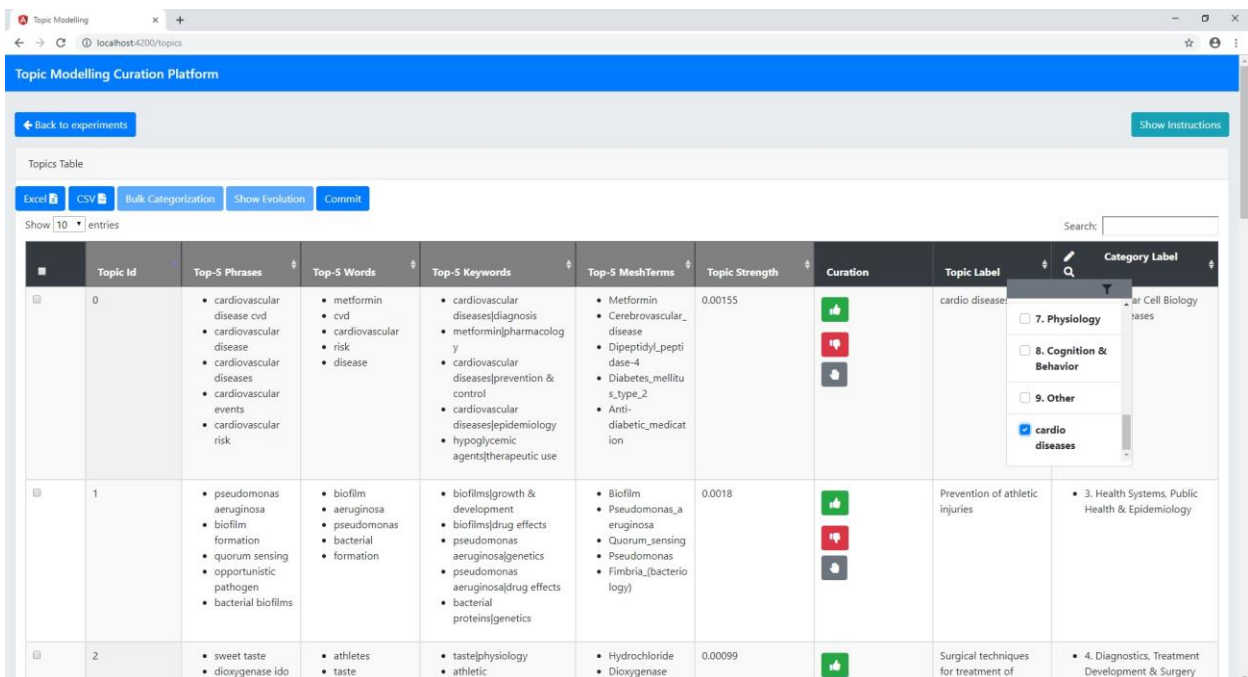
Εικόνα 9: Αναδυόμενο παράθυρο καταχώρησης ετικέτας στο topic



Δύο επιπλέον επιλογές που έχει ο χρήστης όσον αφορά την τελευταία στήλη του πίνακα είναι να φιλτράρει τα δεδομένα του πίνακα βάσει μιας ή περισσότερων κατηγοριών ή ακόμα και να επεξεργαστεί τις υπάρχουσες κατηγορίες στο experiment. Η πρώτη λειτουργικότητα παρέχεται πατώντας πάνω στο εικονίδιο του φακού (δίπλα από το Category Label) και η δεύτερη πατώντας πάνω στο εικονίδιο του μολυβιού. Επιλέγοντας λοιπόν τον φακό, ανοίγει μια dropdown λίστα, όπως φαίνεται στην **Εικόνα 11**, με όλες τις κατηγορίες, όπου ο χρήστης μπορεί να επιλέξει βάσει ποιων κατηγοριών θέλει να φιλτράρει τα δεδομένα του πίνακα. Αφού επιλέξει την/τις κατηγορίες που επιθυμεί να μελετήσει, πατάει το εικονίδιο (χωνί) πάνω δεξιά στο dropdown παράθυρο και όπως φαίνεται στην **Εικόνα 12**, εμφανίζονται μόνο τα topics που ανήκουν στις επιλεγμένες κατηγορίες.

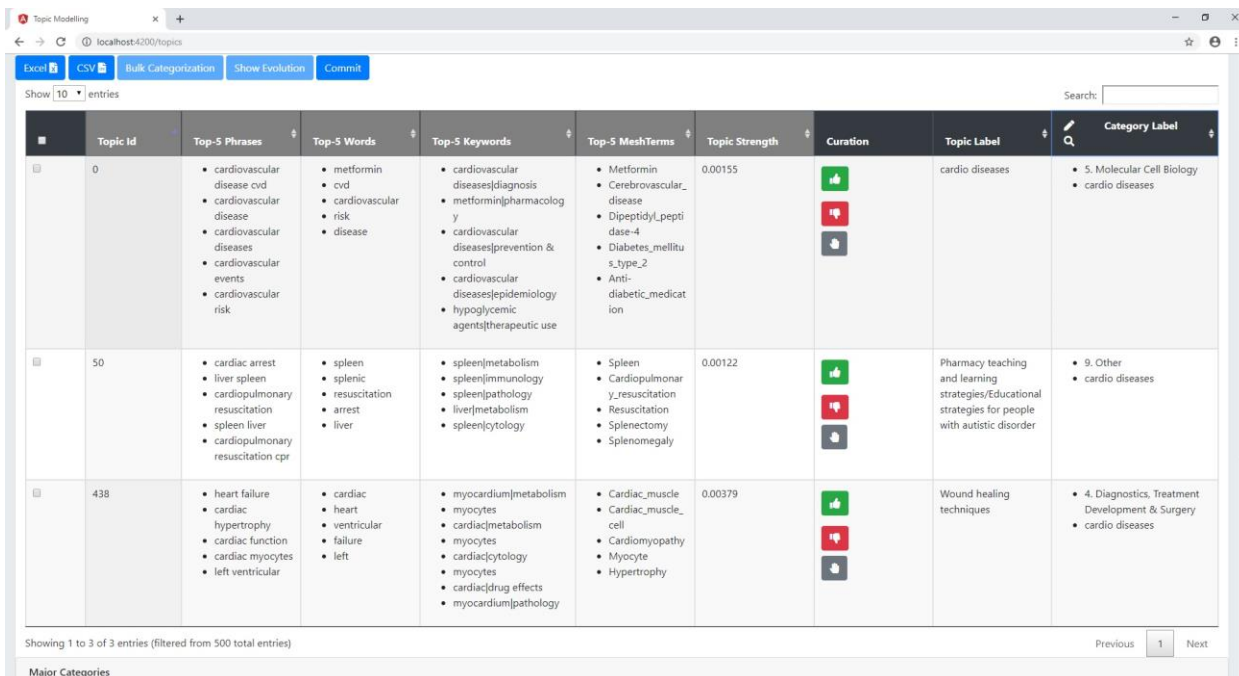


**Εικόνα 10: Αναδυόμενο παράθυρο διαχείρισης κατηγοριών του topic**

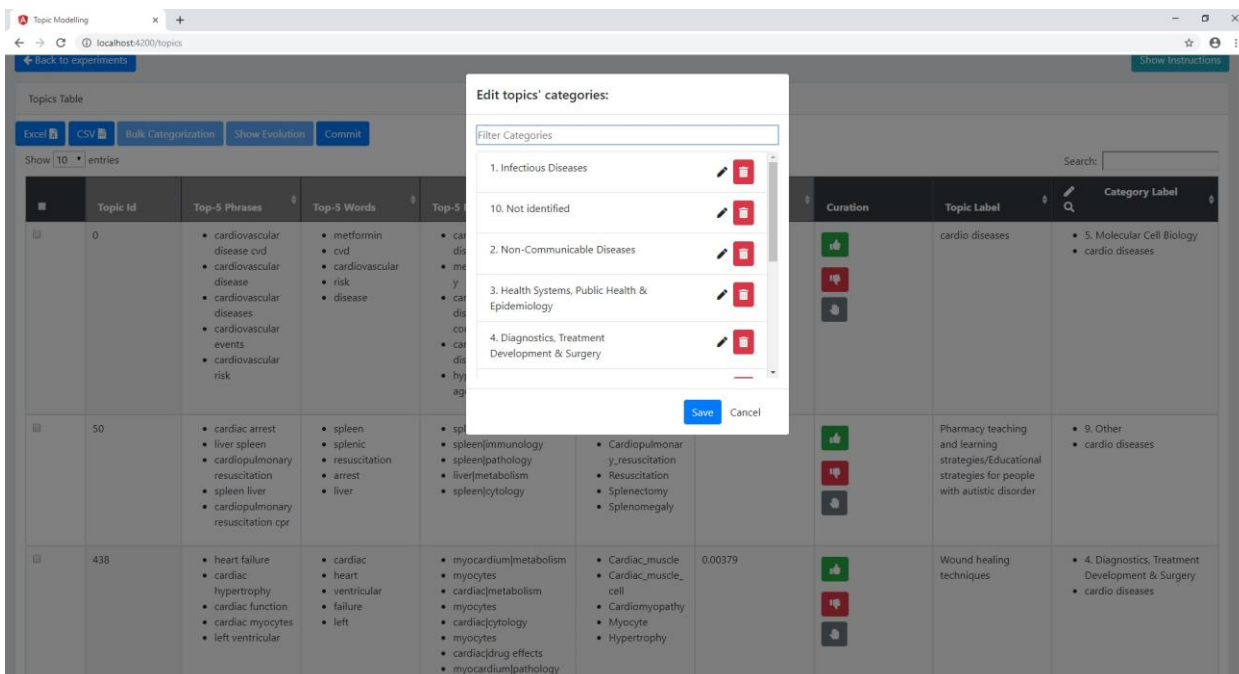


**Εικόνα 11: Φιλτράρισμα των topics βάσει κατηγορίας**

Η δεύτερη λειτουργικότητα όσον αφορά τις κατηγορίες, όπως αναφέρθηκε και προηγουμένως είναι η επεξεργασία τους. Όταν λοιπόν ο χρήστης πατήσει το μολύβι δίπλα από το “Category Label” ανοίγει ένα αναδυόμενο παράθυρο, όπως φαίνεται στην **Εικόνα 13**. Στο συγκεκριμένο παράθυρο υπάρχει μια λίστα με όλες τις κατηγορίες του experiment. Ο curator μπορεί να μετονομάσει ή να διαγράψει οποιαδήποτε κατηγορία από την λίστα. Πατώντας το εικονίδιο του μολυβιού ανοίγει ένα επιπλέον παράθυρο,



**Εικόνα 12: Φιλτράρισμα των topics βάσει κατηγορίας - αποτελέσματα**

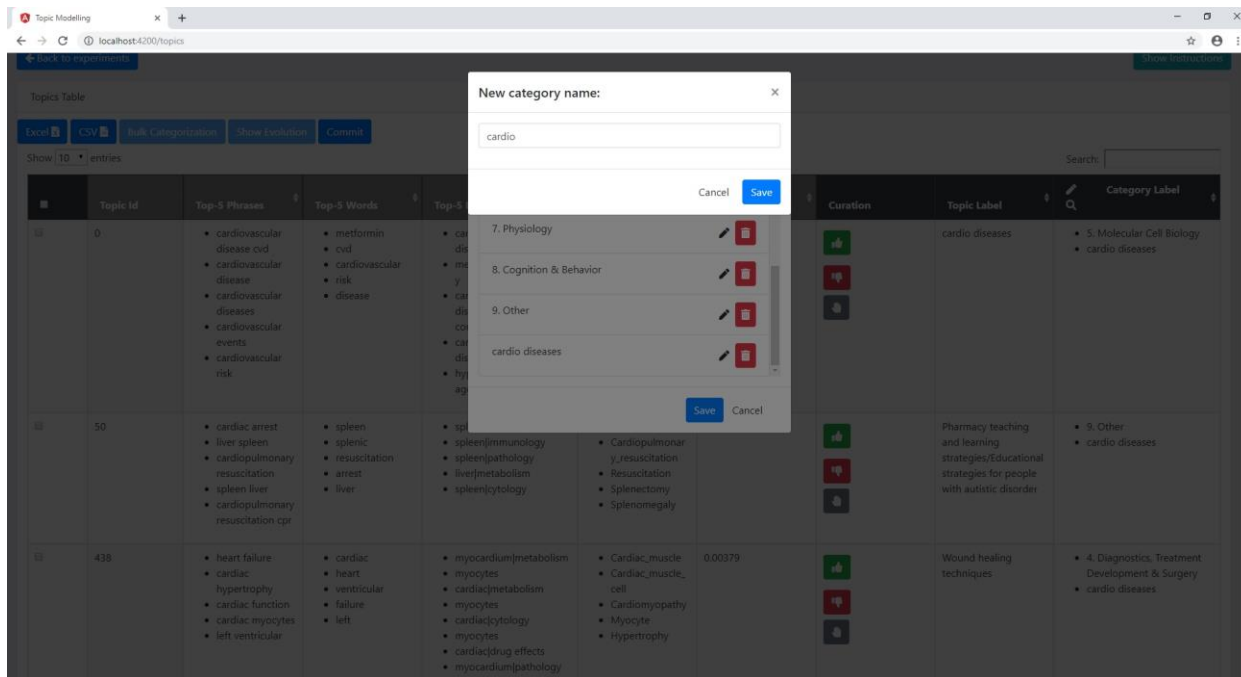


**Εικόνα 13: Επεξεργασία κατηγοριών experiment**

όπως φαίνεται στην **Εικόνα 14**. Σε αυτό το σημείο ο curator καλείται να εισάγει το νέο όνομα της κατηγορίας. Κατά τα γνωστά, μπορεί να σώσει ή να ακυρώσει την αλλαγή του επιλέγοντας το “Save” ή το “Cancel” αντίστοιχα. Πέρα από την μετονομασία όμως των κατηγοριών, ο χρήστης μπορεί να διαγράψει μια κατηγορία πατώντας στο

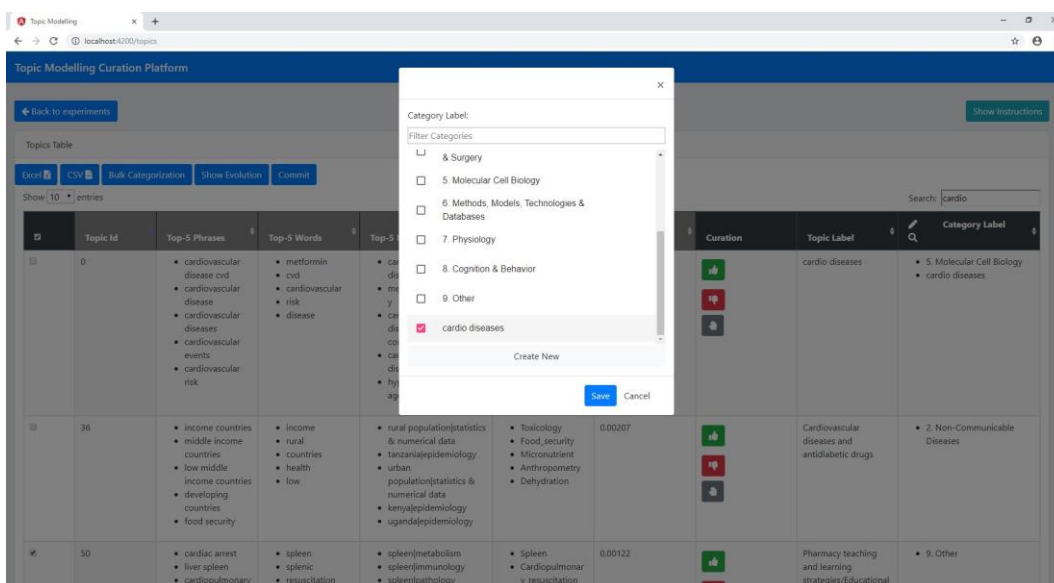
αντίστοιχο κόκκινο κουμπί με τον κάδο. Αν επιλέξει να διαγράψει μια κατηγορία του experiment, η οποία έχει topics που εντάσσονται σε αυτήν, τότε διαγράφοντας την κατηγορία αυτή αφαιρείται από όποιο topic την περιλάμβανε. Τέλος προκειμένου να σώσει τις αλλαγές του πρέπει να πατήσει Save, αλλιώς όλες οι αλλαγές θα χαθούν.

Συνεχίζοντας την παρουσίαση της οθόνης των topics, βλέπουμε ότι ο χρήστης έχει τη δυνατότητα σε αυτόν τον πίνακα όπως και στον προηγούμενο, να κατεβάσει τα δεδομένα του σε Excel ή CSV αρχείο, να επιλέξει πόσες γραμμές θα φαίνονται ανά σελίδα του πίνακα, όπως επίσης και να πραγματοποιήσει αναζήτηση πάνω σε όλο τον πίνακα βάσει συγκεκριμένου keyword.



Εικόνα 14: Μετονομασία κατηγορίας

Μια επιπλέον δυνατότητα που έχει ο curator στη συγκεκριμένη οθόνη είναι η μαζική κατηγοριοποίηση (Bulk Categorization). Επιλέγοντας δύο ή περισσότερα topic με κοινό περιεχόμενο μπορεί να τους προσθέσει μία ή περισσότερες κοινές κατηγορίες όπως φαίνεται στην **Εικόνα 15**.

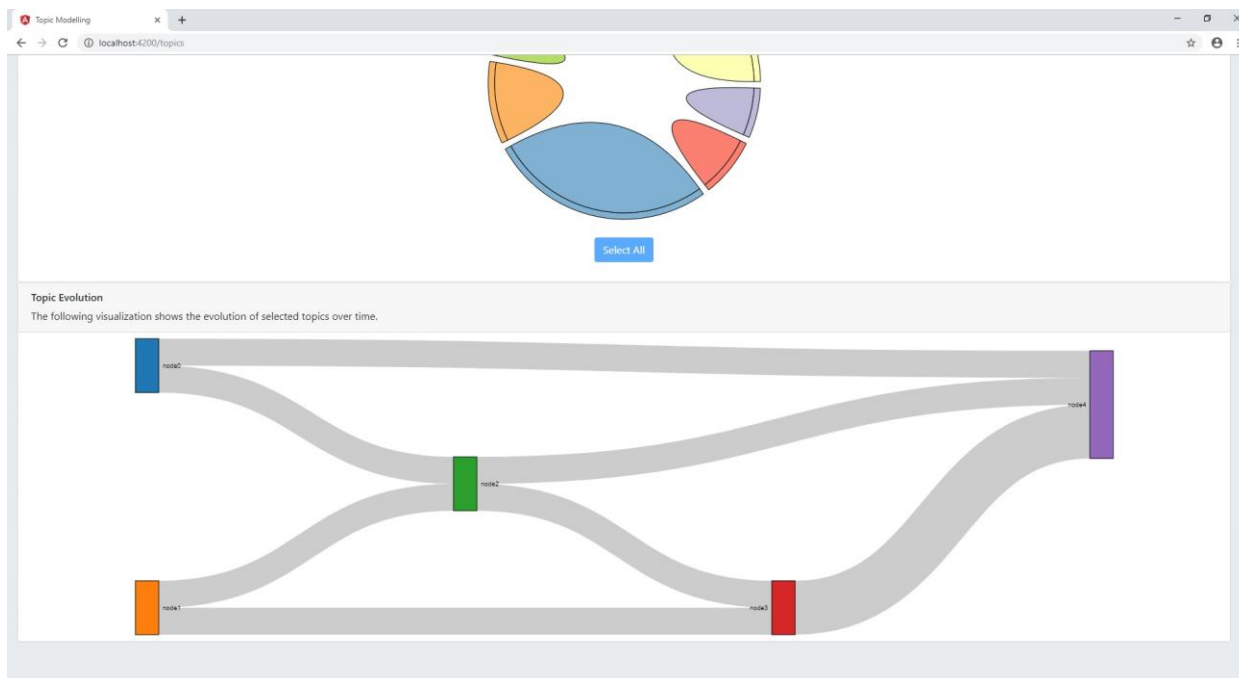


Εικόνα 15: Μαζική κατηγοριοποίηση των topics



Όπως και στην επεξεργασία των κατηγοριών ενός topic (Εικόνα 10) έτσι και εδώ, ο χρήστης μπορεί πέρα από το να καταχωρήσει κάποια από τις υπάρχουσες κατηγορίες να δημιουργήσει κάποια καινούρια, πατώντας το “Create New”.

Δίπλα από το “Bulk Categorization” υπάρχει ένα άλλο κουμπί που γράφει “Show Evolution”. Επιλέγοντας κάποια topics ο χρήστης μπορεί να πατήσει πάνω στο συγκεκριμένο κουμπί και θα εμφανιστεί ένα γράφημα όπως φαίνεται στην **Εικόνα 16** (το δεύτερο γράφημα).

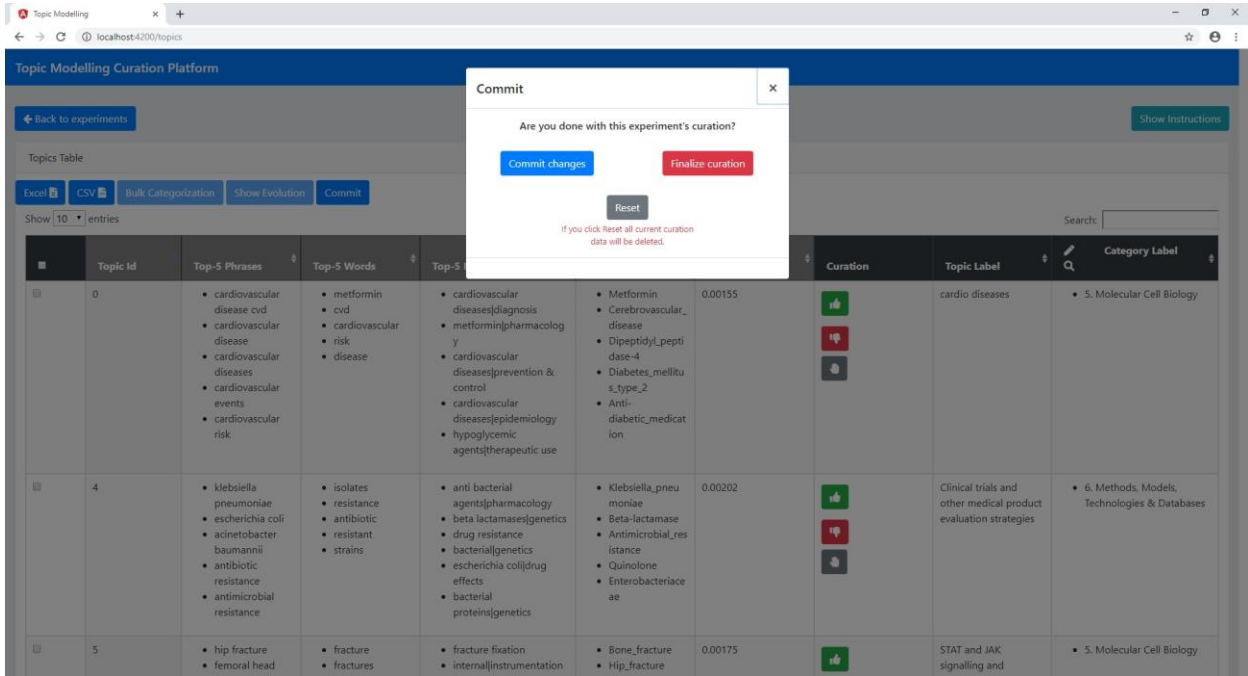


**Εικόνα 16: Διάγραμμα εξέλιξης στον χρόνο των topic**

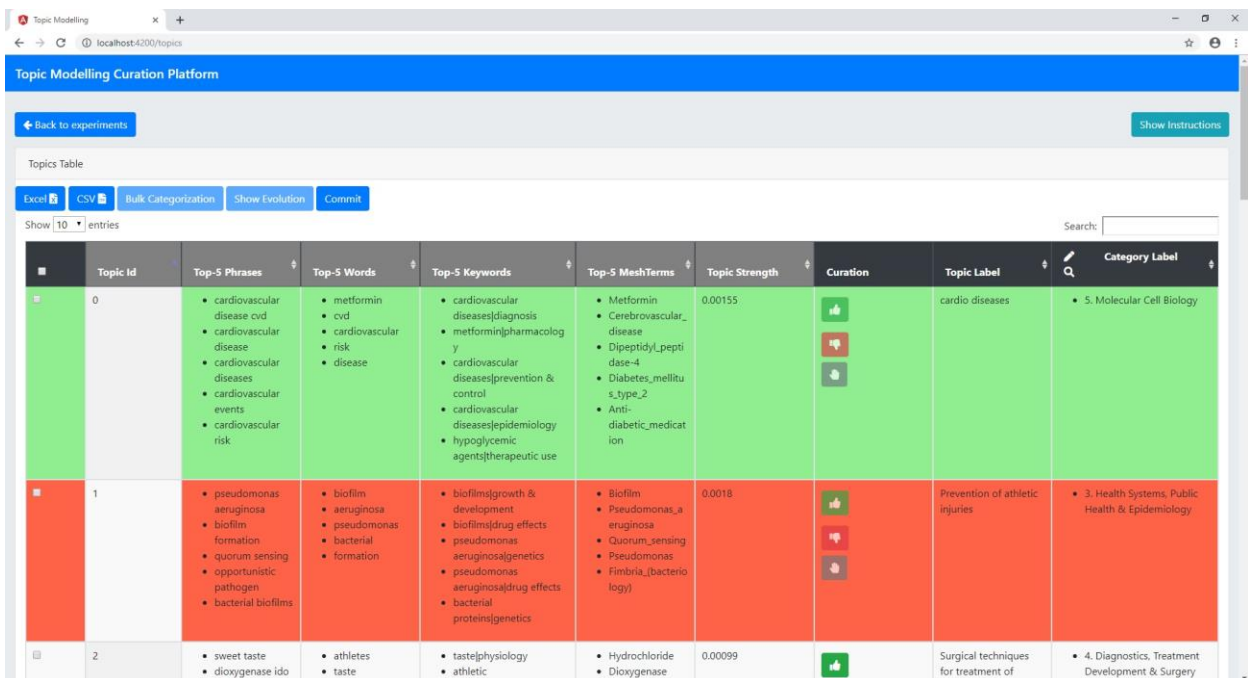
Το συγκεκριμένο γράφημα δείχνει την εξέλιξη των topic στον χρόνο και το κατά πόσο αυξήθηκε, μειώθηκε ή παρέμεινε σταθερή η δυναμικότητα ενός topic στην πάροδο του χρόνου. Η δυναμικότητα ενός topic εκφράζεται από το πάχος της γραμμής. Όσο πιο χοντρή είναι η γραμμή, τόσο πιο σημαντικό είναι το topic στη συγκεκριμένη περίοδο. Το γράφημα αυτό λοιπόν προσφέρεται για συγκρίσεις μεταξύ των topic και για συμπεράσματα για το ποια topic είναι στην επικαιρότητα ανά πάσα χρονική περίοδο.

Μια επιπρόσθετη λειτουργικότητα που παρέχεται μέσω της πλατφόρμας μας είναι το “Commit”. Καθώς ο curator επεξεργάζεται και αξιολογεί τα topics ενός experiment, εμείς κρατάμε τις όποιες αλλαγές του μέσω ενός συστήματος auto save που έχουμε δημιουργήσει. Έτσι μπορεί να σταματήσει ανά πάσα στιγμή το curation και να ξαναρχίσει αργότερα από εκεί ακριβώς που το είχε αφήσει. Πότε όμως αυτές οι τοπικές αλλαγές αποθηκεύονται στη βάση δεδομένων μας; Και πότε ολοκληρώνεται το curation ενός experiment; Η απάντηση δίνεται όταν πατήσουμε το κουμπί του “Commit”. Όπως φαίνεται στην **Εικόνα 17**, ανοίγει ένα αναδυόμενο παράθυρο όπου ο χρήστης έχει τρεις επιλογές: Commit changes, Finalize curation και Reset. Επιλέγοντας το “Commit changes” ο curator δηλώνει ότι θέλει να αποθηκεύσει όλες τις τοπικές αλλαγές στη βάση δεδομένων. Σε αυτό το σημείο γίνεται η επικοινωνία με τη βάση και όλες οι αλλαγές καταγράφονται σε αυτήν. Με το πέρας αυτής της διαδικασίας ο χρήστης δεν μπορεί πλέον να επεξεργαστεί ξανά τα συγκεκριμένα δεδομένα που αποθήκευσε, όπως φαίνεται στην **Εικόνα 18**. Σε αυτήν την εικόνα ο curator επέλεξε να αποθηκεύσει το curation των topic με id 0 και 1 και συνεπώς δεν μπορεί να επεξεργαστεί πλέον τα συγκεκριμένα topic, όπως φαίνεται και από τα απενεργοποιημένα κουμπιά στη στήλη curation. Η δεύτερη επιλογή που δίνεται στον curator στο αναδυόμενο παράθυρο του commit είναι το “Finalize curation”. Πατώντας το συγκεκριμένο κουμπί, ο χρήστης

δηλώνει ότι έχει τελειώσει με το curation του experiment. Οι τελευταίες αλλαγές αποθηκεύονται στη βάση δεδομένων και καμία επιπλέον αλλαγή δεν μπορεί να γίνει έκτοτε στο experiment. Τι γίνεται όμως αν έχουν πάει όλα στραβά και θέλουμε να αρχίσουμε το curation από την αρχή; Σε αυτήν την περίπτωση μπορούμε να επιλέξουμε το “Reset”, να σβήσουμε έτσι το έως εκείνη τη στιγμή curation και να ξεκινήσουμε από την αρχή.



Εικόνα 17: Αναδυόμενο παράθυρο Commit



Εικόνα 18: Παράδειγμα commit τις τρέχουσες αλλαγές

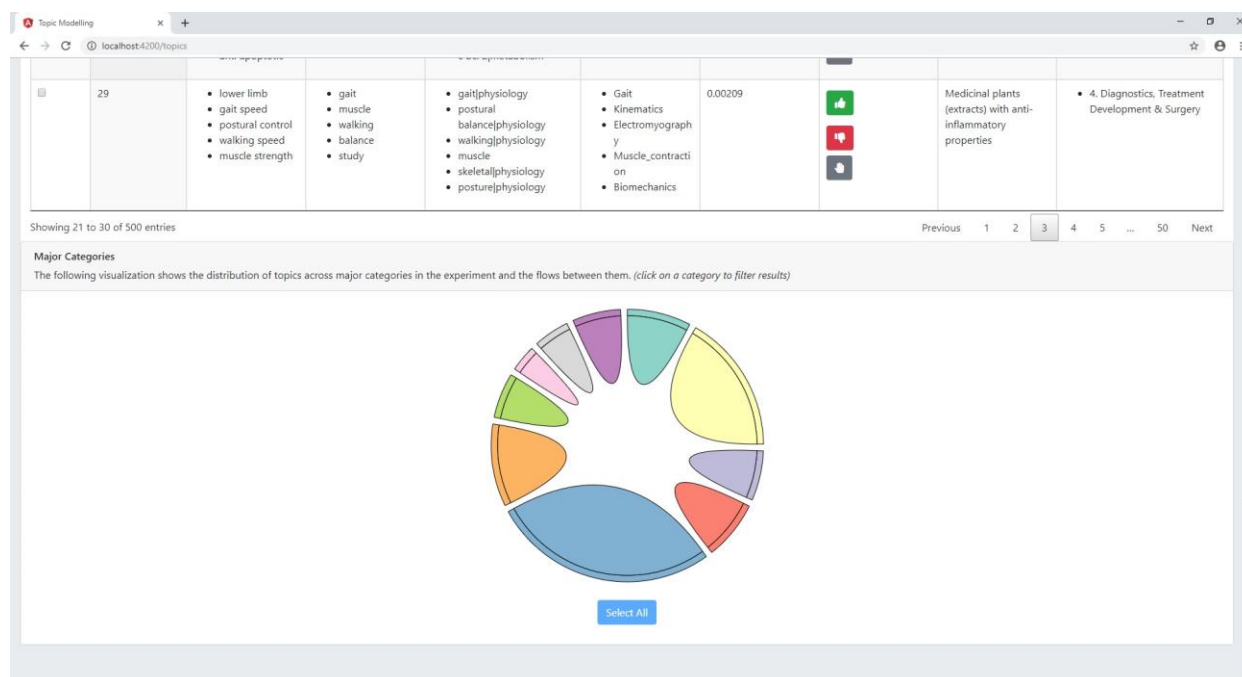
Συνεχίζοντας την περιήγησή μας και πηγαίνοντας πιο κάτω στην οθόνη παρατηρούμε το μενού σελιδοποίησης που παρέχεται στο κάτω δεξιό μέρος του πίνακα, όπως

φαίνεται στην **Εικόνα 19** και μέσω του οποίου ο χρήστης μπορεί να περιηγηθεί στις διάφορες σελίδες του πίνακα.

| ID | Topic  | Related Topics  | Category  |
|----|--|---|---|
| 27 | <ul style="list-style-type: none"> <li>skin lesions</li> <li>cutaneous melanoma</li> <li>human skin</li> <li>epidermal keratinocytes</li> <li>hair follicle</li> </ul> | <ul style="list-style-type: none"> <li>skin</li> <li>melanoma</li> <li>topical</li> <li>keratinocytes</li> <li>cutaneous</li> </ul> | <ul style="list-style-type: none"> <li>keratinocytes cytology</li> <li>melanoma genetics</li> <li>skin metabolism</li> <li>keratinocytes metabolism</li> <li>melanoma drug therapy</li> </ul>                         |
| 28 | <ul style="list-style-type: none"> <li>induced apoptosis</li> <li>cell death</li> <li>cell apoptosis</li> <li>pro apoptotic</li> <li>anti apoptotic</li> </ul>         | <ul style="list-style-type: none"> <li>apoptosis</li> <li>caspase</li> <li>death</li> <li>cells</li> <li>cell</li> </ul>            | <ul style="list-style-type: none"> <li>caspase 3 metabolism</li> <li>apoptosis drug effects</li> <li>caspases metabolism</li> <li>apoptosis physiology</li> <li>proto oncogene proteins c bcl 2 metabolism</li> </ul> |
| 29 | <ul style="list-style-type: none"> <li>lower limb</li> <li>gait speed</li> <li>postural control</li> <li>walking speed</li> <li>muscle strength</li> </ul>             | <ul style="list-style-type: none"> <li>gait</li> <li>muscle</li> <li>walking</li> <li>balance</li> <li>study</li> </ul>             | <ul style="list-style-type: none"> <li>gait physiology</li> <li>postural balance physiology</li> <li>walking physiology</li> <li>muscle</li> <li>skeletal physiology</li> <li>posture physiology</li> </ul>           |

**Εικόνα 19: Σελιδοποίηση πίνακα δεδομένων**

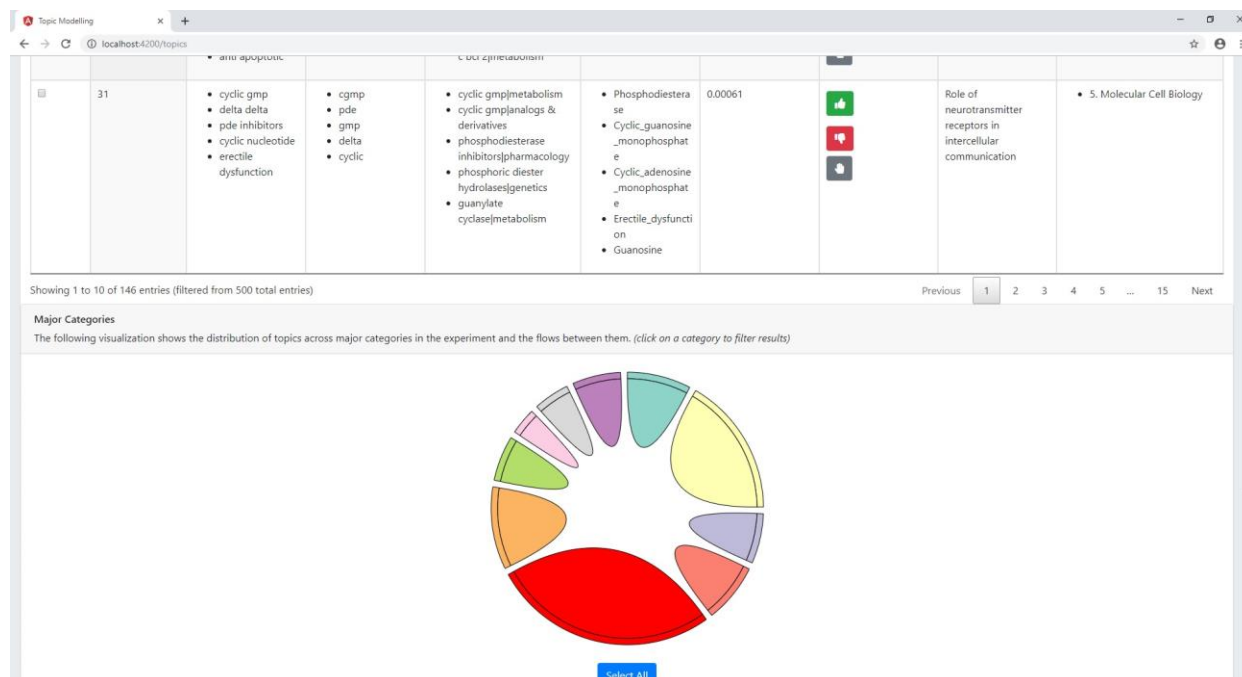
Πηγαίνοντας λίγο πιο κάτω, παρατηρούμε ένα από τα γραφήματα που υπάρχουν στο εργαλείο μας. (**Εικόνα 20**) Το συγκεκριμένο γράφημα χρησιμοποιείται για να δείξει την κατανομή των topics στις διάφορες κατηγορίες του experiment. Μέσα από το συγκεκριμένο γράφημα ο χρήστης μπορεί να δει ποια κατηγορία διαθέτει το μεγαλύτερο και ποια το μικρότερο μερίδιο, περνώντας το ποντίκι πάνω από το κομμάτι του γραφήματος με την μεγαλύτερη και τη μικρότερη έκταση αντίστοιχα.



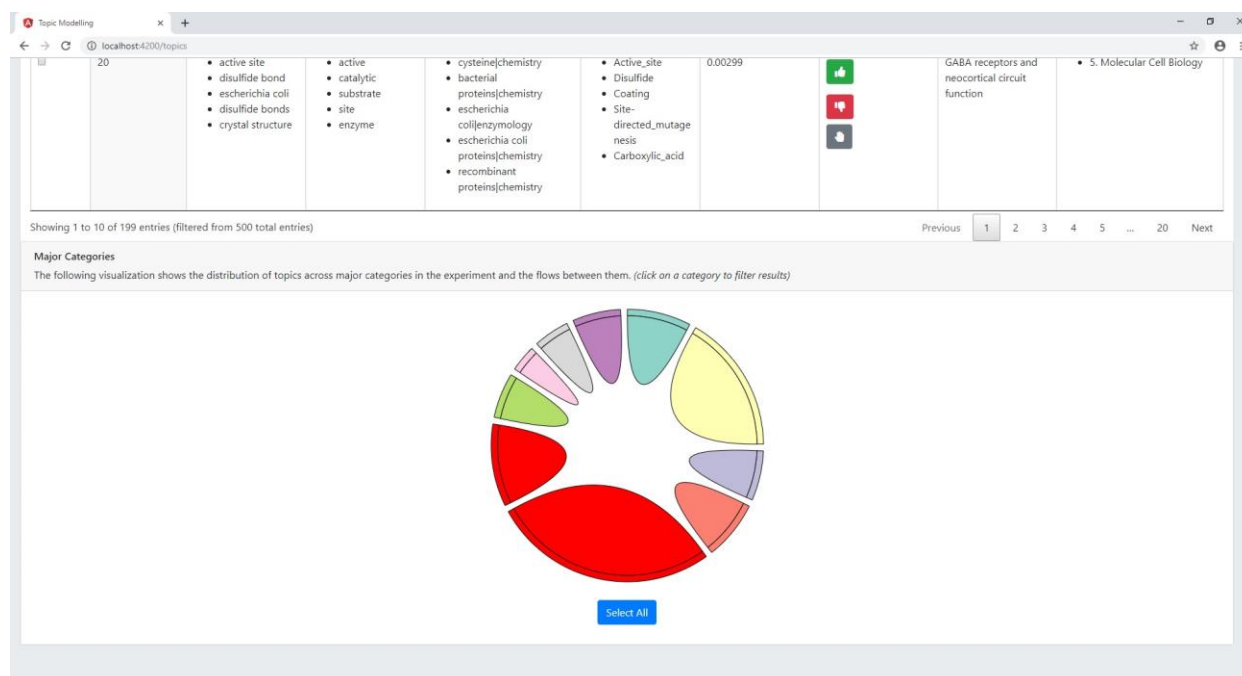
**Εικόνα 20: Γράφημα κατανομής των topics στις κατηγορίες του experiment**

Πατώντας πάνω σε μια συγκεκριμένη κατηγορία μπορεί να φιλτράρει στον από πάνω πίνακα τα topics που ανήκουν μόνο στη συγκεκριμένη κατηγορία, όπως φαίνεται στην **Εικόνα 21**, όπου έχουμε επιλέξει την κατηγορία “5.Molecular Cell Biology” (κόκκινη

περιοχή στο γράφημα) και βλέπουμε στον πίνακα από πάνω ότι έχει φιλτράρει μόνο τα topics της συγκεκριμένης κατηγορίας. Ο χρήστης μπορεί να επιλέξει στο γράφημα περισσότερες από μια κατηγορίες και να φιλτράρει τα δεδομένα του πίνακα βάσει αυτών των κατηγοριών, όπως φαίνεται στην **Εικόνα 22**.



**Εικόνα 21: Επιλογή μιας κατηγορίας από το γράφημα**



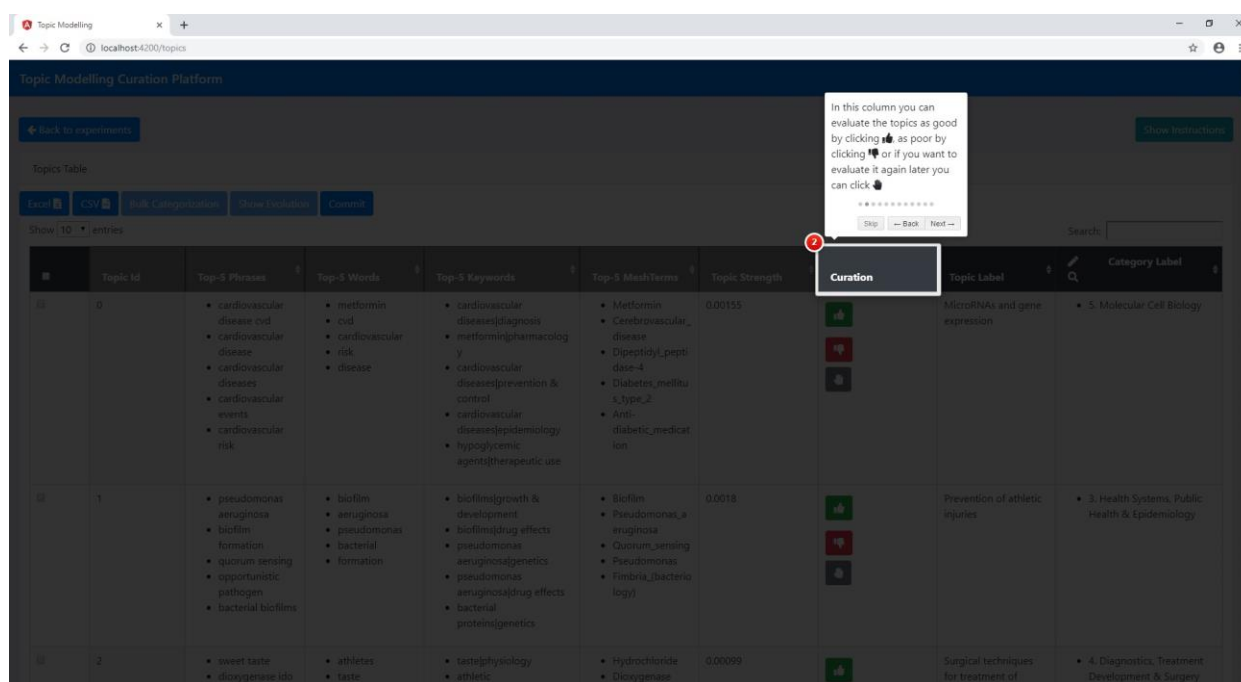
**Εικόνα 22: Επιλογή περισσότερων κατηγοριών από το γράφημα**

Πατώντας πάνω σε μια ήδη επιλεγμένη (κόκκινη) περιοχή, ο χρήστης μπορεί να την απεπιλέξει. Επίσης αξίζει να τονίσουμε ότι η επιλογή κατηγοριών για το φιλτράρισμα δεδομένων συνδέεται με το φιλτράρισμα μέσω του “μεγεθυντικού φακού” που αναφέραμε πιο πάνω. Συνεπώς επιλέγοντας μια κατηγορία στο γράφημα, τσεκάρει τη

συγκεκριμένη κατηγορία στην drop down λίστα, και αντιστρόφως, επιλέγοντας μια κατηγορία από την λίστα, κοκκινίζει τη συγκεκριμένη περιοχή στο γράφημα.

Τέλος, πρέπει να επισημάνουμε ότι το συγκεκριμένο γράφημα προτιμήθηκε έναντι ενός απλού pie chart, διότι πέρα από την κατανομή των topics στις κατηγορίες παρέχει πληροφορία και για το flow και τη σύνδεση μεταξύ των κατηγοριών. Πιο συγκεκριμένα, δεδομένου ότι όπως είδαμε προηγουμένως ένα topic μπορεί να ανήκει σε περισσότερες από μία κατηγορίες θέλαμε κάπως να εκφράσουμε αυτά τα κοινά topics μεταξύ των κατηγοριών. Μέσω του συγκεκριμένου γραφήματος λοιπόν, εκφράζεται και το ποσοστό των topics αυτών που είναι κοινά μεταξύ των κατηγοριών.

Κλείνοντας την ανάλυση της παρούσας οθόνης, πρέπει να αναφέρουμε ότι παρέχεται λεπτομερής καθοδήγηση και βοήθεια για τον χρήστη όσον αφορά τη λειτουργικότητα μέσα στο εργαλείο. Μπορεί λοιπόν, ανά πάσα στιγμή να πατήσει το κουμπί “Show Instructions” πάνω δεξιά, και μέσω μιας αλληλουχίας οδηγιών να ξεναγηθεί και να πληροφορηθεί για το τι μπορεί να κάνει στην κάθε οθόνη του εργαλείου, όπως φαίνεται στην **Εικόνα 23**.



**Εικόνα 23: Οδηγίες για τη χρήση του εργαλείου**

Όπως μπορεί να παρατηρήσει κανείς στη συγκεκριμένη οθόνη επιλέξαμε να παρουσιάσουμε συνοπτική πληροφορία όσον αφορά το topic. Για παράδειγμα επιλέξαμε να δείχνουμε μόνο τα top-5 keywords, phrases κλπ. και όχι περισσότερα, όπως επίσης να μην παρουσιάζουμε καθόλου πληροφορία ως προς τα documents που αναφέρονται σε κάθε topic, τους authors και πολλά άλλα. Ο λόγος ήταν ότι δημιουργήσαμε το συγκεκριμένο εργαλείο για να βοηθήσουμε τους curators να υλοποιούν πιο γρήγορα το curation μιας συλλογής από topics. Συνεπώς αυτό που θέλει ο curator σε αυτό το σημείο είναι τόση πληροφορία όση χρειάζεται προκειμένου να έχει μια ολοκληρωμένη εικόνα με μια γρήγορη ματιά στο topic. Περισσότερη πληροφορία θα τον καθυστερήσει. Επομένως επιλέξαμε σε αυτό το σημείο να έχουμε συνοπτική πληροφορία και αν ο χρήστης χρειάζεται περισσότερες πληροφορίες για την αξιολόγηση του topic, μπορεί να πατήσει πάνω σε αυτό (στον πίνακα) και να μεταφερθεί στην επόμενη οθόνη που έχει περισσότερη πληροφορία για το topic.



### 3.2.3 Οθόνη των Documents

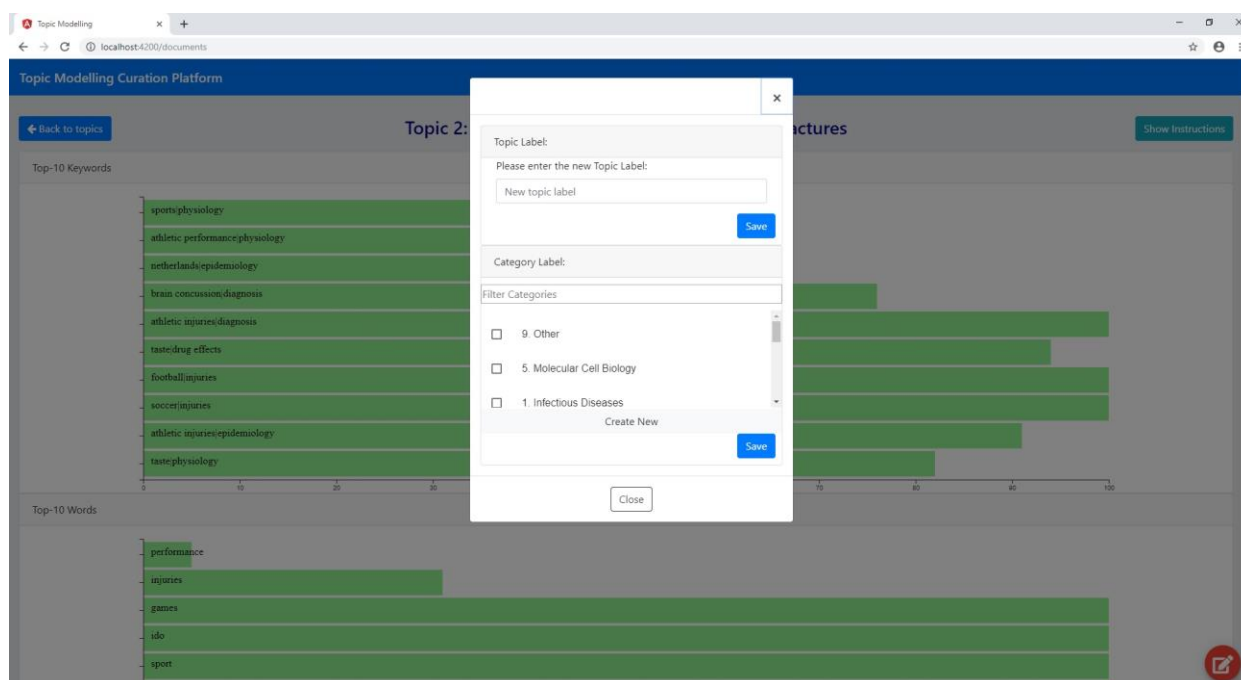
Επιλέγοντας λοιπόν και πατώντας σε ένα συγκεκριμένο topic, ο χρήστης μεταφέρεται στην επόμενη οθόνη, όπως φαίνεται στην **Εικόνα 24**. Στην οθόνη αυτήν, μπορεί να συλλέξει περισσότερη πληροφορία για το topic πριν προχωρήσει στην αξιολόγησή του. Μπαίνοντας λοιπόν στη συγκεκριμένη σελίδα, παρατηρούμε ότι στο επάνω μέρος υπάρχουν τα κουμπιά της πλοήγησης προς τα πίσω (Back to topics) και της βοήθειας (Show Instructions) όπως και στην προηγούμενη. Επιπλέον μεταξύ των δύο αυτών κουμπιών υπάρχει ο τίτλος του επιλεγμένου topic.



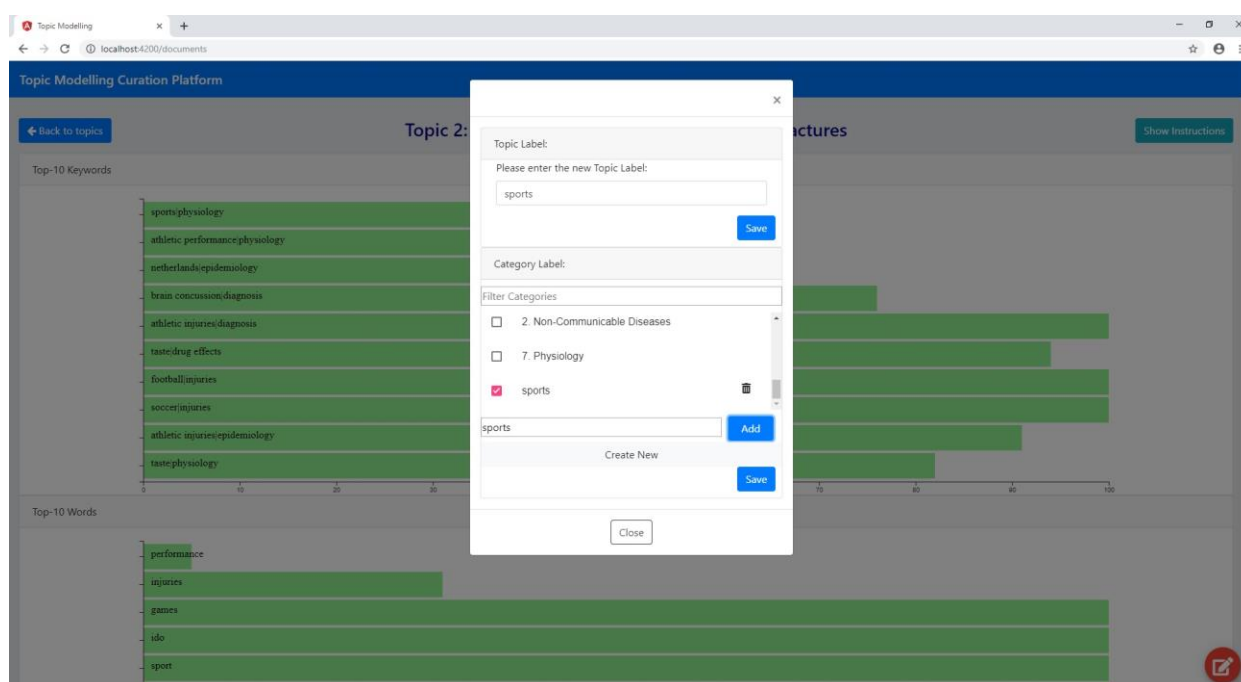
**Εικόνα 24: Οθόνη των documents**

Κάτω δεξιά (μόνιμα τοποθετημένο στην ίδια θέση) υπάρχει ένα κόκκινο κουμπί που αφορά το curation. Δεδομένου ότι το εργαλείο απευθύνεται κυρίως σε curators για την υποστήριξη της διαδικασίας του curation, θέλαμε ο χρήστης να μπορεί να αξιολογήσει το topic και από τη συγκεκριμένη οθόνη, χωρίς να χρειάζεται να πάει πίσω και να το ψάχνει στον πίνακα με τα topics για να το κάνει από εκεί. Πατώντας λοιπόν στο συγκεκριμένο κουμπί ανοίγει ένα αναδυόμενο παράθυρο, όπως φαίνεται στην **Εικόνα 25**. Στο συγκεκριμένο παράθυρο ο χρήστης έχει τη δυνατότητα να εισάγει ετικέτα για το topic και να το κατηγοριοποιήσει σε μία ή περισσότερες κατηγορίες, όπως ακριβώς παρουσιάστηκαν αυτές οι λειτουργικότητες και προηγουμένως στις Εικόνες 9 και 10. Αφού επεξεργαστεί ότι επιθυμεί ο χρήστης, θα πρέπει να πατήσει το “Save” για να αποθηκεύσει τη συγκεκριμένη αλλαγή πριν κλείσει το παράθυρο. Έτσι λοιπόν αν αλλάξει την ετικέτα του topic, θα πρέπει να πατήσει το πάνω “Save”, ενώ αν επεξεργαστεί τις κατηγορίες του topic θα πρέπει να πατήσει το κάτω “Save” προκειμένου να αποθηκεύσει τις αλλαγές του. Έπειτα θα πρέπει να πατήσει το “Close” ώστε να κλείσει το παράθυρο και οι αλλαγές του θα έχουν αποθηκευτεί. Ένα παράδειγμα φαίνεται στην **Εικόνα 26**, όπου έχουμε αλλάξει τόσο την ετικέτα όσο και τις κατηγορίες του topic. Στην περίπτωση αυτή θα πρέπει να πατήσουμε τόσο το πάνω Save, όσο και το κάτω, για να αποθηκεύσουμε τις αλλαγές μας και εν συνεχεία θα πρέπει να πατήσουμε Close για να κλείσουμε το παράθυρο.

Συνεχίζοντας με την ανάλυση της οθόνης, κάτω από τον τίτλο έχουμε barcharts για τα top-10 keywords, words κλπ. του topic. Όπως είπαμε και πιο πάνω, σε αυτήν την οθόνη ο χρήστης μπορεί να συλλέξει περισσότερη πληροφορία για το topic, συνεπώς



Εικόνα 25: Αναδυόμενο παράθυρο για curation του topic



Εικόνα 26: Αναδυόμενο παράθυρο για curation του topic - αλλαγές

παρατηρούμε ότι ενώ στην προηγούμενη οθόνη είχαμε μόνο τα top-5 από κάθε κατηγορία, τώρα έχουμε τα top-10, όπως επίσης και την συχνότητα εμφάνισής τους στα διάφορα documents, όπως φαίνεται στην Εικόνα 24.

Κάτω από τα barcharts για τα keywords, τα words, τα phrases και τα mesh terms έχουμε το βασικό δομικό στοιχείο αυτής της οθόνης που είναι ο πίνακας με τα documents αναφέρονται στο topic που έχουμε επιλέξει να μελετήσουμε. Κάθε γραμμή του πίνακα είναι ένα document. Όπως μπορούμε να δούμε και στην **Εικόνα 27** ο πίνακας αποτελείται από οχτώ στήλες. Η πρώτη στήλη αναφέρει τον τίτλο του document, η δεύτερη τον τύπο του (π.χ P για publication), η τρίτη την ημερομηνία δημοσίευσής τους, η τέταρτη την εφημερίδα στην οποία δημοσιεύθηκε, η πέμπτη τους

authors, η έκτη την περίληψή του, η έβδομη τα link που οδηγούν στις αυθεντικές δημοσιοποιήσεις και η όγδοη τα 10 πιο αντιπροσωπευτικά keywords που ταιριάζουν το document με το επιλεγμένο topic.

| Title  | Type | Publication Date | Journal (View All)   | Authors  | Abstract   | URL   | Top-10 Keywords   |
|--|------|------------------|--|--|--|---|---|
| Randomized controlled trial to compare oral analge...  | P    | 2018-02-01       | Journal of Dental Anesthesia and Pain Medicine   | Goswami, Devalina & Jain, Gaurya & Mohod, Mangesh ...  | Background Third molar extraction is associated with considerable pain and discomfort, which is most ... | <a href="https://www.doi.org/10.17245/jdaptm.2018.18.1.15">https://www.doi.org/10.17245/jdaptm.2018.18.1.15</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5858009">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5858009</a> |   |
| Head Injury and Amyotrophic Lateral Sclerosis          | P    | 2007-07-19       |  | Chen, Honglei & Richard, Marie & Sandler, Dale P. ...  | Recent data showed that soccer players in Italy had an unusually high risk of amyotrophic lateral sc ... | <a href="https://www.doi.org/10.1093/aje/kwm153">https://www.doi.org/10.1093/aje/kwm153</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342</a>                     | New EnglandJepidemiology,SoccerInjuries,CranioCerebral Trauma complications,Amyotrophic Lateral Scl ... |
| Direct reciprocity with costly punishment: generou ... | P    | 2008-10-02       |  | Rand, David G. & Ohtsuki, Hisashi & Nowak, Martin ...  | The standard model for direct reciprocity is the repeated Prisoner's Dilemma, where in each round pl ... | <a href="https://www.doi.org/10.1016/j.tbi.2008.09.015">https://www.doi.org/10.1016/j.tbi.2008.09.015</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2614626">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2614626</a>       |   |
| Relationship between Lower Extremity Tightness and ... | P    | 2014-05-01       | Journal of physical therapy science  | Endo, Yasuhiro & Sakamoto, Masaaki                     | [Purpose] The purpose of this study was to examine the relationship between lower extremity tightnes ... | <a href="https://www.doi.org/10.1589/jpts.26.661">https://www.doi.org/10.1589/jpts.26.661</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4047227">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4047227</a>                   |   |
| Effects of taste stimulation on gastric myoelectri ... | P    | 2018-01-01       | Saudi Journal of Gastroenterology : Official Journal of the Saudi Gastroenterology Association | Waluga, Marek & Jonderko, Krzysztof & Domoslawska, ... | Background/Aim: Sham feeding, reproducing the cephalic phase of digestion, and involving combined vi ... | <a href="https://www.doi.org/10.4103/sjg.SJG_419_17">https://www.doi.org/10.4103/sjg.SJG_419_17</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900469">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900469</a>             | Autonomic Nervous System physiology,Electromyography methods  |
| Relationship between Adaptation after Returning to ... | P    | 2014-11-01       | Journal of physical therapy science  | Tatsumi, Tomonori                                      | [Purpose] The purpose of this study was to examine the relationship between adaptation after returni ... | <a href="https://www.doi.org/10.1589/jpts.26.1813">https://www.doi.org/10.1589/jpts.26.1813</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4242963">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4242963</a>                 |   |
| Anthropometric and                                     | P    | 2011-12-01       | Journal of Human   | Koroponovski, Nenad &                                  | Karate tournaments consist of two equally important  | <a href="https://www.doi.org/10.2478/v">https://www.doi.org/10.2478/v</a>   |   |

Εικόνα 27: Πίνακας των documents

Όπως φαίνεται και στην Εικόνα 27 η πληροφορία κόβεται και υπάρχουν αποσιωπητικά. Ο λόγος είναι ότι προτιμήσαμε να παρουσιάσουμε λιγότερη πληροφορία για να είναι πιο ευανάγνωστος ο πίνακας. Για να διαβάσουμε λοιπόν αναλυτικά όλες τις πληροφορίες για το κάθε document αρκεί να πατήσουμε πάνω σε οποιοδήποτε κελί του document και θα μας ανοίξει ένα παράθυρο όπως φαίνεται στην Εικόνα 28.

| Title   | Type | Publication Date | Journal | Authors   | Abstract  | URL   | Top-10 Keywords  |
|---|------|------------------|---------|---|---|---|--|
| Head Injury and Amyotrophic Lateral Sclerosis | P    | 2007-07-19       |         | Chen, Honglei & Richard, Marie & Sandler, Dale P. & Umbach, David M. & Kamel, Freya (2007), <a href="https://www.doi.org/10.1093/aje/kwm153">https://www.doi.org/10.1093/aje/kwm153</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342</a> | Recent data showed that soccer players in Italy had an unusually high risk of amyotrophic lateral sclerosis (ALS) and repeated head trauma might have contributed to this increase. The authors examined whether head injury was related to ALS risk in a case-control study of 109 ALS cases and 255 controls and conducted a meta-analysis of published literature. Overall ever having a head injury was non-significantly associated with a higher ALS risk. When compared with individuals without a head injury, a statistically significant ALS risk elevation was found for participants with more than one head injury (odds ratio (OR): 3.1, 95 percent confidence intervals (CI): 1.2, 8.1) or with head injury during the past 10 years (OR=3.2, 95 percent CI: 1.0, 10.2). For participants with multiple head injuries in the past 10 years, the risk elevation was more than 11 fold. The meta-analysis also indicated a moderately elevated risk of ALS among individuals with previous head injuries (OR=1.7, 95 percent CI: 1.3, 2.2). In our study population, physical injuries of other body parts, including trunk, arms, or legs, were not related to ALS risk. These data support the notion that head injury may increase the risk of ALS. | <a href="https://www.doi.org/10.1093/aje/kwm153">https://www.doi.org/10.1093/aje/kwm153</a> , <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2239342</a> | New EnglandJepidemiology,SoccerInjuries,CranioCerebral Trauma complications,Amyotrophic Lateral Sclerosis epidemiology |

Εικόνα 28: Παράθυρο με εκτενείς πληροφορίες για το document

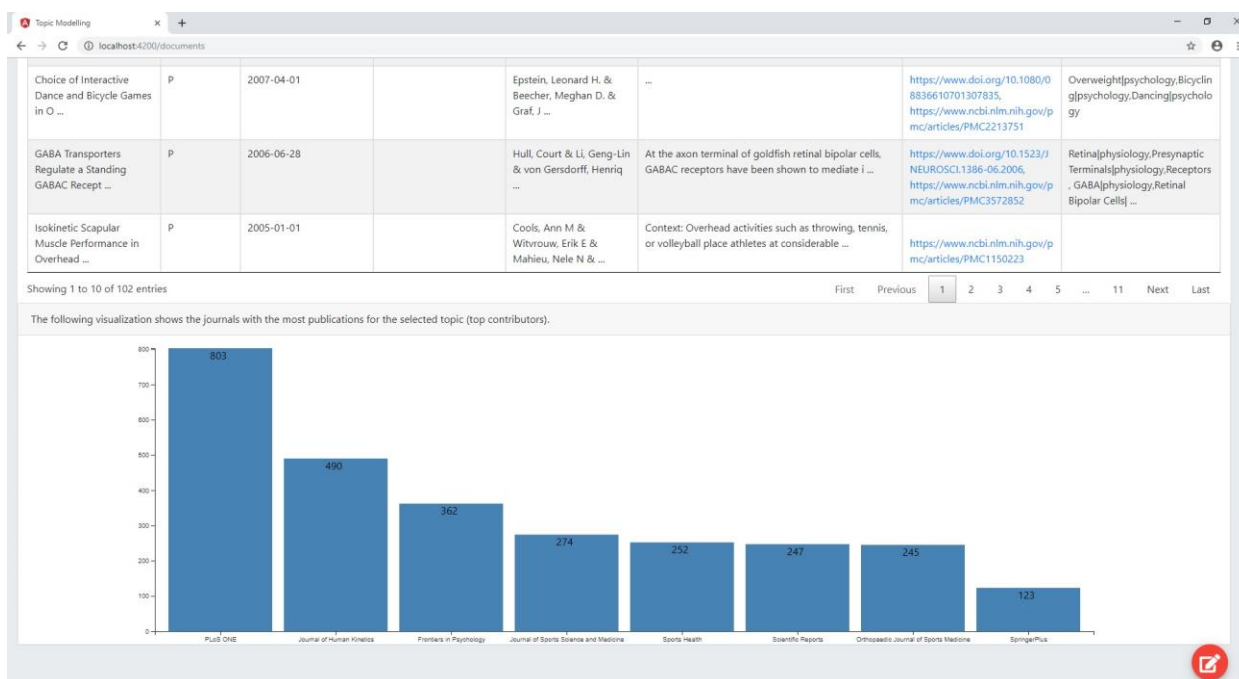
Στο παράθυρο αυτό μπορούμε να διαβάσουμε ολόκληρη την περίληψη του document, να συλλέξουμε πληροφορίες για τους authors, την εφημερίδα που δημοσιοποιήθηκε το topic, τα links που οδηγούν στην αυθεντική δημοσίευση, όπως επίσης και τα 10 πιο



αντιπροσωπευτικά keywords του topic. Πατώντας “x” ή “Close” κλείνουμε το συγκεκριμένο παράθυρο.

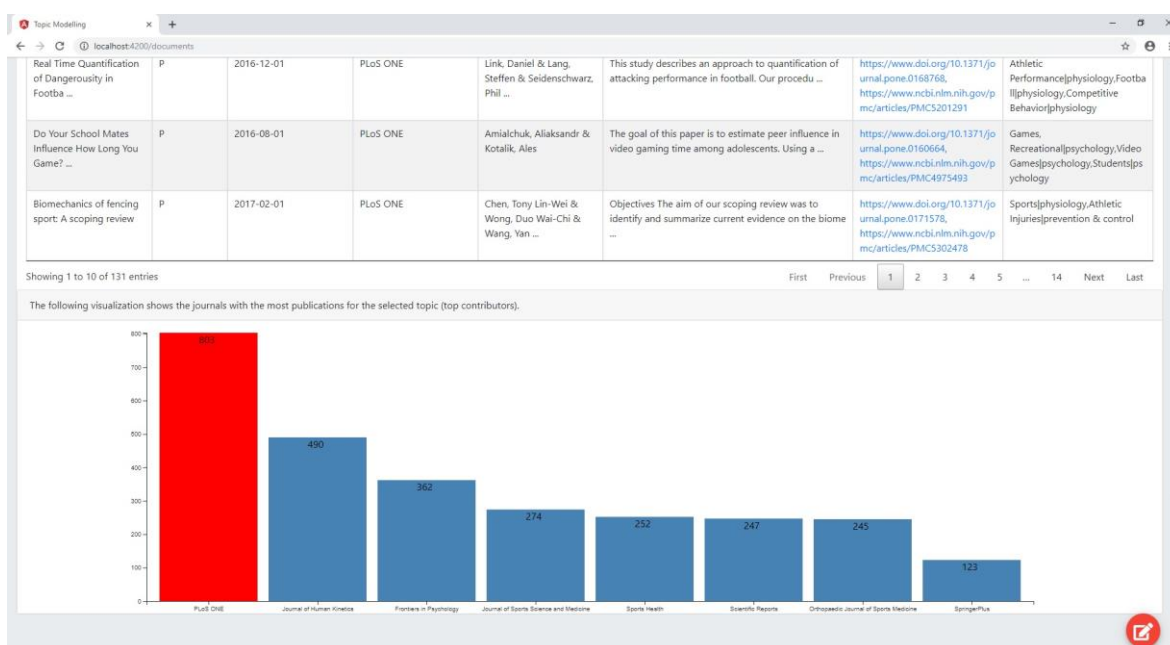
Προχωρώντας πιο κάτω στην οθόνη θα δούμε ότι και σε αυτόν τον πίνακα, όπως και στον προηγούμενο των topics, υπάρχει σελιδοποίηση στο κάτω δεξιό μέρος του πίνακα, ώστε ο χρήστης να μπορεί να πλοηγηθεί στις εγγραφές του πίνακα.

Κάτω από τον πίνακα, όπως φαίνεται και στην **Εικόνα 29** έχουμε ένα ακόμα διαδραστικό γράφημα της πλατφόρμας μας. Πρόκειται για ένα ιστόγραμμα. Το συγκεκριμένο γράφημα το επιλέξαμε για να δείξουμε τις εφημερίδες που έχουν τις περισσότερες δημοσιεύσεις πάνω στο topic που επιλέξαμε να μελετήσουμε.



**Εικόνα 29: Ιστόγραμμα journal-πλήθος documents**

Στον οριζόντιο άξονα λοιπόν φαίνονται τα ονόματα των εφημερίδων και στο πάνω μέρος κάθε ράβδου, φαίνεται το πλήθος των documents που έχουν δημοσιευθεί από τη συγκεκριμένη εφημερίδα και αναφέρονται στο επιλεγμένο topic.



**Εικόνα 30: Φιλτράρισμα των documents συγκεκριμένου journal**

Πατώντας πάνω σε κάποιο από τα journal, φιλτράρονται μόνο τα documents αυτού στον από πάνω πίνακα, όπως φαίνεται στην **Εικόνα 30**. Στην εικόνα αυτή για παράδειγμα επιλέξαμε το PLoS ONE journal στο ιστόγραμμα, με αποτέλεσμα στον πίνακα να εμφανίζονται μόνο τα documents του journal αυτού. Με αυτόν τον τρόπο ο curator μπορεί να επικεντρωθεί στη μελέτη μόνο των documents ενός συγκεκριμένου journal αν το επιθυμεί. Αν θέλει να αφαιρέσει το συγκεκριμένο φίλτρο και να δει πάλι τα documents όλων των journal στον πίνακα, αρκεί απλά να επιλέξει το checkbox με την ετικέτα “View All” που υπάρχει κάτω από την επικεφαλίδα “Journal”, όπως φαίνεται και στην Εικόνα 27.

### 3.3 Συμπεράσματα

Στο κεφάλαιο αυτό κάναμε μια εκτενή παρουσίαση του εργαλείου μας. Αναλύσαμε όλες τις οθόνες και τις λειτουργικότητες χρησιμοποιώντας και εικόνες μέσα από το ήδη υλοποιημένο εργαλείο. Πρόκειται για μια πλατφόρμα που απευθύνεται κυρίως σε curators και δημιουργήθηκε για να τους διευκολύνει τη ζωή και να τους εξοικονομήσει όσο περισσότερο χρόνο γίνεται στη διαδικασία του curation. Όπως μπορούμε να παρατηρήσουμε λοιπόν και από την παραπάνω παρουσίαση, πρόκειται για ένα εργαλείο που προσφέρει ολοκληρωμένη υποστήριξη στη διαδικασία του curation και είναι αρκετά απλό στη δομή του. Ο λόγος που επιλέξαμε να είναι απλό στη δομή του είναι για να μην χρειάζεται ιδιαίτερη βοήθεια ο χρήστης να το χειριστεί και να μην καθυστερεί όταν κάνει curate τα topics. Αυτή είναι λοιπόν η ιδέα μας υλοποιημένη και τώρα αρκεί να το δοκιμάσουν οι curators και να μείνουν έκπληκτοι από το πόσο θα τους βοηθήσει στη δουλειά τους.

## 4. ΣΥΓΚΡΙΣΗ ΜΕ ΑΛΛΑ ΕΡΓΑΛΕΙΑ

Σε αυτό το κεφάλαιο θα παρουσιάσουμε εφαρμογές που έχουν αντικείμενο συναφές με το αντικείμενο της παρούσας εργασίας. Θα ασχοληθούμε λοιπόν με εφαρμογές μέσω των οποίων ο curator μπορεί να υλοποιήσει το curation του topic modeling. Στο τέλος αυτής της ενότητας θα συγκρίνουμε τις εφαρμογές που βρήκαμε με τη δική μας και θα καταλήξουμε με τα συμπεράσματά μας.

### 4.1 Παρουσίαση άλλων εργαλείων

Είναι γεγονός ότι δεν υπάρχουν αυτήν τη στιγμή στην αγορά πολλές εφαρμογές που να ασχολούνται με το topic modeling curation και να προσομοιάζουν με αυτό που επιδιώκουμε να υλοποιήσουμε μέσα από την παρούσα εργασία-εφαρμογή. Ψάχνοντας λοιπόν για μερικές τέτοιες εφαρμογές διαπιστώθηκε ότι οι περισσότερες παρότι βοηθούν τον χρήστη να εκτιμήσει και να αξιολογήσει τα δεδομένα που εξάγει ένα στατικό topic model, ωστόσο δεν καλύπτουν όλη τη διαδικασία του curation που περιλαμβάνει τα ακόλουθα ανεξάρτητα μεταξύ τους βήματα:

- αξιολόγηση των topics ως “good” ή “poor”
- ονοματοδοσία των topics
- κατηγοριοποίηση τους σε μία ή περισσότερες κατηγορίες
- συγχώνευση δύο ή περισσότερων topics με κοινό περιεχόμενο
- διαχωρισμός ενός topic σε περισσότερα, ανάλογα με το πόσα θέματα εμπλέκει

Από τα εργαλεία που θα παρουσιάσουμε τα δύο μόνο μοιάζουν κάπως περισσότερο με αυτό που θέλουμε να φτιάξουμε αλλά και αυτά αν και είναι πολύ κοντά στις προσδοκίες μας, είτε υστερούν σε κάποια από τα παραπάνω βήματα της διαδικασίας του curation είτε δεν είναι τόσο εύχρηστα, είτε δεν παρέχουν αρκετή πληροφορία στον curator.

Στις υποενότητες 5.1.1 – 5.1.4 παρουσιάζουμε τα εργαλεία αυτά.

#### 4.1.1 Interactive Topic Modeling (ITM)

Το πρώτο εργαλείο λοιπόν που βρήκαμε είναι το ITM [1]. Το όνομά του προέρχεται από την σύντμηση των λέξεων Interactive Topic Modeling. Πρόκειται για ένα διαδραστικό εργαλείο που πέρα από την εξερεύνηση των εξαγόμενων από το topic modeling δεδομένων, δίνει τη δυνατότητα στο χρήστη να προσθέσει, να δώσει έμφαση και να αγνοήσει κάποιες λέξεις των topics.

**Διαφορές με τη δική μας πλατφόρμα:** Η διαφορά μεταξύ των δύο εργαλείων είναι μεγάλη, καθώς το ITM περιορίζεται μόνο στη διαχείριση των λέξεων και όχι των ίδιων των topics. Όσον αφορά λοιπόν τη διαδικασία του curation θα μπορούσαμε να πούμε ότι καλύπτει μόνο την αξιολόγηση σε “good” ή “poor” μέσα από την αξιολόγηση των λέξεων του topic, αφού το topic που διαθέτει μεγάλο πλήθος από good λέξεις θα είναι και good στο σύνολό του.

**Ομοιότητες με τη δική μας πλατφόρμα:** Δεν υπάρχουν ομοιότητες.

**Υστερεί** λοιπόν στο μεγαλύτερο μέρος της διαδικασίας του curation και δεν θα μπορούσε σε καμία περίπτωση να αποτελέσει βασική εφαρμογή για curation topic modeling αφού δεν υλοποιεί τέσσερις από τις πέντε βασικές ενέργειες του curation.

| Topic | Words  | Topic | Words  |
|-------|--|-------|--|
| 1     | election, <b>yeltsin</b> , <b>russian</b> , political, party, democratic, <b>russia</b> , president, democracy, <b>boris</b> , country, south, years, month, government, vote, since, leader, presidential, military | 1     | election, democratic, south, country, president, party, <b>africa</b> , <b>lead</b> , <b>even</b> , democracy, leader, presidential, <b>week</b> , <b>politics</b> , <b>minister</b> , <b>percent</b> , <b>voter</b> , <b>last</b> , month, years                |
| 2     | new, york, city, state, mayor, budget, giuliani, council, cuomo, gov, plan, year, rudolph, dinkins, lead, need, governor, legislature, pataki, david   | 2     | new, york, city, state, mayor, budget, council, giuliani, gov, cuomo, year, rudolph, dinkins, legislature, plan, david, governor, pataki, need, cut  |
| 3     | nuclear, arms, weapon, defense, treaty, missile, world, unite, yet, soviet, lead, secretary, would, control, korea, intelligence, test, nation, country, testing   | 3     | nuclear, arms, weapon, treaty, defense, war, missile, may, come, get, american, world, would, need, lead, get, join, yet, clinton, nation  |
| 4     | president, bush, administration, clinton, american, force, reagan, war, unite, lead, economic, iraq, congress, america, iraqi, policy, aid, international, military, see   | 4     | president, administration, bush, clinton, war, unite, force, reagan, american, america, make, nation, military, iraq, iraqi, troops, international, country, yesterday, plan   |
| ⋮     |  | ⋮     |  |
| 20    | <b>soviet</b> , lead, <b>gorbachev</b> , <b>union</b> , west, <b>mikhail</b> , reform, change, europe, leaders, poland, <b>communist</b> , know, old, right, human, washington, western, bring, party                | 20    | <b>soviet</b> , <b>union</b> , economic, reform, <b>yeltsin</b> , <b>russian</b> , lead, <b>russia</b> , <b>gorbachev</b> , leaders, west, president, <b>boris</b> , <b>moscow</b> , europe, poland, <b>mikhail</b> , <b>communist</b> , power, <b>relations</b> |

Εικόνα 31: Επιφάνεια διεπαφής ITM

#### 4.1.2 UTOPIAN

Το δεύτερο εργαλείο που βρήκαμε είναι το UTOPIAN [2]. Το όνομα του είναι μια σύντμηση των λέξεων User-driven Topic modeling based on Interactive Nonnegative Matrix Factorization. Έχει δημιουργηθεί από το Πανεπιστήμιο της Κορέας και αποτελεί μια πλατφόρμα οπτικής ανάλυσης δεδομένων topic modeling.

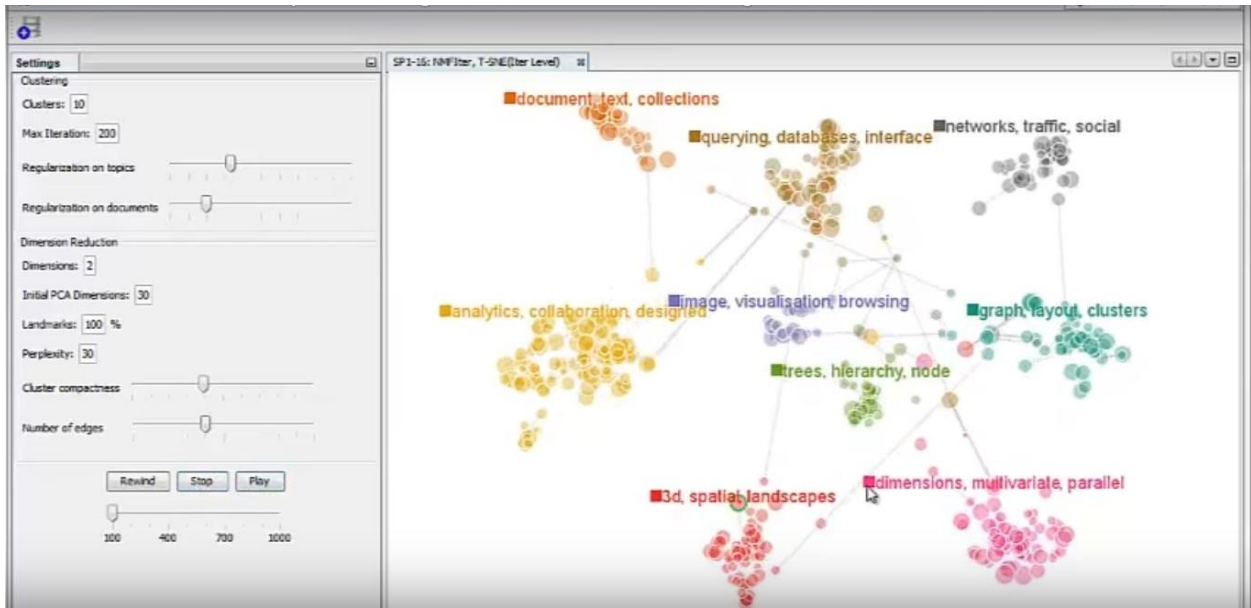
Το UTOPIAN δίνει τη δυνατότητα στον χρήστη να ορίσει το βάρος των λέξεων μέσα σε ένα topic, να κάνει συγχώνευση και διαχωρισμό των topics, όπως και να δημιουργήσει καινούρια topics. Όλες αυτές οι ενέργειες βασίζονται στη διαχείριση των λέξεων που αντιπροσωπεύουν το topic. Δηλαδή, για παράδειγμα η συγχώνευση δυο topic συγχωνεύει τις λέξεις τους. Αντίστοιχα κατά τον διαχωρισμό ενός topic σε δύο ή περισσότερα, κατανέμονται οι λέξεις του topic στα επιμέρους topics. Γενικά λοιπόν όλη η ανάλυση των topic περιορίζεται στο επίπεδο των λέξεων τους.

**Διαφορές με τη δική μας πλατφόρμα:** Η διαφορά μεταξύ των δύο εργαλείων έγκειται στο ότι το UTOPIAN δεν δίνει τη δυνατότητα κατηγοριοποίησης των topic, όπως και στο ότι προσφέρει πολύ λιγότερη πληροφορία όσον αφορά το topic σε σύγκριση με τη δική μας πλατφόρμα.

**Ομοιότητες με τη δική μας πλατφόρμα:** Δεν υπάρχουν ορατές ομοιότητες. Θα μπορούσαμε να πούμε τη συγχώνευση και το διαχωρισμό των topic αλλά ο τρόπος με τον οποίο θα υλοποιηθεί στο δικό μας εργαλείο δεν μοιάζει και τόσο με τον τρόπο που υλοποιείται στο UTOPIAN.

**Υστερεί** λοιπόν στο ότι δεν δίνει τη δυνατότητα κατηγοριοποίησης των topic, όπως επίσης και ονοματοδοσίας τους, καθώς όπως βλέπουμε και στην Εικόνα 6 κάθε topic εκπροσωπείται από τις τρεις πιο σημαντικές λέξεις του. Τέλος στα μείον του μπορεί να αναφερθεί και το ότι δεν παρέχεται λειτουργία αξιολόγησης των topics, με το curation να περιορίζεται μόνο στην επεξεργασία τους. Συνεπώς αν ένα topic είναι “root” και δεν

μπορεί να αξιοποιηθεί ούτε μέσω της συγχώνευσης ούτε μέσω της προσθαφαίρεσης λέξεων σε αυτό, δε δίνεται η δυνατότητα να αγνοηθεί ή ακόμα και να αφαιρεθεί.



Εικόνα 32: Επιφάνεια διεπαφής UTOPIAN

#### 4.1.3 iVisClustering

Το iVisClustering [3] είναι ένα από τα δύο πιο ολοκληρωμένα εργαλεία που εντοπίσαμε κατά την έρευνά μας στην αγορά. Πρόκειται για ένα σύστημα οπτικής ανάλυσης δεδομένων του topic modeling, το οποίο επικεντρώνεται στο διαδραστικό clustering των documents.

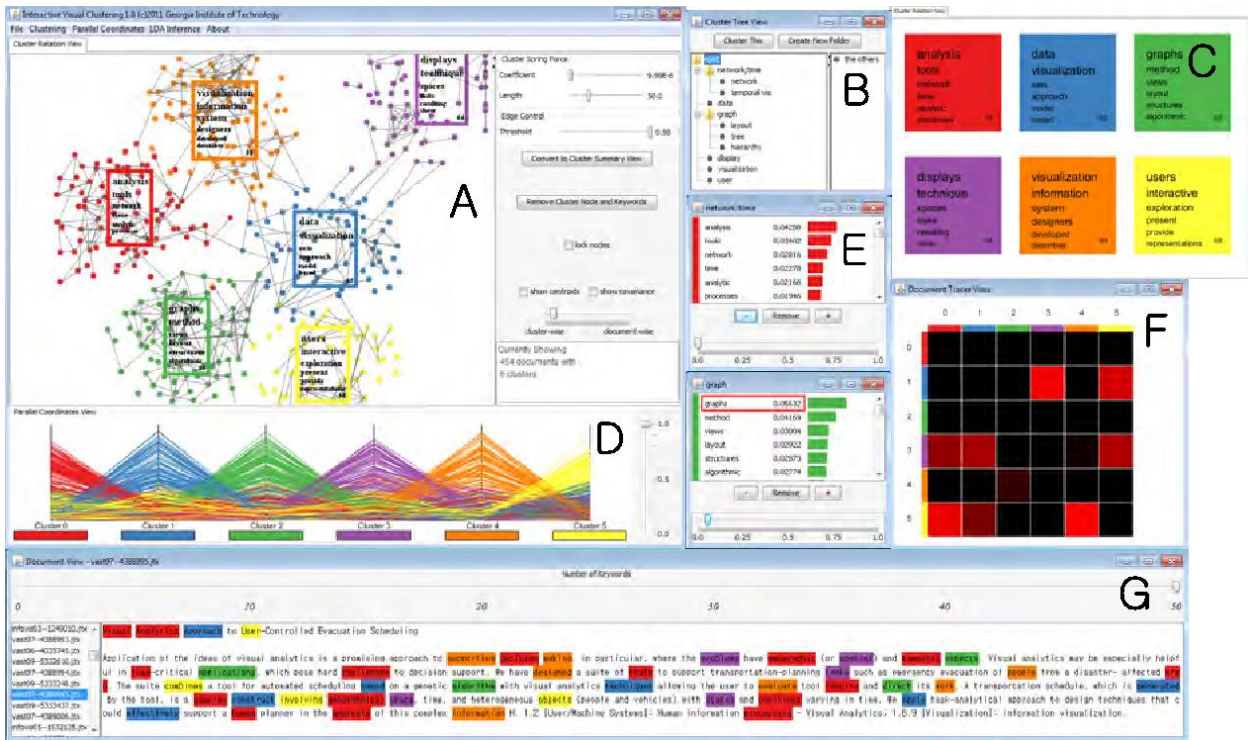
Όσον αφορά τη διαδικασία του curation καλύπτει το μεγαλύτερο μέρος της καθώς επιτρέπει την ονοματοδοσία των topics, την συγχώνευσή τους και τον διαχωρισμό τους. Ο διαχωρισμός των topics βέβαια δεν μπορεί να γίνει απευθείας στην πλατφόρμα, αλλά είναι εφικτός μέσω μιας σειράς ενεργειών αφού προσφέρεται από την πλατφόρμα η δημιουργία και η διαγραφή topic αλλά και η μετακίνηση των documents από ένα topic σε άλλο.

**Διαφορές με τη δική μας πλατφόρμα:** Βασικό χαρακτηριστικό διαφοροποίησης του iVisClustering με τη δική μας πλατφόρμα είναι ότι η παρουσίαση των δεδομένων του topic modeling επικεντρώνεται στα documents, και αναφέρεται στα topics σαν clusters των documents. Αντίθετα το δικό μας εργαλείο βάζει στο επίκεντρο το topic, που είναι και το βασικό αντικείμενο αξιολόγησης, ο πρωταγωνιστής του curation, και απλά χρησιμοποιεί τα documents όταν χρειάζεται ο curator περισσότερη πληροφορία σχετικά με το topic που επιθυμεί να αξιολογήσει. Τα δύο εργαλεία λοιπόν διαφέρουν ως προς την σκοπιά που επιλέγουν να παρουσιάσουν στον χρήστη την πληροφορία που θα χρειαστεί ώστε να βγάλει τα συμπεράσματά του και εν συνεχεία να υλοποιήσει το curation.

**Ομοιότητες με τη δική μας πλατφόρμα:** Υπάρχουν αρκετές ομοιότητες με τη δική μας πλατφόρμα καθώς και τα δύο εργαλεία επιτρέπουν την ονοματοδοσία ενός topic, τη συγχώνευση δύο ή περισσότερων topics, αλλά και τον διαχωρισμό ενός topic σε περισσότερα. Επιπλέον το iVisClustering όπως και η δική μας εφαρμογή παρέχει αρκετή πληροφορία στον χρήστη γραπτή και οπτικοποιημένη, που τον βοηθάει στην κατανόηση των δεδομένων και στην αξιολόγησή τους.



**Υστερεί** λοιπόν στο ότι δεν δίνει τη δυνατότητα κατηγοριοποίησης των topic, όπως επίσης και τη δυνατότητα της άμεσης αξιολόγησης τους ως “good” ή “poor” που διευκολύνει τον curator να φιλτράρει με ποια topic αξίζει να ασχοληθεί και με ποια όχι.



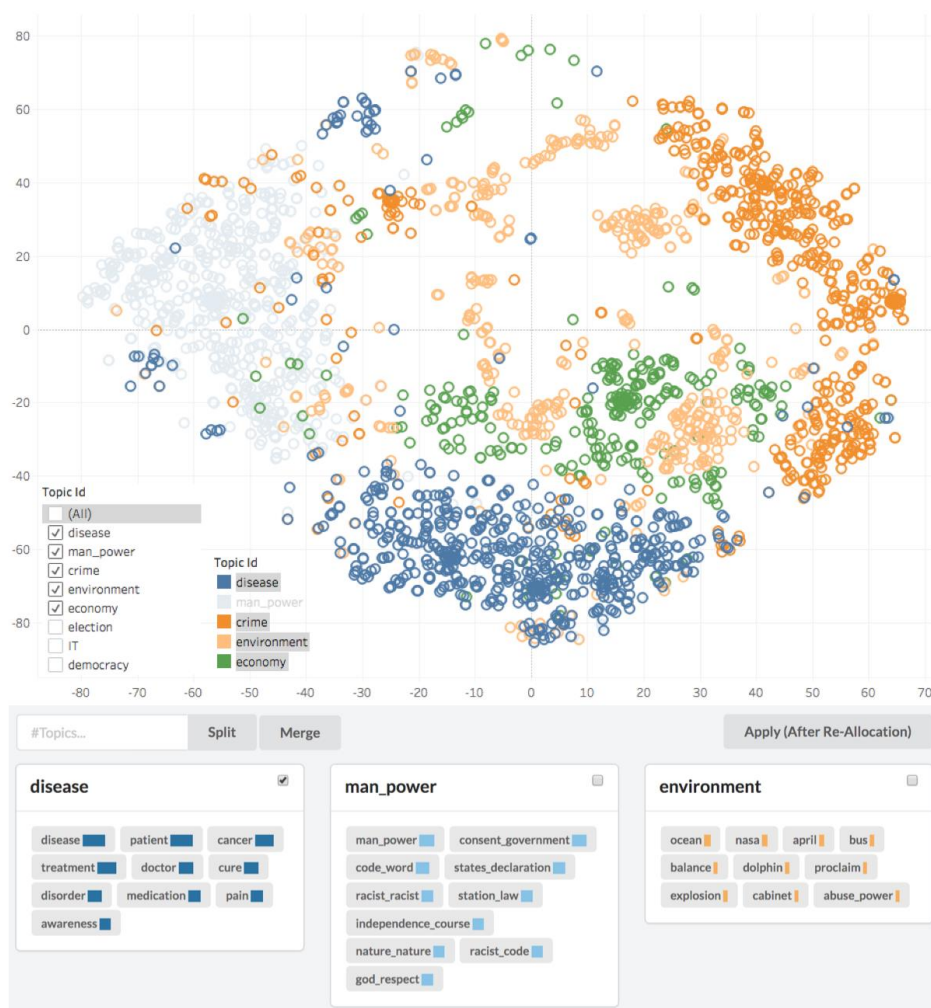
Εικόνα 33: Επιφάνεια διεπαφής iVisClustering

#### 4.1.4 Interactive Visualization for Topic Model Curation

Το συγκεκριμένο εργαλείο [4] είναι το δεύτερο πιο ολοκληρωμένο που εντοπίσαμε όσον αφορά το topic modeling curation, που αποτελεί και το βασικό αντικείμενο της δικής μας πλατφόρμας. Είναι μια κοινή δουλειά του Penn State University, USA και του Lanzhou University, China που παρουσιάστηκε τον Μάρτιο του 2018 στο Τόκιο.

Οι δυνατότητες που προσφέρει στον χρήστη είναι η συγχώνευση δύο ή περισσότερων topic που εμφανίζουν ομοιότητες ως προς το περιεχόμενό τους και αντίστοιχα η διάσπαση ενός topic σε δύο ή περισσότερα, ανάλογα με το πλήθος των θεματικών ενοτήτων που μπερδεύει στο περιεχόμενό του. Επιπλέον παρέχει τη δυνατότητα της αξιολόγησης ενός topic ως “good” ή “poor” μέσω της εποπτείας και της επεξεργασίας των λέξεων που το αντιπροσωπεύουν. Σε περίπτωση που διακρίνει λέξεις που δε συμβαδίζουν με το περιεχόμενο του topic μπορεί να τις αφαιρέσει από αυτό, όπως επίσης και να προσθέσει λέξεις που ταυτίζονται με το περιεχόμενο του topic και οι οποίες φαίνεται εκ των υστέρων ότι κατά το topic modeling κακώς δε συμπεριλήφθηκαν σε αυτό.

**Διαφορές με τη δική μας πλατφόρμα:** Το συγκεκριμένο εργαλείο αν και είναι αυτό που προσεγγίζει περισσότερο την ιδέα μας, διαφέρει αισθητά από το εργαλείο που έχουμε αναπτύξει. Κάποιες από τις διαφορές αυτές είναι ότι δεν παρέχει τη δυνατότητα της κατηγοριοποίησης των topic όπως και τη δυνατότητα της ονοματοδοσίας των topic. Επιπλέον παρέχει πολύ λιγότερη πληροφορία σε σύγκριση με το δικό μας και πέρα από τον γράφο δε διαθέτει άλλα σχήματα που θα βοηθήσουν τον χρήστη να κατανοήσει πιο εύκολα την πληροφορία που του παρέχεται.



Εικόνα 34: Επιφάνεια διεπαφής Interactive Visualization for Topic Model Curation

**Ομοιότητες με τη δική μας πλατφόρμα:** Οι ομοιότητες της συγκεκριμένης πλατφόρμας με τη δικιά μας αφορούν τη διαδικασία του curation. Και στις δύο δίνεται η δυνατότητα της συγχώνευσης και της διάσπασης των topics. Επίσης το συγκεκριμένο εργαλείο, όπως και το δικό μας, παρέχει στον χρήστη τη δυνατότητα της εξερεύνησης των δεδομένων, που όπως είδαμε παραπάνω δε θεωρείται δεδομένη και δεν παρέχεται σε όλα τα παραπάνω εργαλεία.

**Υστερεί** λοιπόν στο ότι δεν δίνει τη δυνατότητα κατηγοριοποίησης των topic, όπως επίσης και τη δυνατότητα της ονοματοδοσίας τους. Επιπλέον δεν παρέχει αρκετή πληροφορία ώστε να μπορεί ο curator να έχει καλύτερη εικόνα των topics καθώς τα αξιολογεί. Τέλος δεν υπάρχει αρκετό υλικό όσον αφορά την οπτικοποίηση των δεδομένων, καθώς το μόνο σχήμα που παρέχεται είναι αυτό του κεντρικού γράφου.

## 4.2 Σύνοψη σύγκρισης με τη δική μας πλατφόρμα

Στο συγκεκριμένο υποκεφάλαιο θα συνοψίσουμε τις διαφορές της δικής μας πλατφόρμας με εκείνες που παρουσιάστηκαν στο υποκεφάλαιο 5.1.

Θα αναφέρουμε τις διαφορές όπως αυτές τονίστηκαν και κατά την παρουσίαση των παραπάνω εφαρμογών και θα επισημάνουμε που υπερτερεί η δικιά μας εφαρμογή έναντι των άλλων.

(Στον πίνακα η δεύτερη στήλη περιέχει τις διαφορές και η τρίτη τα σημεία στα οποία υπερτερεί η εφαρμογή μας)

Στο σημείο αυτό πρέπει να αναφέρουμε ότι στην έκδοση της πλατφόρμας μας, που παρουσιάζουμε στο παρόν κείμενο, δεν είναι υλοποιημένη η συγχώνευση και η διάσπαση topic. Όμως δεδομένου ότι είναι σχεδόν έτοιμα να προστεθούν, και μεν θα τα αναφέρουμε αργότερα ως μελλοντικές επεκτάσεις της πλατφόρμας, διότι δεν έχουν προστεθεί ακόμα, αλλά στο παρόν κεφάλαιο κάνουμε τη σύγκριση με τα υπόλοιπα εργαλεία θεωρώντας δεδομένο ότι έχει γίνει η προσθήκη αυτή. Ο λόγος είναι ότι εφόσον θα προστεθούν, θέλουμε να συγκρίνουμε τη τελική έκδοση της πλατφόρμας μας με τις υπόλοιπες της αγοράς, ώστε να είναι η σύγκριση πιο ολοκληρωμένη.

|  | <b>Διαφορές</b>  | <b>Πού υπερτερούμε</b>  |
|--|--|---|
| <b><u>Interactive Topic Modeling (ITM)</u></b>                   | Το ITM περιορίζεται μόνο στη διαχείριση των λέξεων και όχι των ίδιων των topics.                   | Η πλατφόρμα μας παρέχει τη διαχείριση των topic στο σύνολό τους.                            |
| <b><u>UTOPIAN</u></b>  | Δεν παρέχει αρκετή πληροφορία στον χρήστη.   | Η εφαρμογή μας παρέχει αρκετή πληροφορία.   |
|  | Δεν δίνει τη δυνατότητα κατηγοριοποίησης, ονοματοδοσίας και σήμανσης του topic ως "good" ή "root". | Εμείς παρέχουμε και τις τρεις λειτουργίες μέσω της πλατφόρμας μας.                          |
| <b><u>iVisClustering</u></b>                                     | Σκοπιά παρουσίασης δεδομένων. (documents)  | Σκοπιά παρουσίασης δεδομένων. (topics)<br>Καλύτερη εικόνα του προς αξιολόγηση αντικειμένου. |
|  | Δεν δίνει τη δυνατότητα κατηγοριοποίησης και σήμανσης του topic ως "good" ή "root".                | Οι λειτουργίες αυτές παρέχονται μέσω της πλατφόρμας μας.                                    |
| <b><u>Interactive Visualization for Topic Model Curation</u></b> | Όχι αρκετή πληροφορία (γραπτή ή οπτικοποιημένη) προς τον χρήστη.                                   | Αρκετή και χρήσιμη πληροφορία ώστε να διευκολυνθεί ο χρήστης στην αξιολόγηση των topic.     |
|  | Δεν δίνει τη δυνατότητα κατηγοριοποίησης και ονοματοδοσίας των topic.                              | Η πλατφόρμα μας παρέχει αυτές τις λειτουργίες στον χρήστη.                                  |

### 4.3 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκαν μερικές εφαρμογές που υπάρχουν ήδη στην αγορά και έχουν το ίδιο αντικείμενο με τη δική μας, το curation. Κατά την παρουσίαση τους αναφέρθηκαν οι ομοιότητες τους με τη δική μας εφαρμογή, αλλά και οι διαφορές τους και τα σημεία στα οποία υπερτερεί η δική μας έναντι αυτών. Είδαμε ότι υπάρχουν εφαρμογές που πέρα από την διαχείριση των λέξεων ενός topic δεν προσφέρουν καμία



άλλη λειτουργικότητα στον curator. Είδαμε ότι υπάρχουν εφαρμογές που καλύπτουν αρκετά από τα βήματα της διαδικασίας του curation, χωρίς ωστόσο να παρέχουν αρκετή πληροφορία στον curator για να τον διευκολύνουν με την αξιολόγηση των topic. Τέλος είδαμε και εργαλεία που καλύπτουν αρκετά από τα βήματα του curation και ταυτόχρονα παρέχουν αρκετή πληροφορία προς τον χρήστη και τη δυνατότητα για εξερεύνηση των δεδομένων. Συμπερασματικά, διαπιστώνουμε ότι αυτήν τη στιγμή στην αγορά δεν υπάρχει αυτό το οποίο σχεδιάζουμε να υλοποιήσουμε. Δεν υπάρχει δηλαδή μια πλατφόρμα που θα καλύπτει όλη τη διαδικασία του curation και ταυτόχρονα θα δίνει τη δυνατότητα στον χρήστη για εξερεύνηση των δεδομένων. Αυτός λοιπόν είναι και ο λόγος που η συγκεκριμένη πλατφόρμα θα φανεί πολύ χρήσιμη στους curators και όχι μόνο.

## 5. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Στο κεφάλαιο αυτό κάνουμε μια ανασκόπηση σε όλα όσα παρουσιάστηκαν στα προηγούμενα κεφάλαια. Στη συνέχεια αναφέρουμε τα συμπεράσματά μας όσον αφορά το εργαλείο, τον τρόπο που δουλέψαμε, τη χρήση του και την ανταπόκριση που είχε κατά τη δοκιμή του τουλάχιστον από τα μέλη της ομάδας μας. Τέλος επισημαίνουμε κάποιες μελλοντικές επεκτάσεις στο εργαλείο, που θα του προσδώσουν ακόμα μεγαλύτερη λειτουργικότητα και πιθανόν να εξυπηρετήσουν τα ενδιαφέροντα περισσότερων ομάδων χρηστών, χωρίς ωστόσο να χάνει τον βασικό σκοπό ύπαρξής του, που είναι το curation.

### 5.1 Ανασκόπηση

Σε αυτήν την ενότητα θα κάνουμε μια ανασκόπηση όλων όσων παρουσιάστηκαν στα προηγούμενα κεφάλαια.

Στο πρώτο κεφάλαιο αναφερθήκαμε στο αντικείμενο της εργασίας. Παρουσιάσαμε το περιεχόμενο της και τα κίνητρα που μας ώθησαν στην ενασχόλησή μας με αυτό. Στη συνέχεια αναφερθήκαμε στους στόχους μας για αυτήν την εργασία παρουσιάζοντας την ιδέα μας για ένα εργαλείο που θα απευθύνεται κυρίως σε curators και θα τους προσφέρει μια ολοκληρωμένη υποστήριξη στη διαδικασία του curation, προκειμένου να μειώσουν τους χρόνους ενασχόλησής τους και να αυξήσουν την ποιότητα των αποτελεσμάτων της εργασίας τους.

Στο δεύτερο κεφάλαιο μιλήσαμε για πιο τεχνικά θέματα. Αναφερθήκαμε στην αρχιτεκτονική της πλατφόρμας μας. Τη χωρίσαμε σε δύο κομμάτια, τον εξυπηρετητή και τον πελάτη. Για καθένα από αυτά, παρουσιάσαμε το περιβάλλον ανάπτυξης που χρησιμοποιήσαμε αλλά και τις γλώσσες προγραμματισμού και τις βιβλιοθήκες που αποτέλεσαν τα εργαλεία μας στην υλοποίηση της ιδέας μας, που ήταν η ανάπτυξη της συγκεκριμένης πλατφόρμας.

Στο τρίτο κεφάλαιο παρουσιάσαμε την πλατφόρμα μας υλοποιημένη. Παραθέσαμε εικόνες μέσα από το εργαλείο και αναλύσαμε ένα προς ένα τα δομικά στοιχεία κάθε οθόνης, και τις λειτουργικότητες που προσφέρονται στο χρήστη.

Στο τέταρτο κεφάλαιο αναφερθήκαμε σε άλλα εργαλεία που υπάρχουν αυτήν τη στιγμή στην αγορά και αφορούν το curation των εξαγόμενων δεδομένων από το topic model. Είδαμε αναλυτικά ομοιότητες και διαφορές με το δικό μας εργαλείο και διαπιστώσαμε ότι δεν υπάρχει κάποιο που να προσομοιώνει αυτό ακριβώς που έχουμε στο μυαλό μας.

Στο πέμπτο και τελευταίο κεφάλαιο, που είναι και το παρόν κεφάλαιο, όπως είπαμε θα κάνουμε αρχικά μια ανασκόπηση. Στη συνέχεια θα αναφέρουμε τα συμπεράσματά μας και θα καταλήξουμε παρουσιάζοντας τις μελλοντικές επεκτάσεις της παρούσας εργασίας.

### 5.2 Συμπεράσματα

Φθάνοντας λοιπόν προς το τέλος της εργασίας μας και έχοντας έτοιμο το εργαλείο του curation, τουλάχιστον ως προς το μεγαλύτερο μέρος του, μπορούμε να κοιτάξουμε όλη αυτήν τη διαδρομή και να βγάλουμε κάποια συμπεράσματα.

Μελετώντας τον τρόπο με τον οποίο εργαστήκαμε, και συγκρίνοντας το αποτέλεσμα με αυτό που είχαμε στο μυαλό μας αρχικά πιστεύουμε ότι φθάσαμε στο καλύτερο δυνατό αποτέλεσμα. Σίγουρα λείπουν κάποιες επιπλέον λειτουργικότητες και κάποια επιπλέον

γραφήματα, για να πούμε ότι υλοποιήσαμε πλήρως την ιδέα μας αλλά παραμένουν σαν στόχος για μελλοντική επέκταση.

Όσον αφορά λοιπόν το εργαλείο είδαμε ότι ο σχεδιασμός του είναι απλός, ενώ ταυτόχρονα προσφέρει αρκετές λειτουργικότητες και πλήρη υποστήριξη στους curators. Καθώς το σχεδιάζαμε και το αναπτύσσαμε είχαμε συνέχεια στο μυαλό μας το κοινό στο οποίο θα απευθύνεται. Θέλαμε να υπάρχει αμεσότητα και ευκολία στη χρήση του από τους curators που δεν είναι ειδικοί του topic modeling και συνεπώς θέλουν να βλέπουν την πληροφορία όσο πιο κατανοητά γίνεται ενώ ταυτόχρονα να μπορούν να ολοκληρώσουν τη διαδικασία του curation μέσω αυτή της πλατφόρμας. Έτσι βλέποντας υλοποιημένο το συγκεκριμένο εργαλείο, μπορεί κανείς εύκολα να διαπιστώσει τόσο την ευκολία στην χρήση του όσο και το σκοπό ύπαρξής του. Έχουμε λοιπόν έτοιμο ένα εργαλείο για ολοκληρωμένο curation, το οποίο παραμένει εύκολο στη χρήση του. Αυτό ακριβώς που λείπει αυτήν τη στιγμή από την αγορά. Μια πλατφόρμα που θα αποτελέσει απαραίτητο εργαλείο στα χέρια των curators, ενώ ταυτόχρονα θα μπορεί να χρησιμοποιηθεί και από απλούς ερευνητές που θέλουν να μελετήσουν τα εξαγόμενα από το topic model δεδομένα.

Ήμασταν λοιπόν πολύ περίεργοι να δούμε αν όλα αυτά που σκεφτόμασταν ανταποκρίνονται στην πραγματικότητα. Είναι το εργαλείο μας τόσο απλό στη χρήση του; Μπορεί ο χρήστης να πάρει εύκολα και γρήγορα την πληροφορία που επιθυμεί, χωρίς να χρειάζεται να είναι ειδικός του topic modeling; Βελτιώνει το curation και βοηθάει τους curators να υλοποιήσουν πιο γρήγορα και πιο αποτελεσματικά τη συγκεκριμένη διαδικασία;

Εκτενή αξιολόγηση και διερεύνηση αυτών των ερωτημάτων δεν είχαμε τη δυνατότητα να κάνουμε, δηλαδή να δώσουμε το εργαλείο να δοκιμαστεί από curators. Παρόλα αυτά έχουμε τις πρώτες εντυπώσεις καθώς δοκιμάστηκε από κάποιους ερευνητές εντός της ομάδας μας. Τα συμπεράσματά μας είναι μόνο θετικά. Οι περισσότεροι δεν είχαν να επισημάνουν κάποια μεγάλη αστοχία και περιορίζονταν μονό σε απόψεις και προτάσεις για εναλλακτικές περί των χρωμάτων και των θέσεων των δομικών στοιχείων της πλατφόρμας. Ένα μεγάλο θετικό είναι ότι εξοικειώνονταν γρήγορα με την πλατφόρμα και όπως μας ανέφεραν, μεγάλη συμβολή σε αυτό είχε η επιλογή του “Show Instructions” που παρέχουμε μέσα στο εργαλείο για βοήθεια προς τον χρήστη. Επιπλέον οι περισσότεροι εντυπωσιάστηκαν από τα γραφήματα και την αλληλεπίδραση με τα δεδομένα και συμφώνησαν στο ότι βοηθάνε πολύ στην κατανόηση των δεδομένων και στην εύκολη διαχείρισή τους. Συνοψίζοντας λοιπόν η γενική εικόνα και τα μηνύματα είναι ενθαρρυντικά. Φαίνεται να έχουμε πετύχει τον στόχο μας τουλάχιστον ως προς τη δημιουργία ενός εργαλείου εύκολου στη χρήση που θα προσφέρεται και για εξερεύνηση δεδομένων. Το κομμάτι της διαδικασίας του curation δεν μπορέσαμε να το αξιολογήσουμε για την ώρα, γιατί δεν είχαμε εξειδικευμένους σε αυτόν τον τομέα να το δοκιμάσουν. Μπορεί λοιπόν το δείγμα μας να είναι μικρό και να μην συμπεριλαμβάνεται σε αυτό η άποψη κάποιου curator, η γενική θετική εντύπωση όμως μας έκανε να γεμίσουμε αισιοδοξία και ελπίδα ότι το συγκεκριμένο εργαλείο θα καταφέρει να κλέψει τις εντυπώσεις και ότι θα αποτελέσει σημείο αναφοράς στο χώρο της οπτικοποίησης και του curation των δεδομένων του Topic Modeling μέσα στα επόμενα χρόνια.

### 5.3 ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Σε κάθε εργαλείο πάντα μετά την ολοκλήρωση της υλοποίησής του θεωρούμε ότι θα μπορούσαμε να ενσωματώσουμε και κάτι επιπλέον, κάτι διαφορετικό που δεν το είχαμε συμπεριλάβει στον αρχικό σχεδιασμό. Έτσι και στην περίπτωση του δικού μας τόσο κατά την υλοποίηση όσο και μετά από αυτήν, και πιο συγκεκριμένα όταν δοκιμάστηκε

από εμάς και συνεργάτες μας, προέκυψαν διάφορες ιδέες και πιθανές επεκτάσεις. Αυτές λοιπόν οι επεκτάσεις θα συζητηθούν στην παρούσα ενότητα.

Ξεκινώντας από τις πιο σημαντικές λειτουργικότητες που εντοπίσαμε ότι λείπουν καθώς υλοποιούσαμε το εργαλείο μας θα αναφερθούμε στο Merge και το Split, ήτοι τη συγχώνευση δύο ή περισσότερων topics με κοινό περιεχόμενο σε ένα και τον διαχωρισμό ενός topic, που εμπλέκει δύο ή περισσότερες θεματικές ενότητες, στα επιμέρους topics αντίστοιχα. Με τις συγκεκριμένες λειτουργικότητες το εργαλείο θα είναι πιο ολοκληρωμένο όσον αφορά τη διαδικασία του curation. Οι λειτουργικότητες αυτές λοιπόν θα συμπεριληφθούν στην επόμενη έκδοση της πλατφόρμας μας.

Μια άλλη λειτουργικότητα που έχουμε σκεφτεί να συμπεριλάβουμε και αφορά το curation, είναι η αξιολόγηση ή ακόμα και η επεξεργασία (προσθήκη, διαγραφή) των όρων (λέξεων, φράσεων) που παρουσιάζονται ως αντιπροσωπευτική πληροφορία στα topics. Έχουμε λοιπόν σκεφτεί ο χρήστης να έχει τη δυνατότητα να αξιολογεί (βάσει χρωματισμού) ή καλύτερα να μπορεί να διαγράψει τους μη αντιπροσωπευτικούς και να προσθέσει άλλους πιο αντιπροσωπευτικούς όρους από μια λίστα που θα προκύπτει από την ανάλυση των documents που αναφέρονται στο εκάστοτε topic. Με αυτόν τον τρόπο ο curator θα έχει τη δυνατότητα να επέμβει ακόμα περισσότερο στην επεξεργασία και αξιολόγηση των topics.

Επιπροσθέτως, παρατηρήσαμε ότι ίσως χρειάζεται περισσότερη πληροφορία και επεξήγηση για τον curator όσον αφορά το input και output του Topic Modeling που καλείται να επιμεληθεί. Για παράδειγμα θα ήταν χρήσιμο να προσθέσουμε ένα κουμπί σαν το “Show Instructions” που θα δίνει πληροφορίες στον curator για πιο τεχνικά χαρακτηριστικά, όπως το τι εκφράζει κάθε στήλη (keywords, phrases, words κλπ) στον πίνακα των topics.

Επιπλέον, δεδομένου ότι διαπιστώσαμε ότι τα γραφήματα βοηθούν περισσότερο στην κατανόηση της πληροφορίας από τον χρήστη έχουμε στο πλάνο μας να προσθέσουμε επιπλέον γραφήματα στην επόμενη έκδοση της πλατφόρμας. Δεδομένη θα πρέπει να θεωρείται η προσθήκη ενός γράφου με πληροφορία για τις ομοιότητες μεταξύ των topics και τις κατηγορίες στις οποίες ανήκουν. Στόχος μας είναι ο χρήστης να έχει τη δυνατότητα να υλοποιήσει αρκετές από τις ενέργειες του curation, όπως merge, split, topic labeling, και μέσω αυτού του γράφου. Επίσης θέλουμε να προσθέσουμε και επιπλέον barcharts, histograms και άλλα γραφήματα, ώστε να οπτικοποιήσουμε περισσότερο την πληροφορία που παρέχεται μέσω των πινάκων.

Μια ακόμη σκέψη που έχουμε για επέκταση στο εργαλείο είναι να μπορεί ο χρήστης να επιλέξει τι θέλει να δει και τι όχι. Αυτήν τη στιγμή ο χρήστης βλέπει συγκεκριμένα δεδομένα στους πίνακες και συγκεκριμένα γραφήματα σε κάθε οθόνη. Επιθυμία μας λοιπόν είναι να μπορεί ο χρήστης μέσα από ένα menu να διαλέξει τις στήλες που επιθυμεί να εμφανίζονται σε κάθε πίνακα. Ακόμη θα έχει τη δυνατότητα να επιλέξει από μια λίστα ποια γραφήματα τον ενδιαφέρουν ώστε να εμφανίζονται μόνο αυτά στην οθόνη του. Με αυτήν την προσθήκη, θα μπορεί ο χρήστης να αλληλεπιδράσει ακόμα περισσότερο με την πλατφόρμα, να την παραμετροποιήσει και να την προσαρμόσει στα ενδιαφέροντά του.

Τέλος, μια επιπλέον επέκταση σημαντική για την πλατφόρμα μας θα ήταν η προσθήκη user management. Ο κάθε χρήστης θα δημιουργεί τον δικό του λογαριασμό και θα εισέρχεται στην πλατφόρμα μέσω αυτού. Θα ξεκλειδώσουν έτσι περισσότερες δυνατότητες, όπως το να μπορούν για παράδειγμα δυο curator να εκτελούν curation στην ίδια συλλογή των topics και στη συνέχεια να συγκρίνουν τα αποτελέσματά τους ή ακόμα και να έχουν εσωτερική επικοινωνία μεταξύ τους μέσω μηνυμάτων. Επιπλέον θα μπορούσαμε να δημιουργήσουμε ομάδες χρηστών και να παρέχουμε μόνο τις απαραίτητες λειτουργικότητες για κάθε ομάδα. Έτσι για παράδειγμα ένας ερευνητής που

θα εισέρχεται μέσα στην πλατφόρμα απλά για να εξερευνήσει τα δεδομένα δε χρειάζεται να βλέπει το κομμάτι του curation. Επίσης θα μπορούσαμε να προσθέσουμε δικαιώματα χρήσης, και για παράδειγμα σε κάθε συλλογή από topics να υπάρχει ένας administrator που θα μπορεί να αναθέσει στους curators συγκεκριμένα κομμάτια της συλλογής για curation. Στην περίπτωση αυτή ο κάθε curator θα βλέπει μόνο το κομμάτι που του αντιστοιχεί.

Αυτές λοιπόν είναι κάποιες επεκτάσεις που σκεφτήκαμε βλέποντας την πλατφόρμα μας υλοποιημένη. Κάποιοι ίσως θεωρήσουν ότι μερικές από αυτές ή και όλες θα ήταν απαραίτητες στην εφαρμογή. Αυτό που έχουμε εμείς να πούμε είναι ότι ίσως έχουν δίκιο, απλά όταν σχεδιάζαμε στην αρχή το εργαλείο κάποια δεν τα σκεφτήκαμε και κάποια άλλα τα παραλείψαμε ώστε να διατηρήσουμε την απλότητα στη χρήση της πλατφόρμας μας, που είναι και ένας από τους βασικούς σκοπούς μας. Σε αυτήν τη μορφή που είναι τώρα είναι πολύ απλό στη χρήση του και υποστηρίζει το σύνολο της διαδικασίας του curation, που είναι και ο βασικός λόγος δημιουργίας του. Φυσικά και θα ενσωματώσουμε αν όχι όλες, σίγουρα τις περισσότερες από τις ιδέες που αναφέραμε προηγουμένως στην επόμενη έκδοση της πλατφόρμας, απλά γνώμονάς μας θα είναι να παραμείνει η πλατφόρμα το ίδιο απλή όπως είναι τώρα και να μην ξεφύγει από το βασικό λόγο ύπαρξης της που είναι το curation, όπως δηλώνει και το όνομά της.

#### **5.4 Επίλογος**

Μόλις υλοποιήσαμε μια ιδέα η οποία προέκυψε ξαφνικά και τελικά μας οδήγησε να υλοποιήσουμε κάτι από το μηδέν και να καταφέρουμε να έχουμε ένα όμορφο αποτέλεσμα. Μάθαμε πάρα πολλά κατά τη διάρκεια της διαδικασίας ανάπτυξης, από το σχεδιασμό μιας ολόκληρης πλατφόρμας, μέχρι το ποια είναι τα βήματα μιας διαδικασίας curation. Επιπλέον εντρυφήσαμε σε πολλές γλώσσες προγραμματισμού και βιβλιοθήκες, κάποιες από τις οποίες τις ξέραμε ήδη αλλά τις μάθαμε καλύτερα και κάποιες άλλες τις οποίες τις γνωρίσαμε τώρα. Η μεγαλύτερη χαρά είναι ότι μας έγινε γνώριμος ένας τομέας σχεδόν άγνωστος μέχρι πριν λίγο, αυτός της δημιουργίας web πλατφόρμας. Είχε πραγματικά πολύ κόπο και σκληρή δουλειά η δημιουργία μιας τέτοιας πλατφόρμας από το μηδέν αλλά η ικανοποίηση του αποτελέσματος μας ανταμείβει και μας γεμίζει χαρά και αισιοδοξία.

## ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

| <b>Ξενόγλωσσος όρος</b> | <b>Ελληνικός Όρος</b>                                  |
|-------------------------|--|
| Administrator           | Διαχειριστής   |
| Author                  | Συγγραφέας   |
| Auto save               | Αυτόματη αποθήκευση                                    |
| Blog                    | Ιστολόγιο  |
| Build                   | Χτίσιμο  |
| Button                  | Κουμπί   |
| Category                | Κατηγορία  |
| Cluster                 | Συστάδα  |
| Clustering              | Ομαδοποίηση  |
| Component               | Συνιστώσα  |
| Concluded               | Ολοκληρώθηκε   |
| Curation                | Επιμέλεια  |
| Curator                 | Επιμελητής   |
| Desktop                 | Επιφάνεια εργασίας                                     |
| Document                | Έγγραφο  |
| Experiment              | Πείραμα  |
| Filter                  | Φιλτράρω   |
| Flow                    | Ροή  |
| Form                    | Φόρμα  |
| Framework               | Πλαίσιο Ανάπτυξης Κώδικα (Βιβλιοθήκη κώδικα)           |
| Good                    | Καλό   |
| Grid                    | Πλέγμα   |
| Icons                   | Εικονίδιο  |
| Input                   | Είσοδος  |
| Journal                 | Εφημερίδα  |
| Label                   | Ετικέτα  |
| Link                    | Σύνδεσμο   |
| List                    | Λίστα  |
| Menu                    | Μενού επιλογών   |
| Model                   | Μοντέλο  |
| Module                  | Ενότητα  |
| Output                  | Έξοδος   |
| Partial                 | Μερικός  |
| Pending                 | Εκκρεμής   |
| Pie chart               | Διάγραμμα πίτας  |
| Plug-in                 | Επέκταση   |
| Poor                    | Κακό   |
| Progress bar            | Γραμμή προόδου   |
| Responsive              | Που ανταποκρίνεται σε (αλλαγές στο μέγεθος της οθόνης) |
| Scope                   | Πεδίο  |
| Site                    | Ιστότοπος  |
| Social Media            | Μέσα κοινωνικής δικτύωσης                              |
| Template                | Υπόδειγμα  |
| Toolbar                 | Γραμμή Εργαλείων                                       |
| Topic                   | Θέμα   |
| User management         | Διαχείριση χρηστών                                     |

|            |                     |
|------------|---------------------|
| Validation | Έλεγχος εγκυρότητας |
| Web        | Διαδίκτυο           |
| Website    | Δικτυακός τόπος     |



## ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

|      |  |
|------|--|
| SSD  | Solid State Drive                          |
| HDD  | Hard Disk Drive                            |
| GB   | Gigabyte                                   |
| J2EE | Java 2 Platform Enterprise Edition         |
| XML  | Extensible Markup Language                 |
| ACID | Atomicity Consistency Isolation Durability |
| IDE  | Integrated Development Environment         |
| HTML | Hypertext Markup Language                  |
| HTTP | Hypertext Transfer Protocol                |
| CSS  | Cascading Style Sheets                     |
| SVG  | Scalable Vector Graphics                   |
| CSV  | Comma Separated Values                     |
| π.χ  | Παραδείγματος χάριν                        |
| κλπ  | Και λοιπά                                  |

## ΠΑΡΑΡΤΗΜΑ Ι: ΚΩΔΙΚΑΣ

Σε αυτήν την παράγραφο θα παρουσιάσουμε κάποια από τα endpoints που προσφέρει το REST που καταναλώνει η πλατφόρμα μας για την επικοινωνία με τη βάση δεδομένων. Στην ενότητα που παρουσιάσαμε την αρχιτεκτονική του συστήματος μας, κάναμε έναν διαχωρισμό του εργαλείου σε δύο κομμάτια, τον πελάτη και τον εξυπηρετητή. Το REST λοιπόν που αναφέρουμε εδώ, είναι αυτός ο εξυπηρετητής. Δημιουργήθηκε για να ικανοποιήσει τους σκοπούς της πλατφόρμας μας, αλλά μπορεί να χρησιμοποιηθεί και αυτόνομα από κάποιον άλλο “πελάτη” (εφαρμογή) που θα χρειάζεται τις υπηρεσίες του συγκεκριμένου εξυπηρετητή για αλληλεπίδραση με την βάση. Αυτός είναι και ο λόγος που αποφασίσαμε να παρουσιάσουμε κάποιες από τις πιο σημαντικές υπηρεσίες που προσφέρει το συγκεκριμένο REST.

### Για τα experiments

#### GET /topic\_modeling\_curation\_api/experimentsInfo

Με αυτό το endpoint μπορούμε να φέρουμε από την βάση δεδομένων όλη την πληροφορία που διαθέτουμε για κάθε experiment. Είναι η πληροφορία που παρουσιάζεται και στην πρώτη οθόνη της εφαρμογής μας, δηλαδή id, ημερομηνία εξαγωγής των δεδομένων από το topic model, metadata, κάποια τεχνικά χαρακτηριστικά και άλλα.

### Για τα topics

- Πληροφορίες γενικά για τα topics

#### GET /topic\_modeling\_curation\_api/topics/{expld}/info

Αν βάλουμε όπου expld το id του experiment που θέλουμε να μελετήσουμε, τότε το παραπάνω endpoint θα μας επιστρέψει μια λίστα με πληροφορίες για κάθε topic του συγκεκριμένου experiment.

- Πληροφορίες για το curation των topics

#### GET /topic\_modeling\_curation\_api/topicCurationDetails

Με το συγκεκριμένο endpoint μπορούμε να συλλέξουμε πληροφορία για το curation των topics. Μέσω της συγκεκριμένης υπηρεσίας λοιπόν μπορούμε να πάρουμε πληροφορία για το id του topic, την ετικέτα (τίτλο) που του χει προσθέσει ο curator, την/τις κατηγορίες που ανήκει όπως και τον curator που καταχώρησε τα συγκεκριμένα δεδομένα.

Επιπλέον, δεδομένου ότι στην εφαρμογή μας επιτρέπουμε την επεξεργασία και την διαγραφή κατηγοριών των topic, υπάρχουν και τα αντίστοιχα endpoints:

POST /topic\_modeling\_curation\_api/createCurationTopicCategory

POST /topic\_modeling\_curation\_api/updateCurationTopicCategory

POST /topic\_modeling\_curation\_api/deleteCurationTopicCategory

POST /topic\_modeling\_curation\_api/deleteAllCurationTopicCategories

Τα συγκεκριμένα endpoints τα χρησιμοποιούμε για την διαχείριση των κατηγοριών και θα μπορούσαν να χρησιμοποιηθούν και στην περίπτωση ενός γράφου με topics, όπου θα αναπαριστούσαμε τις κατηγορίες των topics σε clusters και θα επιτρέπαμε τη δημιουργία, την μετονομασία και τη διαγραφή μιας κατηγορίας.

## - Πληροφορίες για τα documents ενός topic

### GET /topic\_modeling\_curation\_api/topicDocuments

Τέλος, αυτό το endpoint είναι ένα άλλο πολύ σημαντικό endpoint που έχουμε υλοποιήσει στο REST προκειμένου να παίρνουμε πληροφορίες για τα documents ενός topic. Είναι η πληροφορία που παρουσιάζεται στην τελευταία οθόνη της εφαρμογής μας, στον πίνακα με τα documents. Η συγκεκριμένη υπηρεσία λοιπόν μας επιστρέφει στοιχεία για το κάθε document, όπως το id του, τον τίτλο του, την περίληψή του, τον τύπο του, την ημερομηνία έκδοσής του, τον εκδότη του, links με το επίσημο περιεχόμενο του και πολλά άλλα.

Αυτά είναι μόνο κάποια από τα endpoints που επιλέξαμε να αναφέρουμε. Λόγω του μεγάλου πλήθους τους (45, όπως φαίνονται παρακάτω) δεν μπορούμε να τα αναλύσουμε όλα. Για όποιον ενδιαφέρεται για περισσότερη πληροφορία μπορεί να πατήσει στο σύνδεσμο:

<http://194.177.192.206:9080/actuator/mappings>

και να μελετήσει με περισσότερη λεπτομέρεια τις υπηρεσίες που του παρέχονται μέσω του REST μας.

Η λίστα αυτών είναι η ακόλουθη:

- 1.GET /topic\_modeling\_curation\_api/curationTopicsCategories
- 2.POST /topic\_modeling\_curation\_api/deleteAllCurationTopicCategories
- 3.POST /topic\_modeling\_curation\_api/deleteCurationTopicCategory
- 4.POST /topic\_modeling\_curation\_api/updateCurationTopicCategory
- 5.POST /topic\_modeling\_curation\_api/createCurationTopicCategory
- 6.GET /topic\_modeling\_curation\_api/curationTopicCategoryByName
- 7.GET /topic\_modeling\_curation\_api/curationTopicLabelsByTopicId
- 8.POST /topic\_modeling\_curation\_api/deleteAllCurationTopicsLabels
- 9.POST /topic\_modeling\_curation\_api/updateCurationTopicLabelsChanged
- 10.POST /topic\_modeling\_curation\_api/createCurationTopicLabels
- 11.POST /topic\_modeling\_curation\_api/updateCurationTopicLabelsSaved
- 12.GET /topic\_modeling\_curation\_api/curationTopicLabels
- 13.POST /topic\_modeling\_curation\_api/updateCurationTopicLabelsTopicLabel
- 14.POST /topic\_modeling\_curation\_api/updateCurationTopicLabelsCategoryLabel
- 15.GET /topic\_modeling\_curation\_api/documentAbstract
- 16.GET /topic\_modeling\_curation\_api/topicDocuments
- 17.GET /topic\_modeling\_curation\_api/experiments
- 18.GET /topic\_modeling\_curation\_api/experimentsInfo
- 19.POST /topic\_modeling\_curation\_api/createExperimentCurationStatusInfo
- 20.POST /topic\_modeling\_curation\_api/updateExperimentCurationStatusInfo
- 21.GET /topic\_modeling\_curation\_api/experimentCurationStatusInfo
- 22.GET /topic\_modeling\_curation\_api/allExperimentCurationStatusInfo
- 23.GET /topic\_modeling\_curation\_api/allExperimentFinalizedInfo

24. GET /topic\_modeling\_curation\_api/experimentFinalizedInfo
25. POST /topic\_modeling\_curation\_api/createExperimentFinalizedInfo
26. POST /topic\_modeling\_curation\_api/updateExperimentFinalizedInfo
27. GET /topic\_modeling\_curation\_api/topicItems/top10PerModality
28. GET /topic\_modeling\_curation\_api/mergeInfo
29. POST /topic\_modeling\_curation\_api/deleteAllTopicsMergeInfo
30. POST /topic\_modeling\_curation\_api/mergeInfo
31. GET /topic\_modeling\_curation\_api/mergeInfoList
32. GET /topic\_modeling\_curation\_api/topicCurations
33. GET /topic\_modeling\_curation\_api/topicCurationsSession
34. POST /topic\_modeling\_curation\_api/topicCurationsSession
35. DELETE /topic\_modeling\_curation\_api/topicCurationsSessionDeletion
36. POST /topic\_modeling\_curation\_api/deleteAllTopicCurationSession
37. POST /topic\_modeling\_curation\_api/deleteAllTopicCurations
38. GET /topic\_modeling\_curation\_api/topicCurationByTopicId
39. POST /topic\_modeling\_curation\_api/createTopicCuration
40. POST /topic\_modeling\_curation\_api/updateTopicCuration
41. GET /topic\_modeling\_curation\_api/topicCuration
42. GET /topic\_modeling\_curation\_api/topicsCurationDetails
43. GET /topic\_modeling\_curation\_api/topics/{expId}/info
44. GET /topic\_modeling\_curation\_api/topicsInfo
45. GET /topic\_modeling\_curation\_api/topicDocsPerJournal

## ΠΑΡΑΡΤΗΜΑ ΙΙ: ΠΛΗΡΟΦΟΡΙΕΣ ΕΚΤΕΛΕΣΗΣ ΚΑΙ ΛΗΨΗΣ ΤΟΥ ΚΩΔΙΚΑ ΤΗΣ ΠΛΑΤΦΟΡΜΑΣ

Σε αυτήν την παράγραφο θα δώσουμε οδηγίες για το πώς μπορεί ο χρήστης να χρησιμοποιήσει την πλατφόρμα, όπως και για το που θα μπορέσει να βρει τον κώδικα της συγκεκριμένης υλοποίησης.

### **Για την εκτέλεση**

Την πλατφόρμα λοιπόν θα την βρει στην διεύθυνση:

[http://194.177.192.206:9080/topic\\_modeling\\_curation\\_platform/](http://194.177.192.206:9080/topic_modeling_curation_platform/)

Βάζοντας τη συγκεκριμένη διεύθυνση σε κάποιον browser, θα μπορέσει αμέσως να αλληλεπιδράσει με την εφαρμογή μας.

### **Για τον κώδικα της πλατφόρμας**

Ο κώδικας που υλοποιεί το συγκεκριμένο εργαλείο, βρίσκεται στο Github και οι ενδιαφερόμενοι μπορούν να τον κατεβάσουν από τις διευθύνσεις:

- Για τον εξυπηρετητή:

[https://github.com/AntwnisKoulalis/TopicModelingVisualizationTool\\_back.git](https://github.com/AntwnisKoulalis/TopicModelingVisualizationTool_back.git)

- Για τον πελάτη:

<https://github.com/AntwnisKoulalis/TopicModelingVisualizationTool.git>

## ΑΝΑΦΟΡΕΣ

- [1] Yuening Hu, Jordan Boyd-Graber, Brianna Satinoff, and Alison Smith. 2014. Interactive topic modeling. *Machine Learning* 95, 3 (2014), 423–469.W.M. Newman and R.F. Sproull, *Principles of Interactive Computer Graphics*, McGraw-Hill, 1979, p. 402.
- [2] Jaegul Choo, Changhyun Lee, Chandan K Reddy, and Haesun Park. 2013. UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization. *IEEE Transactions of Visualization and Computer Graphics* 19, 12 (2013), 1992–2001.
- [3] Hanseung Lee, Jaeyeon Kihm, Jaegul Choo, John Stasko, and Haesun Park. 2012. iVisClustering: An Interactive Visual Document Clustering via Topic Modeling. *Computer Graphics Forum* 31, 3pt3 (2012), 1155–1164.
- [4] Cai G, Sun F, Sha Y. Interactive visualization for topic model curation. *CEUR Workshop Proceedings*. 2018 Jan 1;2068.
- [5] Wikipedia, «Data Curation» [Online], Available: [http://en.wikipedia.org/wiki/Data\\_curation](http://en.wikipedia.org/wiki/Data_curation)
- [6] Wikipedia, « Topic model» [Online], Available: [http://en.wikipedia.org/wiki/Topic\\_model](http://en.wikipedia.org/wiki/Topic_model)
- [7] Bootstrap, «Home Page»[Online], Available: <http://getbootstrap.com/>
- [8] w3schools, «Bootstrap Tutorial»[Online], Available: <http://www.w3schools.com/bootstrap/default.asp>
- [9] IntelliJ IDEA, «Home Page»[Online], Available: <http://www.jetbrains.com/idea/>
- [10] WebStorm, «Home Page»[Online], Available: <http://www.jetbrains.com/webstorm/>
- [11] Wikipedia, «Java (programming language)» [Online], Available: [http://en.wikipedia.org/wiki/Java\\_\(programming\\_language\)](http://en.wikipedia.org/wiki/Java_(programming_language))
- [12] Wikipedia, «Representational state transfer (REST)» [Online], Available: [http://en.wikipedia.org/wiki/Representational\\_state\\_transfer](http://en.wikipedia.org/wiki/Representational_state_transfer)
- [13] Wikipedia, «Angular (web framework)» [Online], Available: [http://en.wikipedia.org/wiki/Angular\\_\(web\\_framework\)](http://en.wikipedia.org/wiki/Angular_(web_framework))
- [14] Angular, «Home Page»[Online], Available: <http://angular.io/>
- [15] Angular Material, «Home» [Online], Available: <https://material.angular.io/>
- [16] javaTpoint, «Angular Tutorial»[Online], Available: <http://www.javatpoint.com/angular-7-tutorial>
- [17] Wikipedia, «Spring Framework» [Online], Available: [http://en.wikipedia.org/wiki/Spring\\_Framework](http://en.wikipedia.org/wiki/Spring_Framework)
- [18] Spring, «Home Page»[Online], Available: <http://spring.io/>
- [19] Spring Boot, «Spring Boot Tutorial»[Online], Available: <http://spring.io/projects/spring-boot>
- [20] Spring Boot, «Spring Boot Tutorial»[Online], Available: [http://www.tutorialspoint.com/spring\\_boot/index.htm](http://www.tutorialspoint.com/spring_boot/index.htm)
- [21] Wikipedia, «Microservices» [Online], Available: <http://en.wikipedia.org/wiki/Microservices>
- [22] Wikipedia, «Client-server model» [Online], Available: [http://en.wikipedia.org/wiki/Client%E2%80%93server\\_model](http://en.wikipedia.org/wiki/Client%E2%80%93server_model)
- [23] Wikipedia, «Web applications» [Online], Available: [http://en.wikipedia.org/wiki/Web\\_application](http://en.wikipedia.org/wiki/Web_application)
- [24] w3schools, «JavaScript Tutorial»[Online], Available: <http://www.w3schools.com/js/>
- [25] tutorialspoint, «Typescript Tutorial»[Online], Available: <http://www.tutorialspoint.com/typescript/index.htm>
- [26] D3.js, «Home Page»[Online], Available: <http://d3js.org/>
- [27] Intro.js, «Home Page»[Online], Available: <http://introjs.com/>
- [28] PostgreSQL, «Home Page»[Online], Available: <http://www.postgresql.org/>

- [29] Apache Maven, «Home Page»[Online], Available:  
<http://maven.apache.org/>
- [30] Wikipedia, «Data binding» [Online], Available:  
[http://en.wikipedia.org/wiki/Data\\_binding](http://en.wikipedia.org/wiki/Data_binding)
- [31] Wikipedia, «Build automation» [Online], Available:  
[http://en.wikipedia.org/wiki/Build\\_automation](http://en.wikipedia.org/wiki/Build_automation)
- [32] Wikipedia, «Microservices» [Online], Available:  
<http://en.wikipedia.org/wiki/Microservices>
- [33] Wikipedia, «Scripting language» [Online], Available:  
[http://en.wikipedia.org/wiki/Scripting\\_language](http://en.wikipedia.org/wiki/Scripting_language)