



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

**ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ & ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ**

## **Διπλωματική Εργασία**

**Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική**



**RAPIDMINER**

People x Expertise x Data. Boom.

**Φοιτήτρια: Αργυρώ-Αντωνία Καραμητσιάνη**  
**AM: 18387063**

**Επιβλέπουσα Καθηγήτρια**

**Μαρία Ραγκούση**  
**Καθηγήτρια**

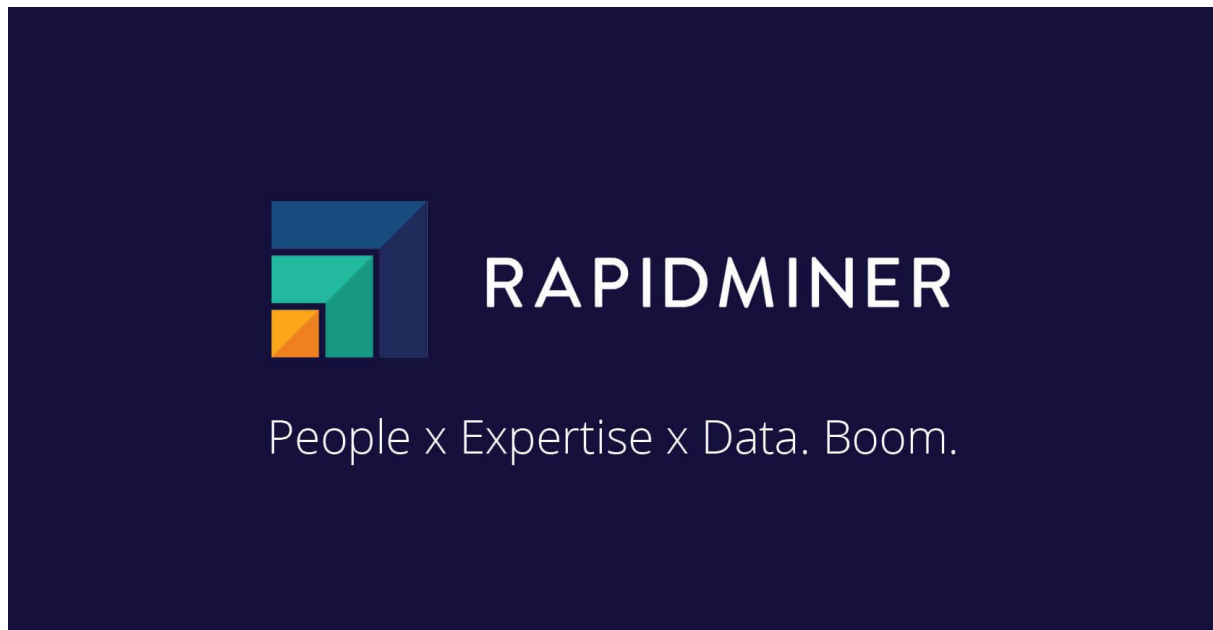
**ΑΘΗΝΑ-ΑΙΓΑΛΕΩ, ΣΕΠΤΕΜΒΡΙΟΣ 2023**



**UNIVERSITY OF WEST ATTICA**  
**FACULTY OF ENGINEERING**  
**DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING**

## **Diploma Thesis**

### **Text Mining applications in Biomedical Engineering**



**Student: Argyro-Antonia Karamitsiani**  
**Registration Number: 18387063**

**Supervisor**

**Maria Rangoussi**  
**Professor**

**ATHENS-EGALEO, SEPTEMBER 2023**

Η Διπλωματική Εργασία έγινε αποδεκτή και βαθμολογήθηκε από την εξής τριμελή επιτροπή:

Μαρία Ραγκούση Καθηγήτρια (επιβλέπουσα)	Χαράλαμπος Πατρικάκης Καθηγητής	Ηλίας Ζώης Αναπλ. Καθηγητής
(Υπογραφή)	(Υπογραφή)	(Υπογραφή)

**Copyright ©** Με επιφύλαξη παντός δικαιώματος. All rights reserved.

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ και Αργυρώ-Αντωνία Καραμητσιάνη,  
Σεπτέμβριος 2023**

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον/την συγγραφέα του και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις θέσεις του επιβλέποντος, της επιτροπής εξέτασης ή τις επίσημες θέσεις του Τμήματος και του Ιδρύματος.

### **ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ**

Η κάτωθι υπογεγραμμένη **Αργυρώ-Αντωνία Καραμητσιάνη** του **Γεωργίου**, με αριθμό μητρώου **18387063** φοιτήτρια του Πανεπιστημίου Δυτικής Αττικής της Σχολής **ΜΗΧΑΝΙΚΩΝ** του Τμήματος **ΗΛΕΚΤΡΟΛΟΓΩΝ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΙΚΩΝ**,

**δηλώνω υπεύθυνα ότι:**

«Είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.»

Η Δήλωση



Αργυρώ-Αντωνία Καραμητσιάνη

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω την καθηγήτριά μου, Μαρία Ραγκούση, για την υπέροχη συνεργασία μας και την απόλυτη βοήθεια και στήριξη που μου προσέφερε. Με την πολύτιμη καθοδήγησή της, ολοκληρώθηκε η διπλωματική μου εργασία και ο ακαδημαϊκός μου κύκλος στο Πανεπιστήμιο Δυτικής Αττικής.

Θέλω να ευχαριστήσω τους γονείς μου, Γιώργο και Ντίνα και την αδερφή μου Παναγιώτα, για την στήριξη και την εμπύχωσή τους όλα αυτά τα χρόνια των σπουδών μου. Χωρίς αυτούς δεν θα είχα καταφέρει όσα έχω καταφέρει μέχρι σήμερα.

Τέλος, θέλω να ευχαριστήσω όλους τους φίλους που έκανα κατά την διάρκεια των σπουδών μου και ειδικότερα την Μαρία-Αμαλία Μεσίρη, που ήμασταν μαζί στις εύκολες και τις δύσκολες στιγμές.

## Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει το πεδίο των εφαρμογών της ανάλυσης κειμένων βιοϊατρικού περιεχομένου (biomedical text mining), που αφορούν στη νόσο Alzheimer. Η εργασία επιχειρεί να αναλύσει επιστημονικά κείμενα (δημοσιευμένες ερευνητικές εργασίες) των οποίων η αναζήτηση και ανάκτηση γίνεται από την βάση PubMed. Αρχικά, δίνεται έμφαση σε κάποιες βασικές ορολογίες με σκοπό την κατανόηση της έννοιας της εξόρυξης δεδομένων και πιο συγκεκριμένα της εξόρυξης κειμένου, ως σύγχρονης τεχνολογίας με ευρύτατες εφαρμογές. Οι έννοιες που αναλύονται είναι η εξόρυξη δεδομένων, η επεξεργασία της φυσικής γλώσσας και η εξόρυξη κειμένου. Ένας από τους κεντρικούς πυλώνες της διπλωματικής εργασίας είναι η κατανόηση και η χρήση του αλγορίθμου Latent Dirichlet Allocation (LDA), ο οποίος παίζει κρίσιμο ρόλο στην ανακάλυψη κρυμμένων αλλά σημαντικών πληροφοριών. Επίσης παρουσιάζεται το κύριο εργαλείο της ανάλυσης, το ανοικτό λογισμικό RapidMiner, το οποίο χρησιμοποιήθηκε για την διεξαγωγή των πειραμάτων. Η δημιουργία και η εκτέλεση μίας πιλοτικής διεργασίας αποτέλεσε το βασικό πρώτο βήμα, το οποίο συνεχίστηκε με μία πιο ολοκληρωμένη πειραματική σχεδίαση και υλοποίηση. Τα (θετικά) αποτελέσματα που λήφθηκαν αποτέλεσαν τη βάση για να επιχειρηθεί μία απάντηση στο κύριο ερώτημα με το οποίο ασχολείται η διπλωματική εργασία.

## Λέξεις – κλειδιά

Βιοϊατρική Μηχανική, Εξαγωγή θέματος, Επεξεργασία φυσικής γλώσσας, Latent Dirichlet Allocation (LDA), RapidMiner, Νόσος Alzheimer.

## **Abstract**

The present diploma thesis examines the field of biomedical text mining applications, focusing on Alzheimer's disease. In this context, the thesis analyzes scientific texts (research publications) retrieved from the PubMed bibliographic database. Initially, some basic terminology is outlined, in order to explain and clarify the concept of data mining and more specifically of text mining. The concepts covered include data mining, natural language processing and text mining. One of the central pillars of the thesis is the in-depth comprehension and correct utilization of the Latent Dirichlet Allocation (LDA) algorithm, which plays a significant role in discovering hidden yet meaningful information. Subsequently, the RapidMiner open software platform is introduced; this is the tool employed for all experiments carried out. The development and execution of a pilot experiment for the verification of the process setup on artificially created, non-biomedical data, was the first step. Subsequently, a more detailed experimental setup was designed and executed on selected biomedical text data, with positive results. These results served as the ground on which an attempt was made to answer the research question posed and investigated in the present diploma thesis.

## **Keywords**

Biomedical Engineering, Topic Extraction, Natural Language Process, Latent Dirichlet Allocation (LDA), RapidMiner, Alzheimer's disease.

## Περιεχόμενα

<b>Κατάλογος Πινάκων.....</b>	<b>10</b>
<b>Κατάλογος Εικόνων .....</b>	<b>10</b>
<b>ΕΙΣΑΓΩΓΗ.....</b>	<b>14</b>
Αντικείμενο της διπλωματικής εργασίας.....	14
Σκοπός και στόχοι .....	14
Μεθοδολογία.....	15
Δομή της διπλωματικής εργασίας.....	15
<b>ΚΕΦΑΛΑΙΟ 1: Τεχνολογίες εξόρυξης κειμένου .....</b>	<b>17</b>
1.1 Γενικά στοιχεία για την Εξόρυξη Δεδομένων (Data Mining) .....	17
1.2 Η επεξεργασία της φυσικής γλώσσας (Natural Language Processing) .....	21
1.3 Η εξόρυξη κειμένου (Text Mining) .....	24
<b>ΚΕΦΑΛΑΙΟ 2: Μέθοδοι και εργαλεία text mining .....</b>	<b>29</b>
2.1 Το λογισμικό RapidMiner .....	29
2.1.1 Εγκατάσταση RapidMiner .....	29
2.1.2 Περιγραφή περιβάλλοντος .....	31
2.1.3 Ανάλυση κατηγοριών χειριστών (operators).....	39
2.2 Προ-επεξεργασία κειμένων .....	39
2.3 Αλγόριθμος Latent Dirichlet Allocation (LDA) για εξαγωγή θέματος (topic extraction) .....	46
2.3.1 Η εξαγωγή θέματος (topic extraction).....	46
2.3.2 Εξήγηση του αλγορίθμου με ένα παράδειγμα .....	52
<b>ΚΕΦΑΛΑΙΟ 3: Βιοϊατρική Μηχανική και ο ρόλος των δημοσιεύσεων .....</b>	<b>54</b>
3.1 Γενική εισαγωγή περί Βιοϊατρικής Μηχανικής.....	54
3.2 Ο ρόλος των δημοσιεύσεων στην προαγωγή της έρευνας.....	55
3.3 Το ερευνητικό ερώτημα της διπλωματικής εργασίας.....	58
<b>ΚΕΦΑΛΑΙΟ 4: Πειραματικό μέρος και αποτελέσματα .....</b>	<b>60</b>
4.1 Ανάπτυξη της δομημένης διεργασίας στο RapidMiner .....	60
4.2 Πιλοτική αξιολόγηση σε τεχνητό σύνολο μη βιοϊατρικών κειμένων .....	65
4.2.1 Πιλοτική αξιολόγηση, k = 2 θέματα .....	66
4.2.2 Πιλοτική αξιολόγηση, k = 3 θέματα .....	68
4.2.3 Πιλοτική αξιολόγηση, k = 4 θέματα .....	70
4.2.4 Πιλοτική αξιολόγηση, σχολιασμός αποτελεσμάτων .....	73
4.3 Εφαρμογή σε περιλήψεις βιοϊατρικών κειμένων και αποτελέσματα .....	77
4.3.1 Αναζήτηση και άντληση βιοϊατρικών κειμένων.....	77
4.3.2 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας άρθρων Clinical Trial I .....	78
4.3.3 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας άρθρων Clinical Trial II .....	84
4.3.4 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας Controlled Clinical Trial .....	92
4.4 Εφαρμογή σε βιοϊατρικά πλήρη κείμενα και αποτελέσματα .....	112
4.4.1 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας Clinical Trial I.....	112
4.4.2 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας Clinical Trial II.....	118
4.4.3 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας Controlled Clinical Trial.....	125
4.5 Σύγκριση αποτελεσμάτων.....	143
4.5.1 Clinical Trial I .....	144
4.5.2 Clinical Trial II .....	150
4.5.3 Controlled Clinical Trial .....	156
<b>ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα – Περαιτέρω έρευνα.....</b>	<b>165</b>



<b>Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές .....</b>	<b>167</b>
<b>ΠΑΡΑΡΤΗΜΑ Ι Άρθρα της ομάδας Clinical Trial I.....</b>	<b>170</b>
<b>ΠΑΡΑΡΤΗΜΑ ΙΙ Άρθρα της ομάδας Clinical Trial ΙΙ.....</b>	<b>173</b>
<b>ΠΑΡΑΡΤΗΜΑ ΙΙΙ Άρθρα της ομάδας Controlled Clinical Trial.....</b>	<b>177</b>

## Κατάλογος Πινάκων

Πίνακας 4.1: Οι παράμετροι των μοντέλων για 3 διαφορετικά εμπειρικά k. ....	73
Πίνακας 4.2: Τιμές παραμέτρων του μοντέλου των πλήρων κειμένων και των περιλήψεών τους, Clinical Trial I (L = Loglikelihood, P = Perplexity). ....	144
Πίνακας 4.3: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο k = 2 (Clinical Trial I). ....	145
Πίνακας 4.4.α: Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεών τους για k = 2, Clinical Trial I. ....	145

## Κατάλογος Εικόνων

Εικόνα 2.1: Η εργαλειοθήκη για την εγκατάσταση του προγράμματος RapidMiner. ....	29
Εικόνα 2.2: Σύνδεση μέσω ατομικού λογαριασμού (account). ....	30
Εικόνα 2.3: Συμπλήρωση στοιχείων για τη δημιουργία ατομικού λογαριασμού. ....	30
Εικόνα 2.4: Επιλογή λειτουργικού συστήματος. ....	31
Εικόνα 2.5: Αρχική οθόνη – περιβάλλον εργασίας RapidMiner. ....	31
Εικόνα 2.6: Η εισαγωγή νέας διεργασίας. ....	32
Εικόνα 2.7: Τα είδη των χειριστών. ....	33
Εικόνα 2.8: Η επιλογή των επεκτάσεων (Extensions) από την εργαλειοθήκη. ....	33
Εικόνα 2.9: Άνοιγμα παραθύρου Marketplace του RapidMiner. ....	34
Εικόνα 2.10: Η λίστα με τις δημοφιλέστερες επεκτάσεις. ....	34
Εικόνα 2.11: Επιλογή πακέτου επέκτασης για εγκατάσταση. ....	35
Εικόνα 2.12: Επιβεβαίωση εγκατάστασης του πακέτου της επέκτασης. ....	36
Εικόνα 2.13: Εγκατάσταση παραπάνω από ένα πακέτο επέκτασης. ....	37
Εικόνα 2.14: Εμφάνιση παραθύρου επανεκκίνησης της εφαρμογής. ....	37
Εικόνα 2.15: Ο πίνακας των παραμέτρων. ....	38
Εικόνα 2.16: Το παράθυρο παροχής βοήθειας. ....	38
Εικόνα 2.17: Παράθυρο έναρξης της εφαρμογής RapidMiner. ....	41
Εικόνα 2.18: Η δημιουργία διεργασίας για προ-επεξεργασία κειμένου. ....	43
Εικόνα 2.19: Παραμετροποίηση του χειριστή Tokenize. ....	43
Εικόνα 2.20: Παραμετροποίηση του χειριστή Stem (Snowball). ....	44
Εικόνα 2.21: Παραμετροποίηση του χειριστή Filter Stopwords (English). ....	44
Εικόνα 2.22: Αποθήκευση διεργασίας με τίτλο “diploma_trial” ....	44
Εικόνα 2.23: Αποτελέσματα διεργασίας. ....	45
Εικόνα 2.24: Γραφική αναπαράσταση του αλγορίθμου LDA. ....	48
Εικόνα 2.25: Υπολογισμός πιθανότητας συνόλου εγγράφων. ....	49
Εικόνα 2.26 : Παράδειγμα γραφικής παράστασης για εύρεση βέλτιστου αριθμού θεμάτων. [Πηγή: <a href="https://community.rapidminer.com/discussion/57377/whats-the-best-way-to-determine-the-number-of-topics-in-the-extract-topics-from-data-lda-operator">https://community.rapidminer.com/discussion/57377/whats-the-best-way-to-determine-the-number-of-topics-in-the-extract-topics-from-data-lda-operator</a> ]. ....	50
Εικόνα 4.1: Η δομημένη διεργασία στο περιβάλλον RapidMiner. ....	61
Εικόνα 4.2: Παραμετροποίηση του χειριστή “Loop Files” ....	62
Εικόνα 4.3: Παραμετροποίηση του χειριστή “Loop Collection” ....	62
Εικόνα 4.4: Παραμετροποίηση του χειριστή “Extract Topics from Documents (LDA)”. ....	63
Εικόνα 4.5: Ο χειριστής “Read Document” ως υπό-διεργασία του χειριστή “Loop Files”. ....	64

Εικόνα 4.6: Οι χειριστές “Tokenize”, “Stem (Snowball)” και “Filter Stopwords (English)” ως υπό-διεργασία του χειριστή “Loop Collection” .....	65
Εικόνα 4.7: Τα κατηγοριοποιημένα κείμενα στον προσωπικό υπολογιστή.....	66
Εικόνα 4.8: Η τιμή Loglikelihood για $k = 2$ θέματα. ....	67
Εικόνα 4.9: Η τιμή perplexity για $k = 2$ θέματα.....	67
Εικόνα 4.10: Το μοντέλο LDA για $k = 2$ θέματα. ....	67
Εικόνα 4.11: Οι 3 πιο συχνές λέξεις και τα βάρη τους για $k = 2$ θέματα.....	68
Εικόνα 4.12: Η κατηγοριοποίηση των κειμένων σε $k = 2$ θέματα. ....	68
Εικόνα 4.13: Η τιμή Loglikelihood για $k = 3$ θέματα. ....	69
Εικόνα 4.14: Η τιμή perplexity για $k = 3$ θέματα.....	69
Εικόνα 4.15: Το μοντέλο LDA για $k = 3$ θέματα. ....	69
Εικόνα 4.16: Οι 3 πιο συχνές λέξεις και τα βάρη τους για $k = 3$ θέματα.....	70
Εικόνα 4.17: Η κατηγοριοποίηση των κειμένων σε $k = 3$ θέματα. ....	70
Εικόνα 4.18: Η τιμή Loglikelihood για $k = 4$ θέματα. ....	71
Εικόνα 4.19: Η τιμή perplexity για $k = 4$ θέματα.....	71
Εικόνα 4.20: Το μοντέλο LDA για $k = 4$ θέματα. ....	72
Εικόνα 4.21: Οι 3 πιο συχνές λέξεις και τα βάρη τους για $k = 4$ θέματα.....	72
Εικόνα 4.22: Η κατηγοριοποίηση των κειμένων σε $k = 4$ θέματα. ....	73
Εικόνα 4.23: Αλλαγή directory για εύρεση αρχείων περιλήψεων των άρθρων της ομάδας Clinical Trial I.....	78
Εικόνα 4.24: Η τιμή Loglikelihood για $k = 2$ θέματα, Clinical Trial I - περιλήψεις.....	79
Εικόνα 4.25: Η τιμή Perplexity για $k = 2$ θέματα, Clinical Trial I – περιλήψεις. ....	79
Εικόνα 4.26: Το μοντέλο LDA για $k = 2$ θέματα, Clinical Trial I – περιλήψεις.....	79
Εικόνα 4.27: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 2$ θέματα, Clinical Trial I – περιλήψεις. ....	80
Εικόνα 4.28: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε $k = 2$ θέματα.....	80
Εικόνα 4.29: Η τιμή Loglikelihood για $k = 3$ θέματα, Clinical Trial I – περιλήψεις.....	81
Εικόνα 4.30: Η τιμή Perplexity για $k = 3$ θέματα, Clinical Trial I – περιλήψεις. ....	81
Εικόνα 4.31: Το μοντέλο LDA για $k = 3$ θέματα, Clinical Trial I – περιλήψεις. ....	81
Εικόνα 4.32: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 3$ θέματα, Clinical Trial I – περιλήψεις. ....	82
Εικόνα 4.33: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε $k = 3$ θέματα.....	82
Εικόνα 4.34: Η τιμή Loglikelihood για $k = 4$ θέματα, Clinical Trial I – περιλήψεις.....	83
Εικόνα 4.35: Η τιμή Perplexity για $k = 4$ θέματα, Clinical Trial I – περιλήψεις. ....	83
Εικόνα 4.36: Το μοντέλο LDA για $k = 4$ θέματα, Clinical Trial I – περιλήψεις.....	83
Εικόνα 4.37: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 4$ θέματα, Clinical Trial I – περιλήψεις. ....	84
Εικόνα 4.38: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε $k = 4$ θέματα.....	84
Εικόνα 4.39: Αλλαγή directory για εύρεση αρχείων περιλήψεων των άρθρων της ομάδας Clinical Trial II.....	85
Εικόνα 4.40: Η τιμή Loglikelihood για $k = 2$ θέματα, Clinical Trial II – περιλήψεις.....	85
Εικόνα 4.41: Η τιμή Perplexity για $k = 2$ θέματα, Clinical Trial II – περιλήψεις. ....	85
Εικόνα 4.42: Το μοντέλο LDA 2 θέματα Clinical Trial II – περιλήψεις.....	86
Εικόνα 4.43: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 2$ θέματα, Clinical Trial II – περιλήψεις. ....	86
Εικόνα 4.44: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε $k = 2$ θέματα.....	87
Εικόνα 4.45: Η τιμή Loglikelihood για $k = 3$ θέματα, Clinical Trial II – περιλήψεις.....	88
Εικόνα 4.46: Η τιμή Perplexity για $k = 3$ θέματα, Clinical Trial II – περιλήψεις.....	88
Εικόνα 4.47: Το μοντέλο LDA για $k = 3$ θέματα, Clinical Trial II – περιλήψεις.....	88

Εικόνα 4.48: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 3$ θέματα, Clinical Trial II – περιλήψεις. ....	89
Εικόνα 4.49: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε $k = 3$ θέματα. ....	89
Εικόνα 4.50: Η τιμή Loglikelihood για $k = 4$ θέματα, Clinical Trial II – περιλήψεις. ....	90
Εικόνα 4.51: Η τιμή Perplexity για $k = 4$ θέματα, Clinical Trial II – περιλήψεις. ....	90
Εικόνα 4.52: Το μοντέλο LDA για $k = 4$ θέματα, Clinical Trial II – περιλήψεις. ....	90
Εικόνα 4.53: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 4$ θέματα, Clinical Trial II – περιλήψεις. ....	91
Εικόνα 4.54: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε $k = 4$ θέματα. ....	91
Εικόνα 4.55: Αλλαγή directory για εύρεση αρχείων περιλήψεων της ομάδας Controlled Clinical Trial .....	92
Εικόνα 4.56: Η τιμή Loglikelihood για $k = 2$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	92
Εικόνα 4.57: Η τιμή Perplexity για $k = 2$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	92
Εικόνα 4.58: Το μοντέλο LDA για $k = 2$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	93
Εικόνα 4.59: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 2$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	93
Εικόνα 4.60: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε $k = 2$ θέματα. ....	99
Εικόνα 4.61: Η τιμή Loglikelihood για $k = 3$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	100
Εικόνα 4.62: Η τιμή Perplexity για $k = 3$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	100
Εικόνα 4.63: Το μοντέλο LDA για $k = 3$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	100
Εικόνα 4.64: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 3$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	101
Εικόνα 4.65: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε $k = 3$ θέματα. ....	106
Εικόνα 4.66: Η τιμή Loglikelihood για $k = 4$ θέματα Controlled Clinical Trial – περιλήψεις. ....	106
Εικόνα 4.67: Η τιμή Perplexity για $k = 4$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	106
Εικόνα 4.68: Το μοντέλο LDA για $k = 4$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	107
Εικόνα 4.69: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 4$ θέματα, Controlled Clinical Trial – περιλήψεις. ....	107
Εικόνα 4.70: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε $k = 4$ θέματα. ....	112
Εικόνα 4.71: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων του Clinical Trial I. ....	113
Εικόνα 4.72: Η τιμή Loglikelihood για $k = 2$ θέματα, Clinical Trial I - πλήρη κείμενα. ....	113
Εικόνα 4.73: Η τιμή Perplexity για $k = 2$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	113
Εικόνα 4.74: Το μοντέλο LDA για $k = 2$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	114
Εικόνα 4.75: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 2$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	114
Εικόνα 4.76: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε $k = 2$ θέματα. ....	114
Εικόνα 4.77: Η τιμή Loglikelihood για $k = 3$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	115
Εικόνα 4.78: Η τιμή Perplexity για $k = 3$ θέματα Clinical Trial I - πλήρη κείμενα. ....	115
Εικόνα 4.79: Το μοντέλο LDA για $k = 3$ θέματα, Clinical Trial – πλήρη κείμενα. ....	115
Εικόνα 4.80: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 3$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	116
Εικόνα 4.81: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε $k = 3$ θέματα. ....	116
Εικόνα 4.82: Η τιμή Loglikelihood για $k = 4$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	116
Εικόνα 4.83: Η τιμή Perplexity για $k = 4$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	117
Εικόνα 4.84: Το μοντέλο LDA για $k = 4$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	117
Εικόνα 4.85: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 4$ θέματα, Clinical Trial I – πλήρη κείμενα. ....	117
Εικόνα 4.86: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε $k = 4$ θέματα. ....	118
Εικόνα 4.87: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων της ομάδας Clinical Trial II. ....	118

Εικόνα 4.88: Η τιμή Loglikelihood για $k = 2$ θέματα, Clinical Trial II – πλήρη κείμενα.....	119
Εικόνα 4.89: Η τιμή Perplexity για $k = 2$ θέματα, Clinical Trial II – πλήρη κείμενα.....	119
Εικόνα 4.90: Το μοντέλο LDA για $k = 2$ θέματα, Clinical Trial II – πλήρη κείμενα.....	119
Εικόνα 4.91: : Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 2$ θέματα, Clinical Trial II – πλήρη κείμενα.....	120
Εικόνα 4.92: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε $k = 2$ θέματα. ....	120
Εικόνα 4.93: Η τιμή Loglikelihood για $k = 3$ θέματα, Clinical Trial II – πλήρη κείμενα.....	121
Εικόνα 4.94: Η τιμή Perplexity για $k = 3$ θέματα, Clinical Trial II – πλήρη κείμενα.....	121
Εικόνα 4.95: Το μοντέλο LDA για $k = 3$ θέματα, Clinical Trial II – πλήρη κείμενα.....	121
Εικόνα 4.96: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 3$ θέματα, Clinical Trial II – πλήρη κείμενα.....	122
Εικόνα 4.97: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε $k = 3$ θέματα. ....	122
Εικόνα 4.98: Η τιμή Loglikelihood για $k = 4$ θέματα, Clinical Trial II – πλήρη κείμενα.....	123
Εικόνα 4.99: Η τιμή Perplexity για $k = 4$ θέματα, Clinical Trial II – πλήρη κείμενα.....	123
Εικόνα 4.100: Το μοντέλο LDA για $k = 4$ θέματα, Clinical Trial II – πλήρη κείμενα.....	123
Εικόνα 4.101: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για $k = 4$ θέματα, Clinical Trial II – πλήρη κείμενα.....	124
Εικόνα 4.102: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε $k = 4$ θέματα. ....	124
Εικόνα 4.103: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων της ομάδας Controlled Clinical Trial. ....	125
Εικόνα 4.104: Η τιμή Loglikelihood για $k = 2$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	125
Εικόνα 4.105: Η τιμή Perplexity για $k = 2$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	126
Εικόνα 4.106: Το μοντέλο LDA για $k = 2$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	126
Εικόνα 4.107: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 2$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	126
Εικόνα 4.108: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε $k = 2$ θέματα.....	131
Εικόνα 4.109: Η τιμή Loglikelihood για $k = 3$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	131
Εικόνα 4.110: Η τιμή Perplexity για $k = 3$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	132
Εικόνα 4.111: Το μοντέλο LDA για $k = 3$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	132
Εικόνα 4.112: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 3$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	132
Εικόνα 4.113: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε $k = 3$ θέματα.....	137
Εικόνα 4.114: Η τιμή Loglikelihood για $k = 4$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	137
Εικόνα 4.115: Η τιμή Perplexity για $k = 4$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	138
Εικόνα 4.116: Το μοντέλο LDA για $k = 4$ θέματα, Controlled Clinical Trial – πλήρη κείμενα. ....	138
Εικόνα 4.117: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για $k = 4$ θέματα, Controlled Clinical Trial – πλήρη κείμενα.....	138
Εικόνα 4.118: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε $k = 4$ θέματα.....	143
Εικόνα 4.119: Ονομασία αρχείου Excel “2 tops_abstract”.....	158
Εικόνα 4.120: Ονομασία αρχείου Excel “2 tops_full text”.....	158

## **ΕΙΣΑΓΩΓΗ**

Το πεδίο της βιοϊατρικής μηχανικής έχει υποστεί τα τελευταία χρόνια μία εντυπωσιακή εξέλιξη και μεταμόρφωση, που οφείλεται κατά το μεγαλύτερο μέρος της στην είσοδο τεχνολογιών που βασίζονται στα δεδομένα. Ανάμεσα σε αυτές τις τεχνολογίες, η ανάλυση κειμένου έχει αναδυθεί ως ένα ισχυρό εργαλείο με την δυνατότητα να επαναπροσδιορίσει τον τρόπο που προσεγγίζεται η υγεία και η ιατρική έρευνα. Μεταξύ άλλων παραδειγμάτων, η μελέτη της νόσου Alzheimer είναι ένα από τα πεδία της ιατρικής το οποίο έχει προσεγγιστεί με την τεχνολογία της ανάλυσης κειμένου. Η νόσος αυτή αποτελεί ένα εκτεταμένο, σοβαρό αλλά και πολύμορφο-πολύπλευρο πρόβλημα, και απαιτεί καινοτόμες λύσεις που υπερβαίνουν την παραδοσιακή ιατρική μεθοδολογία. Η τεχνολογία της ανάλυσης κειμένου (text mining) είναι ένας κλάδος της γενικότερης εξόρυξης δεδομένων (data mining) που προσφέρει ένα ελπιδοφόρο δρόμο για την εξαγωγή αξιόλογων ερευνητικών συμπερασμάτων από τα τεράστια αποθετήρια κειμενικών δεδομένων στο βιοϊατρικό πεδίο γενικότερα και στο πλαίσιο της νόσου Alzheimer ειδικότερα.

### **Αντικείμενο της διπλωματικής εργασίας**

Η διπλωματική εργασία θα προσπαθήσει να απαντήσει το ερευνητικό ερώτημα αν η ανάγνωση περιλήψεων (abstracts) βιοϊατρικών κειμένων παρέχει τις ίδιες βασικές πληροφορίες με την ανάγνωση των πλήρων βιοϊατρικών κειμένων (full texts), προκειμένου οι ερευνητές να επιλέξουν και να κατατάξουν σε θεματολογικές κατηγορίες μία συγκεκριμένη δημοσίευση, για περαιτέρω μελέτη.

Η απάντηση στο ανωτέρω ερώτημα αναμένεται να διευκολύνει το έργο των επιστημόνων, ερευνητών και μελετητών στο αντίστοιχο πεδίο, προσφέροντάς τους οικονομία δυνάμεων και εξοικονόμηση πόρων γενικότερα.

### **Σκοπός και στόχοι**

Το προαναφερόμενο βασικό ερευνητικό ερώτημα που έχει τεθεί προς διερεύνηση, αποτελεί και τον κύριο στόχο της πειραματικής σχεδίασης. Επιπλέον, ο σκοπός αυτός γέννησε και κάποια ακόμη ερωτήματα που μπορούν να θεωρηθούν ως δευτερεύοντες στόχοι της εργασίας, όπως:

- Υπάρχει όριο στο πλήθος ή/και στον όγκο των κειμένων που μπορούν να αναλυθούν;

- Υπάρχει βέλτιστος αριθμός θεμάτων που πρέπει να δίνεται ως είσοδος στον αλγόριθμο – και αν ναι, πώς καθορίζεται;
- Είναι απαραίτητη η πιλοτική αξιολόγηση, π.χ. μέσω μιας demo διεργασίας, ώστε να επαληθευτεί η ορθή λειτουργία μίας αναπτυσσόμενης εφαρμογής λογισμικού;

καθώς επίσης και ορισμένα γενικότερης φύσης ερωτήματα όπως:

- Είναι αναγκαία η γνώση κάποιας γλώσσας προγραμματισμού ώστε να αναπτυχθεί μία λειτουργική εφαρμογή εξαγωγής θεμάτων (topic extraction);
- Μπορεί ένας μηχανικός να ασχοληθεί επιτυχώς με τον κλάδο της βιοϊατρικής χωρίς να έχει βασικές ιατρικές / βιοϊατρικές σπουδές;

## Μεθοδολογία

Η μεθοδολογία που ακολουθήθηκε ήταν πειραματική. Αρχικά, ήταν αναγκαία η μελέτη της σχετικής βιβλιογραφίας για την απόκτηση κάποιων βασικών θεωρητικών γνώσεων ώστε να είναι πιο κατανοητή η περιοχή της εξόρυξης κειμένων, και ειδικότερα η λειτουργία του αλγορίθμου LDA. Στην συνέχεια, απαραίτητο βήμα ήταν η επιλογή εργαλείου και η εξοικείωση με αυτό (RapidMiner). Ακολούθησε το βήμα της σχεδίασης και ανάπτυξης της διεργασίας (process) στο εργαλείο αυτό, η πιλοτική αξιολόγησή της, και τέλος το κυρίως πείραμα για εξαγωγή αποτελεσμάτων.

## Δομή της διπλωματικής εργασίας

Κεφάλαιο 1<sup>ο</sup>: Αναλύονται οι έννοιες εξόρυξη δεδομένων, επεξεργασία της φυσικής γλώσσας και εξόρυξη κειμένου. Δίνονται οι ορισμοί, τα μέσα και τα εργαλεία με τα οποία υλοποιούνται καθώς και τα πλεονεκτήματα και τα μειονεκτήματα.

Κεφάλαιο 2<sup>ο</sup>: Παρουσιάζεται το ανοικτό λογισμικό RapidMiner και ειδικότερα η δομή του και η λειτουργία του. Στην συνέχεια, αναλύεται η έννοια της προ-επεξεργασίας των κειμένων και παρατίθεται ένα παράδειγμα για την καλύτερη κατανόησή της. Τέλος, γίνεται αναφορά στη δομή και τα χαρακτηριστικά του αλγορίθμου LDA.

Κεφάλαιο 3<sup>ο</sup>: Γίνεται εισαγωγή για την βιοϊατρική μηχανική. Ακολούθως, αναφέρεται ο ρόλος των δημοσιεύσεων στην διεξαγωγή των ερευνών και τέλος, περιγράφεται και στοιχειοθετείται το πρόβλημα με το οποίο ασχολείται η διπλωματική εργασία.

Κεφάλαιο 4<sup>ο</sup>: Είναι το κυρίως πειραματικό μέρος της εργασίας. Αρχικά, αναπτύσσεται βήμα-βήμα το σύστημα εξαγωγής θεμάτων εντός του περιβάλλοντος RapidMiner. Στη συνέχεια

επιβεβαιώνεται η ορθή λειτουργία του μέσω μίας πιλοτικής αξιολόγησης με demo διεργασία, στηριγμένη σε τεχνητά, μη βιοϊατρικά σύνολα δεδομένων (κειμένων) και σχολιάζονται τα αποτελέσματά της. Ακολούθως σχεδιάζεται και εκτελείται μία πιο εκτεταμένη πειραματική διαδικασία, με στόχο την εξαγωγή αποτελεσμάτων από πραγματικά βιοϊατρικά κείμενα. Παρατίθενται τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν και στο τέλος συγκρίνονται ώστε να διατυπωθεί μία απάντηση στο ερευνητικό ερώτημα που είχε τεθεί αρχικά.

Κεφάλαιο 5<sup>ο</sup>: Η εργασία κλείνει με τα συμπεράσματα και προτάσεις για περαιτέρω έρευνα.



## ΚΕΦΑΛΑΙΟ 1: Τεχνολογίες εξόρυξης κειμένου

Οι τεχνολογίες εξόρυξης κειμένου αποτελούν ένα σημαντικό και αναπτυσσόμενο κλάδο της επιστήμης της πληροφορικής και της ανάλυσης δεδομένων. Αυτές οι τεχνολογίες επιτρέπουν την αυτόματη εξαγωγή πληροφοριών, μοτίβων και γνώσης από κείμενα γραπτής γλώσσας. Στον σύγχρονο ψηφιακό κόσμο, όπου η πληροφορία πληθαίνει με αστρονομικούς ρυθμούς, οι τεχνολογίες εξόρυξης κειμένου προσφέρουν έναν ισχυρό τρόπο ανάλυσης και αξιοποίησης αυτής της πλούσιας πληροφορίας.

### 1.1 Γενικά στοιχεία για την Εξόρυξη Δεδομένων (Data Mining)

Η εξόρυξη δεδομένων (data mining) ορίζεται ως η διαδικασία κατηγοριοποίησης μεγάλου όγκου δεδομένων, με σκοπό να εντοπιστούν μοτίβα και συσχετίσεις, τα οποία θα αποτελέσουν το βασικό θεμέλιο για την αντιμετώπιση προβλημάτων σε διάφορα πεδία, μέσω της ανάλυσης δεδομένων. Αυτό θα έχει ως αποτέλεσμα την πρόβλεψη σε διάφορους τομείς, όπως αυτός των επιχειρήσεων, την αναζήτηση για μελλοντικές τάσεις και την λήψη πιο ενημερωμένων αποφάσεων. Ο όρος Data Mining χρησιμοποιήθηκε για πρώτη φορά την δεκαετία του 1990 – μια εποχή που ήδη το σύνολο των διαθέσιμων δεδομένων αυξανόταν ραγδαία – και αποτελεί την ένωση των κλάδων της τεχνητής νοημοσύνης, της μηχανικής μάθησης και της στατιστικής. Εφαρμόζεται μεταξύ άλλων στον έλεγχο του πιστωτικού κινδύνου, στην ανίχνευση απάτης και στο φιλτράρισμα ανεπιθύμητων μηνυμάτων κατά τις επαγγελματικές και προσωπικές επικοινωνίες.

Ο μεγάλος όγκος δεδομένων μετατρέπεται σε χρήσιμη πληροφορία με την βοήθεια διάφορων αλγορίθμων και τεχνικών. Τα πιο γνωστά προβλήματα στην περιοχή αυτή είναι:

- **Η ταξινόμηση (Classification):** Η τεχνική αυτή χρησιμοποιεί τις ήδη υπάρχουσες ετικέτες (labels) πάνω στα δεδομένα ώστε να γίνει προσεκτικά η κατηγοριοποίησή τους με βάση τα κοινά τους χαρακτηριστικά. Έτσι, δημιουργείται ένα μοντέλο που μπορεί να προβλέψει σε ποια κατηγορία θα ταξινομηθεί ένα νέο δεδομένο (χωρίς ετικέτα), βάσει των προκαθορισμένων ετικετών.
- **Η ομαδοποίηση (Clustering):** Η διαδικασία αυτή είναι παρόμοια με την ταξινόμηση με την διαφορά ότι το μοντέλο της ομαδοποίησης δεν είναι εκπαιδευμένο να βάζει σε κατηγορίες δεδομένα σύμφωνα με τις ετικέτες αλλά να εντοπίζει μοτίβα ή ομάδες που υπάρχουν στο σύνολο των δεδομένων χωρίς αυτές.

- **Η ανάλυση πρόβλεψης (Prediction):** Η αξιοποίηση πληροφοριών που περιέχουν τον άξονα της ιστορικότητας (ή του χρόνου, π.χ. χρονοσειρές), όπως είναι η θερμοκρασία, οι τιμές μετοχών κ.ά., βοηθάει το μοντέλο της πρόβλεψης, γραμμικής ή μη γραμμικής, να προβλέψει τα προσεχή (μελλοντικά) αποτελέσματα/τιμές/δεδομένα.

Οι πιο γνωστές οικογένειες αλγορίθμων για την εξόρυξη δεδομένων είναι οι ακόλουθες:

- **Οι κανόνες συσχέτισης (Association Rule Mining):** Εξετάζουν τις σχέσεις μεταξύ των μεταβλητών. Έτσι, προστίθεται νέα αξία στα υπάρχοντα δεδομένα καθώς γίνεται η προσπάθεια σύνδεσής τους. Παραδείγματος χάριν, μία εταιρία ερευνά το ιστορικό πωλήσεων με σκοπό να διαπιστώσει ποια προϊόντα πωλούνται μαζί ώστε να μπορέσει να προγραμματίσει, προωθήσει και προβλέψει τις επόμενες πωλήσεις. Για αυτόν τον λόγο ονομάζεται και 'ανάλυση του καλάθιού αγοράς'.
- **Τα δέντρα αποφάσεων (Decision Trees, DT):** Με την βοήθεια ιεραρχικών ερωτήσεων, τα δεδομένα διαδοχικά μοιράζονται σε υποσύνολα και τελικά τοποθετούνται στις κλάσεις που ανήκουν, ανάλογα με τις απαντήσεις που δίνονται.
- **Ο k-κοντινότερος γείτονας (k-Nearest Neighbor, k-NN):** Ο αλγόριθμος αυτός υποθέτει ότι τα δεδομένα που βρίσκονται σε μικρή απόσταση μεταξύ τους, έχουν μεγάλη ομοιότητα και βρίσκονται στην ίδια κατηγορία. Ο k-NN προβλέπει τα χαρακτηριστικά μιας ομάδας βάσει μεμονωμένων σημείων των δεδομένων.
- **Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks, ANN):** Είναι μία εποπτευόμενη τεχνική<sup>1</sup>, όπου γίνεται η χαρτογράφηση των δεδομένων και οι είσοδοι, τα βάρη και η έξοδος αποτελούν τους κόμβους των νευρωνικών δικτύων.

Οι αναλυτές δεδομένων, χρησιμοποιώντας μία από τις τεχνικές που παρουσιάστηκαν, ακολουθούν μία συγκεκριμένη διαδικασία εξόρυξης δεδομένων, η οποία ελαχιστοποιεί την εμφάνιση προβλημάτων κατά την ανάλυση δεδομένων. Η διαδικασία απαρτίζεται από τα εξής βήματα:

1. **Κατανόηση της επιχείρησης:** Είναι αναγκαίο να έχουν κατανοηθεί εξαρχής οι στόχοι και η κατάσταση της επιχείρησης καθώς έτσι θα καθοριστεί η επιτυχία της μεθόδου.
2. **Κατανόηση των δεδομένων:** Αφού έχει καθοριστεί το πρόβλημα, προκύπτουν ερωτήματα σχετικά με την διαθεσιμότητα των πηγών των δεδομένων, την ασφάλειά τους, την αποθήκευσή τους και πως αναμένεται να είναι το τελικό αποτέλεσμα.

---

<sup>1</sup> Ο ορισμός της εποπτευόμενης τεχνικής (supervised learning) παρατίθεται στην ενότητα 2.3.1.

Επιπλέον, θέτονται οι κατάλληλοι περιορισμοί από τους οποίους επηρεάζεται η διαδικασία της εξόρυξης.

3. **Προετοιμασία των δεδομένων:** Τα δεδομένα αφού συλλεχθούν, φορτωθούν και εξαχθούν, υποβάλλονται σε διαδικασία καθαρισμού, αφαίρεσης ακραίων τιμών, εκτίμησης σφαλμάτων και ελέγχου εγκυρότητας. Κατά την διάρκεια αυτού του σταδίου, ελέγχεται το μέγεθος των δεδομένων, καθώς ένα υπέρογκο σύνολο μπορεί να οδηγήσει στην επιβράδυνση των υπολογισμών και της ανάλυσής τους.
4. **Δημιουργία μοντέλου:** Σε αυτό το βήμα, γίνεται χρήση μιας τεχνικής εξόρυξης. Τα «καθαρισμένα» δεδομένα ελέγχονται για τάσεις, συνεχή μοτίβα ή συσχετίσεις. Τα δεδομένα μπορούν να εισαχθούν σε ένα μοντέλο πρόβλεψης ώστε να γίνει η αξιολόγηση προηγούμενων κομματιών πληροφοριών που μπορούν να ερμηνευθούν σε μελλοντικά δεδομένα.
5. **Αξιολόγηση των αποτελεσμάτων:** Η φάση της εξόρυξης δεδομένων που επικεντρώνεται στα δεδομένα, συμπληρώνεται με την αξιολόγηση της απόδοσης του μοντέλου. Τα αναλυόμενα αποτελέσματα συγκεντρώνονται, ερμηνεύονται και παρουσιάζονται στους ειδικούς που έχουν αναλάβει να λαμβάνουν αποφάσεις.
6. **Πιθανές αλλαγές και παρακολούθηση:** Στο τελευταίο στάδιο, επεμβαίνει η διοίκηση που λαμβάνει τα κατάλληλα μέτρα με βάση τα αποτελέσματα. Μετά το πέρας, δημιουργούνται νέοι κύκλοι εξόρυξης δεδομένων, διότι προκύπτουν νέα επιχειρηματικά προβλήματα.

Τα δεδομένα που μπορούν να επεξεργαστούν, καταλαμβάνουν μια ευρεία γκάμα κατηγοριών και χωρίζονται σε διάφορα είδη, ανάλογα με τον τύπο τους, την δομή τους και τον τρόπο επεξεργασίας τους. Κάποια από τα βασικά είδη δεδομένων περιλαμβάνουν:

- **Δεδομένα Αριθμητικών Τιμών (Numerical Data):** Αποτελούνται από αριθμητικές τιμές, που αναπαριστούν ποσότητες, μετρήσεις, θερμοκρασίες, κλπ.
- **Δεδομένα Κατηγορικών Μεταβλητών (Categorical Data):** Περιλαμβάνουν κατηγορίες ή ετικέτες, όπως είναι τα χρώματα, είδη προϊόντων, κωδικοί περιοχής κ.ά.
- **Δεδομένα Χρονοσειρών (Time Series Data):** Αφορούν δεδομένα που σχετίζονται με τον χρόνο, για παράδειγμα οι τιμές των μετοχών σε διάφορα χρονικά σημεία, κλιματολογικά δεδομένα και χρονολογικές σειρές γεγονότων.
- **Δεδομένα Κειμένου (Text Data):** Τα δεδομένα αυτά είναι τα κείμενα και οι λέξεις, που προέρχονται από άρθρα, βιβλία, σχόλια σε μέσα κοινωνικής δικτύωσης, κλπ.

- **Δεδομένα Εικόνων και Βίντεο (Image and Video Data):** Οι εικόνες και τα βίντεο χρησιμοποιούνται σε τομείς όπως είναι η επεξεργασία εικόνας, η αναγνώριση προτύπων και η «μηχανική» όραση (machine vision).
- **Δεδομένα Ήχου (Audio Data):** Στα δεδομένα αυτά περιλαμβάνονται οι ήχοι και τα φωνητικά σήματα που χρησιμοποιούνται σε εφαρμογές αναγνώρισης φωνής και επεξεργασίας ήχου.
- **Δεδομένα Γράφων (Graph Data):** Τα δεδομένα αναπαριστούν τις σχέσεις μεταξύ διάφορων στοιχείων και εντοπίζονται σε διάφορες εφαρμογές, όπως είναι τα κοινωνικά δίκτυα και η δρομολόγηση.
- **Δεδομένα Γεωγραφικής Τοποθεσίας (Geospatial Data):** Περιλαμβάνονται πληροφορίες σχετικά με γεωγραφικές θέσεις όπου η χρήση τους γίνεται σε χαρτογραφικές εφαρμογές και περιβαλλοντικές μελέτες.

Η εξέλιξη και η συνεισφορά του πεδίου της εξόρυξης δεδομένων βελτίωσε την λήψη επιχειρηματικών αποφάσεων και του σχεδιασμού στρατηγικής. Η συγκέντρωση και η ανάλυση των δεδομένων έχουν γίνει πιο αξιόπιστες καθώς ανιχνεύονται τυχόν απάτες και απειλές. Παράλληλα, αναλύοντας τα δεδομένα εντοπίζονται διάφορες σχέσεις και συσχετίσεις κρυμμένων μοτίβων, που θεωρούνται επωφελή για μία εταιρία ή οργανισμό, καθώς η εκμετάλλευση των εξαγόμενων πληροφοριών μπορεί να επιφέρει κέρδη ή άλλα οφέλη. Επιπλέον, κρίνεται απαραίτητη η γρήγορη ανάλυσή τους, η οποία επιτυγχάνεται με την διαδικασία της εξόρυξης.

Από την άλλη πλευρά, η εξόρυξη δεδομένων μπορεί να είναι αρκετά πολύπλοκη. Οι τεχνικοί οφείλουν να έχουν συγκεκριμένες τεχνικές ικανότητες και software εξοπλισμό. Για τις μικρότερες επιχειρήσεις, αυτό αποτελεί ένα δύσκολο εμπόδιο που πρέπει υπερνικήσουν. Όσον αφορά τα αποτελέσματα, η εξόρυξη δεδομένων δεν εγγυάται πάντα την ορθότητά τους. Επηρεάζονται από πολλούς παράγοντες, όπως είναι οι ανακριβείς πηγές των δεδομένων, οι ημερήσιες αλλαγές στην αγορά και τα σφάλματα των μοντέλων. Τέλος, η εξόρυξη είναι μια κοστοβόρα διαδικασία. Οι εταιρίες είναι υποχρεωμένες να πληρώνουν μηνιαία συνδρομή τόσο σε software λογισμικό όσο και σε αποθήκες ανάκτησης δεδομένων. Συμπληρωματικά, η επεξεργασία και ανάλυση είναι υπολογιστικά ακριβείς διαδικασίες, όταν τα σύνολα δεδομένων είναι πολύ μεγάλα.

Εν κατακλείδι, ναι μεν υπάρχουν κάποια μειονεκτήματα στην χρήση της εξόρυξης δεδομένων, όμως αυτά τελικά δεν πρέπει να εμποδίζουν έναν τεχνικό αναλυτή να επεξεργαστεί

δεδομένα. Η εξέλιξη της τεχνολογίας θα καταφέρει να αντιμετωπίσει αυτά τα προβλήματα, καθώς το εργαλείο της εξόρυξης συνεισφέρει στην ανάπτυξη βιώσιμων λύσεων και στρατηγικών στόχων για την μελλοντική δράση μιας εταιρίας. Στην σύγχρονη εποχή, είναι ένας πολύτιμος κλάδος που συνεχώς αναπτύσσεται διότι επιφέρει θετική απόδοση και καινοτομία σε πολλούς τομείς, συμπεριλαμβανομένων και των επιχειρήσεων.

## **1.2 Η επεξεργασία της φυσικής γλώσσας (Natural Language Processing)**

Η επεξεργασία φυσικής γλώσσας (NLP) αντιπροσωπεύει ένα σημαντικό και δυναμικό πεδίο της επιστήμης της τεχνητής νοημοσύνης και της πληροφορικής. Επιτρέπει στις μηχανές (υπολογιστές) να αλληλοεπιδρούν με τους ανθρώπους με πιο φυσικό τρόπο και να κατανοούν τα κείμενα και την ομιλία όπως κάνει ένας άνθρωπος. Η NLP αποτελεί έναν κρίσιμο πυλώνα στην ανάπτυξη εφαρμογών που συμπεριλαμβάνουν την αυτόματη μετάφραση, την αναγνώριση προθέσεων σε φυσική γλώσσα, συστήματα συνομιλίας (chatbots), ανάλυση συναισθημάτων και πολλές άλλες εφαρμογές.

Η ανάπτυξη του τομέα της επεξεργασίας φυσικής γλώσσας ανθίζει από το 1950, οπότε ο Alan Turing δημιούργησε το Turing Test με σκοπό να αποδείξει πόσο έξυπνοι ήταν οι υπολογιστές. Το Turing Test αφορούσε την ικανότητα ενός υπολογιστή να διερμηνεύει αυτόματα και να δημιουργεί λόγο σε φυσική γλώσσα – σε βαθμό ώστε να μην διακρίνεται πρακτικά από τον άνθρωπο-συνομιλητή. Στις δεκαετίες που ακολούθησαν μέχρι το 1990, οι γλωσσολόγοι ανέπτυξαν κανόνες για την επεξεργασία της γλώσσας από τους υπολογιστές, οι οποίοι αποτελούσαν την βάση της NLP. Μετά το 1990, έγινε αντικατάσταση της γλωσσολογικής προσέγγισης της NLP με μία στατιστική κυρίως προσέγγιση, λόγω των εξελίξεων στον χώρο της πληροφορικής που κατέστησαν την στατιστική προσέγγιση πιο αποτελεσματική. Οι υπολογιστές έγιναν ταχύτεροι και η χρήση τους γινόταν για την ανάπτυξη γλωσσικών μοντέλων βασισμένων σε στατιστικά δεδομένα, χωρίς την ανάγκη ανάπτυξης καινούργιων κανόνων από γλωσσολόγους. Έτσι, η στατιστική προσέγγιση μετέτρεψε την επεξεργασία φυσικής γλώσσας από διαδικασία που βασιζόταν σε γλωσσολόγους σε διαδικασία που βασίζεται σε μηχανικούς και στηρίζεται σε μία ευρύτερη ποικιλία επιστημονικών πεδίων. Όμως με το πέρασμα του χρόνου και την εξέλιξη της τεχνολογίας των υπολογιστών, η επεξεργασία φυσικής γλώσσας είναι πλέον η σύνθεση της κλασσικής γλωσσολογίας και των στατιστικών μεθόδων.

Η χρήση της NLP από τους υπολογιστές, αποτελεί το θεμέλιο για την κατανόηση της ανθρώπινης φυσικής γλώσσας. Είτε πρόκειται για προφορική είτε για γραπτή γλώσσα, η NLP

εκμεταλλεύεται την τεχνική νοημοσύνη για να λαμβάνει δεδομένα από τον πραγματικό κόσμο, να τα επεξεργάζεται και να τα κατανοεί με τον τρόπο που μπορεί να κατανοήσει ένας υπολογιστής. Όπως οι άνθρωποι διαθέτουν διάφορους αισθητήρες, τα αφτιά για την ακοή και τα μάτια για την όραση, έτσι και οι υπολογιστές διαθέτουν ειδικό λογισμικό και διατάξεις για την ανάγνωση / εισαγωγή κειμένου και μικρόφωνα για την καταγραφή του ήχου. Κατά την διάρκεια της επεξεργασίας, γίνεται μετατροπή της εισόδου σε κώδικα, που μπορεί να καταλάβει ο υπολογιστής.

Τα δύο βασικά στάδια για την επεξεργασία της φυσικής γλώσσας είναι η προ-επεξεργασία των δεδομένων και η ανάπτυξη ενός αλγορίθμου. Η διαδικασία της προ-επεξεργασίας αποτελείται από την προετοιμασία και τον «καθαρισμό» των δεδομένων ώστε να μπορούν να αναλυθούν από τον υπολογιστή. Αρχικά, τα δεδομένα διασπώνται σε μικρότερα μέρη (*Tokenization*), ύστερα γίνεται αφαίρεση των κοινών λέξεων όπως είναι στα ελληνικά οι λέξεις «να», «το», «και», «δεν», κ.ά., με σκοπό να απομείνουν οι λέξεις που εμπεριέχουν σημαντική πληροφορία (*Stop-word removal*), από τις εναπομείνουσες λέξεις κόβονται οι καταλήξεις και διατηρείται το θέμα μόνο (*Stemming and Lemmatization*) και τέλος, οι λέξεις διαχωρίζονται σε ουσιαστικά, επίθετα και ρήματα (*Part-Of-Speech tagging - POS*). Εν συνεχεία, αφού έχει ολοκληρωθεί η διαδικασία της προ-επεξεργασίας, υιοθετείται και παραμετροποιείται ένας αλγόριθμος που θα επεξεργαστεί τα δεδομένα. Ο πρώτος αλγόριθμος είναι ο **Rules-based system**, κατά τον οποίο αναπτύσσονται γλωσσικοί κανόνες. Ο δεύτερος αλγόριθμος, **Machine learning-based system**, κάνει χρήση των στατιστικών μεθόδων. Εκτελεί διεργασίες βάσει εκπαιδευμένων δεδομένων που έχουν εκχωρηθεί και γίνεται προσαρμογή των μεθόδων όσο τα δεδομένα επεξεργάζονται. Οι αλγόριθμοι της επεξεργασίας φυσικής γλώσσας χρησιμοποιώντας την μηχανική μάθηση σε συνδυασμό με την βαθιά μάθηση και τα νευρωνικά δίκτυα, παράγουν δικούς τους κανόνες μέσω των επαναλήψεων κατά την εκπαίδευσή τους.

Η επεξεργασία φυσικής γλώσσας αξιοποιεί δύο τεχνικές: την συντακτική και την σημασιολογική ανάλυση. Στην πρώτη ανάλυση, οι προτάσεις μετασχηματίζονται για να έχουν ορθή γραμματική διότι η NLP αναζητά το νόημα μέσα σε μία πρόταση, βάσει των κανόνων γραμματικής. Οι συντακτικές τεχνικές είναι οι εξής:

- **Ανάλυση:** Με βάση την γραμματική ανάλυση μιας πρότασης, ο αλγόριθμος NLP καταλαβαίνει τι μέρος του λόγου είναι κάθε λέξη.
- **Τμηματοποίηση λέξεων:** Κατά την ανάλυση ενός κειμένου, ο αλγόριθμος NLP αντιλαμβάνεται ότι οι λέξεις χωρίζονται με ένα άσπρο κενό μεταξύ τους.

- **Διαχωρισμός προτάσεων:** Ο αλγόριθμος NLP αναγνωρίζει ότι ο διαχωρισμός των προτάσεων γίνεται με τελεία, θαυμαστικό ή ερωτηματικό.
- **Μορφολογική κατάτμηση:** Κατά την τεχνική αυτή, οι λέξεις χωρίζονται μορφολογικά, δηλαδή η λέξη αλγόριθμος θα σπάσει στις μορφές ‘αλ’, ‘γο’, ‘ριθ’ και ‘μος’.
- **Στελεχοποίηση:** Αφαιρούνται οι καταλήξεις από τις λέξεις και μένουν τα στελέχη τους.

Στην δεύτερη ανάλυση, η NLP χρησιμοποιεί αλγόριθμους ώστε να κατανοήσει το νόημα και την δομή μιας πρότασης. Οι σημασιολογικές τεχνικές είναι οι ακόλουθες:

- **Αποσαφήνιση σημασίας λέξης:** Με αυτή την τεχνική, ο αλγόριθμος προσπαθεί να εξαγάγει το νόημα της λέξης βάσει του περιεχομένου.
- **Αναγνώριση ονομαστικής οντότητας:** Οι λέξεις ταξινομούνται σε ομάδες.
- **Παραγωγή φυσικής γλώσσας:** Χρησιμοποιείται μία βάση δεδομένων για να αναγνωρίσει την σημασία πίσω από τις λέξεις, ώστε να δημιουργηθεί εξολοκλήρου ένα νέο κείμενο.

Οι υφιστάμενες προσεγγίσεις στην επεξεργασία φυσικής γλώσσας στηρίζονται στην βαθιά μάθηση, μία μορφή τεχνητής νοημοσύνης που εξετάζονται και χρησιμοποιούνται πρότυπα στα δεδομένα, τα οποία βελτιώνουν την κατανόηση ενός υπολογιστικού προγράμματος. Τα μοντέλα αυτά έχουν ανάγκη μεγάλες ποσότητες δεδομένων που ήδη φέρουν χαρακτηρισμό (ετικέτες) για την εκπαίδευση του αλγορίθμου της NLP και την αναγνώριση συσχετίσεων.

Οι αλγόριθμοι της επεξεργασία φυσικής γλώσσας έχουν κάποιες βασικές λειτουργίες:

- **Ταξινόμηση κειμένου:** Στην λειτουργία αυτή, αντιστοιχίζονται ετικέτες σε κείμενα ώστε να κατηγοριοποιηθούν. Έτσι ο αλγόριθμος μπορεί να αναγνωρίσει την σημασία ή το συναίσθημα που κρύβεται στο κείμενο.
- **Εξαγωγή κειμένου:** Με την βοήθεια αυτής της λειτουργίας, πραγματοποιείται η αυτόματη περίληψη ενός κειμένου και η ανακάλυψη σημαντικών δεδομένων. Για παράδειγμα, ανακαλύπτονται λέξεις-κλειδιά, οι οποίες βελτιστοποιούν τις μηχανές αναζήτησης.
- **Μηχανική μετάφραση:** Γίνεται η μετάφραση ενός κειμένου από μία γλώσσα σε μία άλλη με την βοήθεια του υπολογιστή.

- **Παραγωγή φυσικής γλώσσας:** Μη δομημένα δεδομένα αναλύονται και δημιουργείται αυτόματα μία βάση δεδομένων.

Ένα από τα πλεονεκτήματα που επέφερε η ανάπτυξη του πεδίου της επεξεργασίας της φυσικής γλώσσας, είναι η βελτίωση της επικοινωνίας μεταξύ του ανθρώπου και του υπολογιστή. Επιπλέον, τα έγγραφα είναι πιο ακριβή και αποτελεσματικά στην τεκμηρίωση πληροφοριών και δημιουργούνται αυτόματα περιλήψεις από μεγάλα και σύνθετα κείμενα. Οι ψηφιακοί προσωπικοί βοηθοί, όπως είναι το λογισμικό ‘Alexa’, κατανοούν καλύτερα την προφορική γλώσσα καθώς η ανάπτυξη των chatbots έχει βελτιώσει την υποστήριξη πελατών των επιχειρήσεων. Τέλος, είναι πλέον εφικτό να εξάγονται σύνθετες πληροφορίες σε προβλήματα και εφαρμογές όπου προηγουμένως ήταν αδύνατον, εξαιτίας των υπερόγκων συνόλων των δεδομένων.

Παρ’όλ’αυτά, η επεξεργασία της φυσικής γλώσσας έχει και κάποια μειονεκτήματα. Οι υπολογιστές, εξαιτίας της απαιτούμενης ακριβούς και ξεκάθαρης γλώσσας προγραμματισμού τους, αρκετές φορές καθίσταται δύσκολο να κατανοήσουν την ανθρώπινη γλωσσική δομή λόγω της ύπαρξης της αργκό και των διαλέκτων. Επίσης, ο τόνος και η κλίση της φωνής επηρεάζουν αρκετά την επεξεργασία της φυσικής γλώσσας, καθώς ο αλγόριθμος δεν μπορεί να αναγνωρίσει την σημασία του τονισμού των λέξεων. Ένα ακόμη μειονέκτημα για την επεξεργασία της φυσικής γλώσσας είναι το γεγονός ότι η γλώσσα εξελίσσεται. Η αλλαγή του τρόπου χρήσης της γλώσσας μετά από μια περίοδο καταλήγει να μη συνάδει πλέον με τους υπολογιστικούς κανόνες που έχουν δημιουργηθεί, οπότε χρειάζονται νέοι κανόνες.

Η επεξεργασία της φυσικής γλώσσας διαδραματίζει κρίσιμο ρόλο στην τεχνολογία και στην αλληλεπίδραση των ανθρώπων με αυτή, ενώ βρίσκει εφαρμογή στον επιχειρηματικό και καταναλωτικό κλάδο. Παρά τους περιορισμούς που μπορεί να υπάρχουν, είναι πλέον ένα ζωτικής σημασίας εργαλείο, τόσο στην βιομηχανία όσο και στην καθημερινή ζωή.

### **1.3 Η εξόρυξη κειμένου (Text Mining)**

Με την εξέλιξη της τεχνολογίας έχει παρατηρηθεί ο διαρκώς αυξανόμενος όγκος των ηλεκτρονικών δεδομένων. Σε πολλούς τομείς, όπως η ιατρική, η βιομηχανία, η επιστήμη των υπολογιστών, καθίσταται αναγκαία η άντληση σημαντικών πληροφοριών, οι οποίες δύσκολα μπορούν να εντοπιστούν από έναν άνθρωπο-αναγνώστη. Έτσι, αναπτύχθηκε η τεχνολογία Text Mining (εξόρυξη κειμένου), με την βοήθεια της οποίας γίνεται ανάκτηση σημαντικών πληροφοριών, που εντοπίζονται μέσα σε δομημένα ή μη-δομημένα κείμενα, χρησιμοποιώντας υπολογιστικές μεθόδους και τεχνικές. Η εξόρυξη κειμένου αποτελεί απαραίτητο θεμέλιο για



την αξιοποίηση μη ανιχνεύσιμων δεδομένων, καθώς τα κείμενα κατανοούνται καλύτερα, βελτιώνεται η λήψη αποφάσεων και ανακαλύπτονται νέες γνώσεις και πληροφορίες που είναι χρήσιμες σε διάφορους τομείς.

Η επίτευξη της εξαγωγής της εκάστοτε πληροφορίας ακολουθεί κάποιες βασικές αρχές της εξόρυξης κειμένου. Αρχικά, είναι αναγκαίο να γίνει η σωστή προετοιμασία των δεδομένων κατά την οποία τα δεδομένα συλλέγονται, επιλέγονται, κατηγοριοποιούνται και επεξεργάζονται από τα εργαλεία της εξόρυξης κειμένου, στα οποία θα γίνει αναφορά παρακάτω. Στην συνέχεια, η πληροφορία εξάγεται και αναλύεται με την βοήθεια των διαφόρων τεχνικών της εξόρυξης κειμένου. Κατά την διαδικασία της ανάλυσης των δεδομένων γίνεται αναγνώριση θεμάτων, συσχετίσεων και τάσεων.

Όπως αναφέρθηκε παραπάνω, στην διαδικασία της σωστής προετοιμασίας των δεδομένων κρίνεται απαραίτητο να χρησιμοποιηθούν κάποια Text Mining εργαλεία. Ένα από αυτά είναι η διαδικασία μετατροπής κειμένου από ανθρώπινη γλώσσα σε μορφή που μπορεί να αναλυθεί από υπολογιστές, όπως είναι η αναγνώριση κειμένου, η αναγνώριση χαρακτηρισμένων οντοτήτων κ.ά. Ακολούθως, γίνεται ανάλυση της συχνότητας των λέξεων, διαδικασία κατά την οποία εντοπίζονται οι λέξεις που εμφανίζονται με μεγαλύτερη συχνότητα και απομακρύνονται οι άχρηστες λέξεις (π.χ. «και», «να», «θα», κτλ.) που προκαλούν «θόρυβο». Ακόμα δύο χρήσιμα εργαλεία εξόρυξης κειμένου είναι η μέθοδος LSA (Latent Semantic Analysis), βασιζόμενη στον αλγόριθμο SVD (Singular Value Decomposition), που αναλύει το εκάστοτε κείμενο και βρίσκει κρυμμένα σημασιολογικά μοτίβα και η μέθοδος LDA (Latent Dirichlet Allocation), βασιζόμενη στον αλγόριθμο Bayes, που αναγνωρίζει τις κύριες ιδέες ή θέματα (topics) που περιλαμβάνονται στα κείμενα. Τέλος, το εργαλείο NER (Named Entity Recognition) παρέχει την δυνατότητα της τεχνικής αναγνώρισης και κατηγοριοποίησης ονοματισμένων οντοτήτων, όπως ονόματα προσώπων, οργανισμών, τοποθεσιών κ.ά.

Μία από τις βασικές αρχές της εξόρυξης κειμένου είναι η ανάλυση της πληροφορίας με την χρήση κάποιων τεχνικών. Κρίνεται απαραίτητο να σημειωθεί ότι οι τεχνικές Text Mining είναι ποικίλες και συνεχώς εξελίσσονται. Στην παρούσα αναφορά, θα αναλυθούν οι τρεις βασικότερες τεχνικές εκ των οποίων η μία είναι η ανάλυση συχνότητας λέξεων η οποία εμφανίζεται και ως εργαλείο. Στην προκειμένη περίπτωση, η ανάλυση συχνότητας εμφάνισης μιας λέξης σε ένα υπό εξέταση κείμενο υποδηλώνει την σημασία της μέσα στο κείμενο. Εν συνεχεία, η τεχνική της κατηγοριοποίησης (classification) ταξινομεί τα κείμενα σε κατηγορίες σύμφωνα με ένα εκπαιδευμένο μοντέλο που αναγνωρίζει τη σημασία του κάθε κειμένου. Η ομαδοποίηση (clustering) είναι άλλη μία τεχνική Text Mining, κατά την οποία τα κείμενα

ταξινομούνται σε ομάδες με βάση κάποια κοινά χαρακτηριστικά τους. Μέσα από την κατηγοριοποίηση αυτή αναδεικνύεται η σχέση μεταξύ των διαφορετικών κειμένων όπως και κοινά θέματα ή προβλήματα. Η τεχνική αυτή δεν χρησιμοποιεί τη λογική της εκπαίδευσης, ούτε απαιτεί ένα επισημασμένο (labeled) σύνολο δεδομένων (training data set).

Η χρήση αυτού του εξελιγμένου εργαλείου τεχνολογίας είναι αρκετά βοηθητική στους διάφορους τομείς που χρησιμοποιείται, δεν παύει όμως να αντιμετωπίζει κάποια σημαντικά προβλήματα που αφορούν την βέλτιστη και αποτελεσματική χρήση του. Ένα από αυτά αφορά την ποιότητα των δεδομένων, καθώς κάποιες πληροφορίες μπορεί να είναι ασαφείς ή αντιφατικές με αποτέλεσμα να δυσχεραίνεται η ανάλυση ενός κειμένου. Τα στοιχεία «θορύβου», όπως είναι τα ορθογραφικά ή/και τα συντακτικά λάθη, οι σύντομες φράσεις κλπ., και η απουσία κειμενικών δεδομένων αποτελούν δύο επιπλέον προβλήματα ποιότητας των δεδομένων. Εξίσου σημαντικά είναι τα προβλήματα που σχετίζονται με την ανάλυση πολύ μεγάλων ή πολύπλοκων δεδομένων. Πιο συγκεκριμένα, απαιτείται πολύ μεγάλη υπολογιστική ισχύς με αποτέλεσμα να καθίσταται δύσκολη η επεξεργασία των δεδομένων για μεγάλο χρονικό διάστημα.

Γίνεται αντιληπτό ότι η εξόρυξη κειμένου είναι ένα πολύ ισχυρό εργαλείο για την άντληση χρήσιμων πληροφοριών που εντοπίζονται μέσα σε μεγάλα κείμενα. Στον κάθε τομέα ξεχωριστά, αυτό το τεχνολογικό επίτευγμα είναι χρήσιμο στην λήψη αποφάσεων και στην επίλυση διαφόρων προβλημάτων. Οι περιορισμοί που μπορεί να προκύψουν, διορθώνονται με την επαρκή γνώση της τεχνολογίας και την σωστή επιλογή τεχνικών ανάλυσης κειμένου.

Το Text Mining βρίσκει εφαρμογές σε πολλούς τομείς, εκ των οποίων ένας είναι η βιοϊατρική. Ο τομέας αυτός έχει ιδιαίτερες απαιτήσεις που συνδέονται με τα αντικείμενα της ιατρικής και της βιολογίας και για αυτόν τον λόγο προέκυψε η ειδικότερη περιοχή της Εξόρυξης Βιοϊατρικών Κειμένων (Biomedical Text Mining). Πιο αναλυτικά, κατά την εξόρυξη βιοϊατρικού κειμένου εξάγονται πληροφορίες από την ιατρική βιβλιογραφία και τα ηλεκτρονικά αρχεία υγείας με σκοπό την ανάλυση των διαθέσιμων ιατρικών δεδομένων και την ανακάλυψη νέων πληροφοριών και γνώσεων. Οι ιατρικοί επιστήμονες και ερευνητές, μέσω της άντλησης νέων γνώσεων, αποκτούν την δυνατότητα να αναπτύξουν νέες θεραπείες και φάρμακα, βασιζόμενοι σε δεδομένα και πληροφορίες που θα ήταν αδύνατο ή εξαιρετικά κοπώδες να ανακαλυφθούν με άλλους μη αυτοματοποιημένους τρόπους. Επιπλέον, πρέπει να σημειωθεί ότι η χρήση του Biomedical Text Mining συμβάλλει στην βελτίωση τόσο της ποιότητας θεραπείας των ασθενών όσο και της πρόληψης των ασθενειών.

Η εξόρυξη βιοϊατρικού κειμένου συναντάται σε ποικίλες εφαρμογές. Μία από αυτές είναι η αναζήτηση και η ανακάλυψη νέων φαρμάκων διότι αναζητούνται τα συστατικά των φαρμάκων στην ιατρική βιβλιογραφία ενώ παράλληλα ανακαλύπτονται οι νέοι στόχοι των φαρμάκων. Η πρόβλεψη ασθενειών αποτελεί μία ακόμα εφαρμογή του Biomedical Text Mining, όπου μετά την ανάλυση μεγάλου όγκου κλινικών δεδομένων ασθενών, είναι εφικτό να προβλεφθεί ο κίνδυνος εμφάνισης ασθενειών. Σημαντική εφαρμογή που πρέπει να αναφερθεί είναι η αντιμετώπιση του καρκίνου, όπου με την χρήση του Biomedical Text Mining αναγνωρίζονται συγκεκριμένες γενετικές ανωμαλίες ή αλλαγές στις πρωτεΐνες που σχετίζονται με την ασθένεια.

Αντίστοιχα με το Text Mining, υπάρχουν κάποιες βασικές αρχές και στο Biomedical Text Mining. Αρχικά, απαραίτητη κρίνεται η κατανόηση του βιοϊατρικού πεδίου ώστε να γίνεται με σωστό τρόπο η αντιμετώπιση και η επεξεργασία των δεδομένων. Η συλλογή δεδομένων πρέπει να προέρχεται από αξιόπιστες πηγές όπως είναι οι ιατρικές εγγραφές, τα επιστημονικά άρθρα κτλ., ώστε να μπορέσει να γίνει ο σωστός καθαρισμός τους από τα δεδομένα που προκαλούν θόρυβο.

Για την επίτευξη της χρήσης της βιοϊατρικής εξόρυξης κειμένου, πρέπει να γίνει αναφορά στις τεχνικές που χρησιμοποιούνται. Αρχικά, η τεχνική της ανάλυσης δικτύων εξετάζει τις σχέσεις μεταξύ των διαφόρων μορίων, των πρωτεϊνών και των ασθενειών και τις χρησιμοποιεί ώστε να ανακαλυφθούν νέες συσχετίσεις και αλληλεπιδράσεις που θα οδηγήσουν σε νέες θεραπείες και αποτελεσματικότερες διαγνώσεις. Επιπρόσθετα, η κατηγοριοποίηση, η οποία εξυπηρετεί την κατανόηση και την εύρεση πληροφοριών και η ονοματολογία, κατά την οποία γίνεται αναγνώριση μορίων και ασθενειών με βάση τις ειδικές ονομασίες τους, αποτελούν σημαντικές τεχνικές του Biomedical Text Mining.

Παρόλο την μεγάλη χρησιμότητα του Biomedical Text Mining, αυτό το τεχνολογικό εργαλείο αντιμετωπίζει τα ίδια προβλήματα με το γενικότερο πεδίο του Text Mining, αν και με κάποιες μικρές επιπλέον διαφορές. Όσον αφορά την ποιότητα των δεδομένων, είναι γνωστό ότι τα ιατρικά δεδομένα είναι περίπλοκα καθώς περιλαμβάνουν πολλούς και ειδικούς όρους. Έτσι, η ποιότητά τους επηρεάζεται τόσο από τον τρόπο συλλογής τους όσο και καταγραφής τους. Παράλληλα προκύπτει και η πρόκληση του μεγάλου όγκου δεδομένων. Τα δύο αυτά στοιχεία κάνουν αναγκαία τη χρήση εξειδικευμένων αλγορίθμων και τεχνικών επεξεργασίας δεδομένων.

Παρά την ύπαρξη των προβλημάτων και των προκλήσεων, παρατηρείται ότι η τεχνολογία του Biomedical Text Mining ανοίγει νέους ορίζοντες στην ιατρική έρευνα και

πρακτική διότι δίνεται η δυνατότητα εξαγωγής πληροφοριών από μεγάλα σύνολα δεδομένων και ανακάλυψης νέων, χρήσιμων και αξιοποιήσιμων γνώσεων σχετικά με ασθένειες και φάρμακα.

## ΚΕΦΑΛΑΙΟ 2: Μέθοδοι και εργαλεία text mining

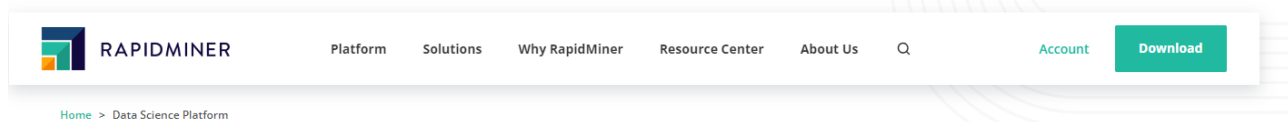
Στο παρόν κεφάλαιο θα παρουσιαστεί αναλυτικά το περιβάλλον της εφαρμογής RapidMiner, ένα ανοικτό λογισμικό που επιλέχθηκε και χρησιμοποιήθηκε για την υλοποίηση των πειραμάτων στην παρούσα εργασία. Επίσης θα γίνει και αναφορά στον τρόπο της απαιτούμενης προ-επεξεργασίας των κειμένων και του αλγορίθμου LDA.

### 2.1 Το λογισμικό RapidMiner

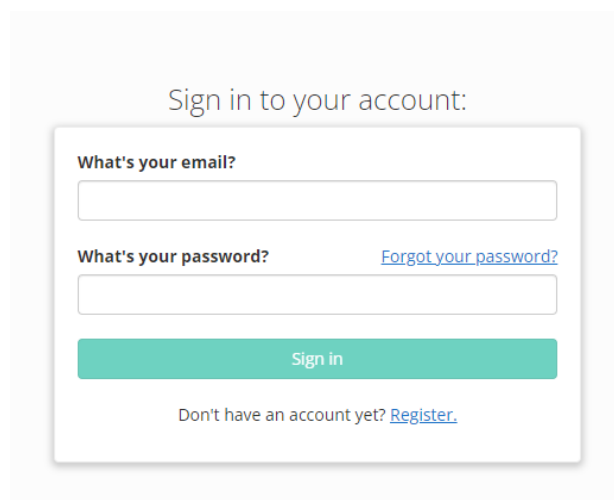
#### 2.1.1 Εγκατάσταση RapidMiner

Το RapidMiner είναι ένα περιβάλλον οπτικού προγραμματισμού, στο οποίο μπορεί εύκολα ο χρήστης να δημιουργήσει μοντέλα και να τα εφαρμόσει σε επιλεγμένα σύνολα δεδομένων ώστε να εξάγει αποτελέσματα και να κάνει προβλέψεις. Άλλες εφαρμογές και περιβάλλοντα λογισμικού και προσομοιώσεων που χρησιμοποιούνται για εξόρυξη και ανάλυση δεδομένων είναι το Matlab, το Apache Spark, το SAS κ.ά. Το πρόγραμμα RapidMiner επιλέχθηκε για την εκπόνηση της παρούσας διπλωματικής εργασίας διότι δεν απαιτεί την γνώση ή τη χρήση κάποιας γλώσσας προγραμματισμού για την ανάπτυξη κώδικα. Είναι ένα ‘οριζόντιο’ και πολύ ευέλικτο εργαλείο, που απευθύνεται τόσο σε αρχάριους όσο και σε προχωρημένους χρήστες, για την ανάλυση και την αξιοποίηση δεδομένων από οποιοδήποτε επιστημονικό πεδίο.

Για να γίνει η εγκατάσταση της εφαρμογής, αρκεί ο χρήστης να αναζητήσει ηλεκτρονικά στον Παγκόσμιο Ιστό τον όρο “Rapid miner download”. Αφού εμφανιστεί η αρχική σελίδα, πατώντας την επιλογή “Download RapidMiner Studio” γίνεται ανακατεύθυνση στις οδηγίες εγκατάστασης. Ακολουθώντας τον σύνδεσμο “RapidMiner website”, ο χρήστης δημιουργεί έναν καινούργιο λογαριασμό (Account), κάνοντας εγγραφή (Register) και συμπληρώνοντας τα στοιχεία του, όπως φαίνεται διαδοχικά στις εικόνες 2.1 – 2.3.



Εικόνα 2.1: Η εργαλειοθήκη για την εγκατάσταση του προγράμματος RapidMiner.



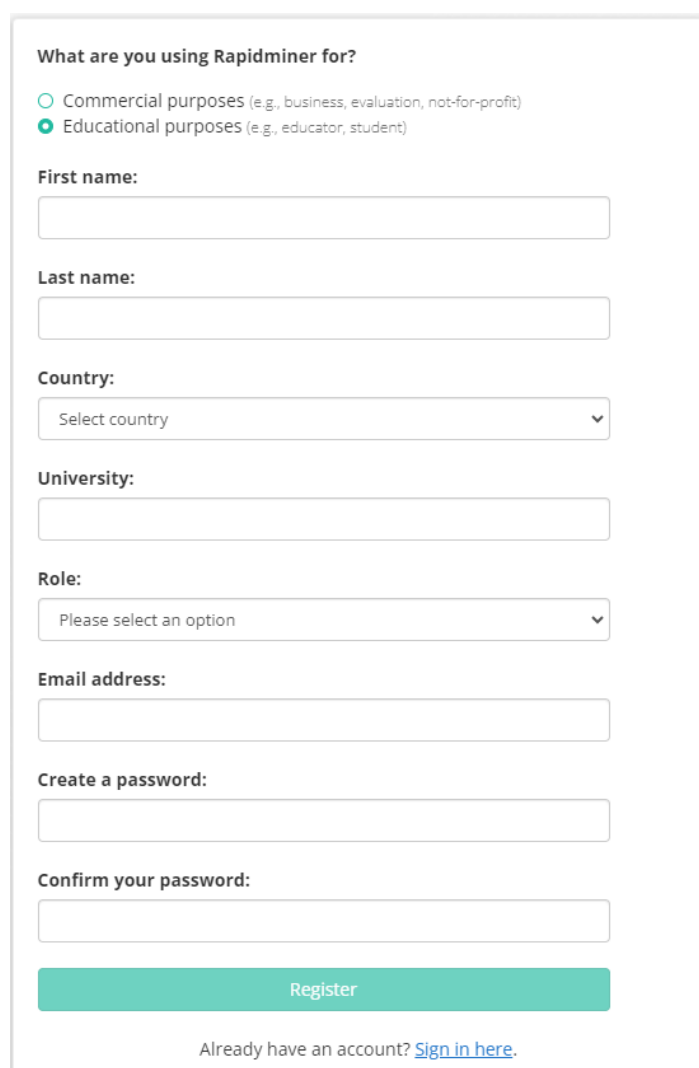
Sign in to your account:

What's your email?

What's your password? [Forgot your password?](#)

Don't have an account yet? [Register.](#)

Εικόνα 2.2: Σύνδεση μέσω ατομικού λογαριασμού (account).



What are you using Rapidminer for?

Commercial purposes (e.g., business, evaluation, not-for-profit)

Educational purposes (e.g., educator, student)

First name:

Last name:

Country:

University:

Role:

Email address:

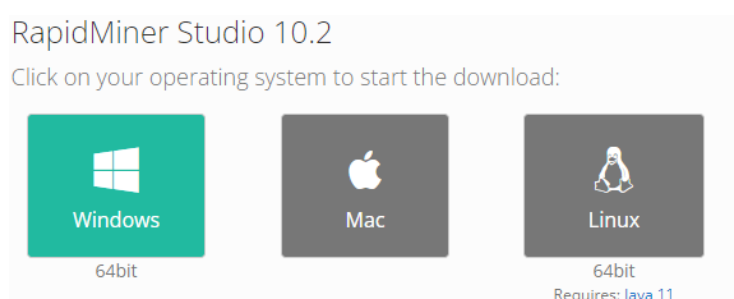
Create a password:

Confirm your password:

Already have an account? [Sign in here.](#)

Εικόνα 2.3: Συμπλήρωση στοιχείων για τη δημιουργία ατομικού λογαριασμού.

Πρέπει να σημειωθεί ότι για να γίνει η εγκατάσταση του RapidMiner Studio στον προσωπικό υπολογιστή του χρήστη, δεν είναι απαραίτητη η σύνδεση στον λογαριασμό, παρά μόνο να επιλεγεί το κατάλληλο λειτουργικό σύστημα (εικόνα 2.4). Τα στοιχεία σύνδεσης θα χρησιμοποιηθούν κατά την είσοδο του χρήστη στην εφαρμογή, μετά την εγκατάσταση.

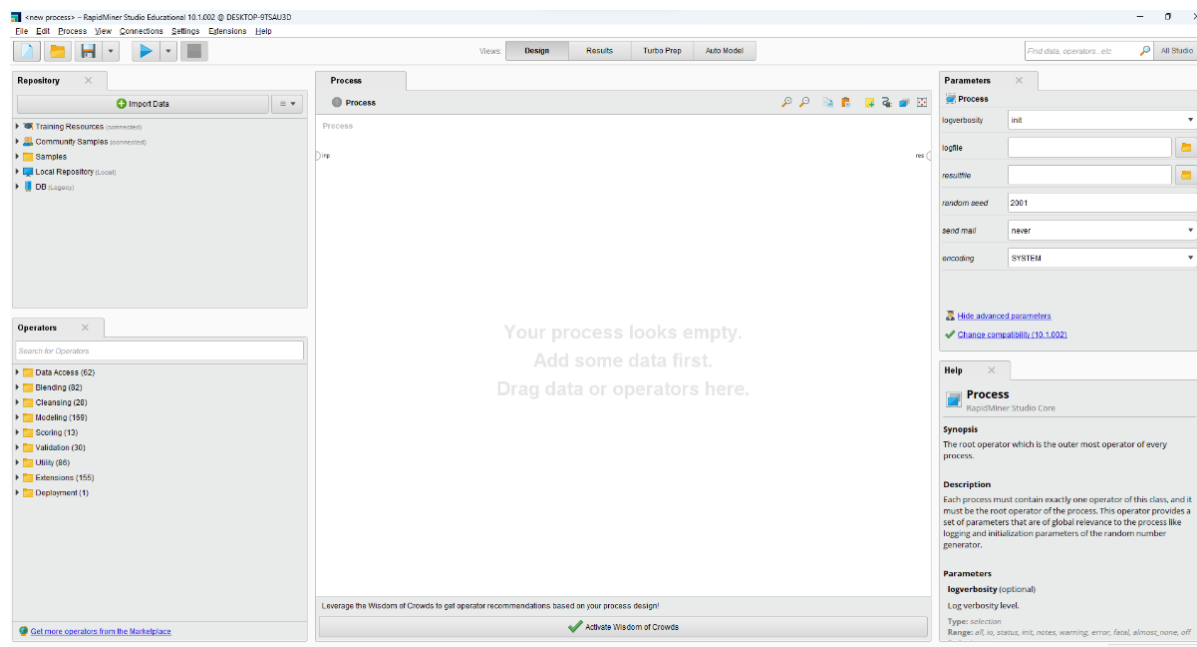


Εικόνα 2.4: Επιλογή λειτουργικού συστήματος.

Ολοκληρώνοντας την διαδικασία της εγκατάστασης, η εφαρμογή ανοίγει και στην οθόνη του υπολογιστή εμφανίζεται η καρτέλα σύνδεσης, στην οποία ο χρήστης θα χρησιμοποιήσει τα στοιχεία σύνδεσης του λογαριασμού που δημιούργησε.

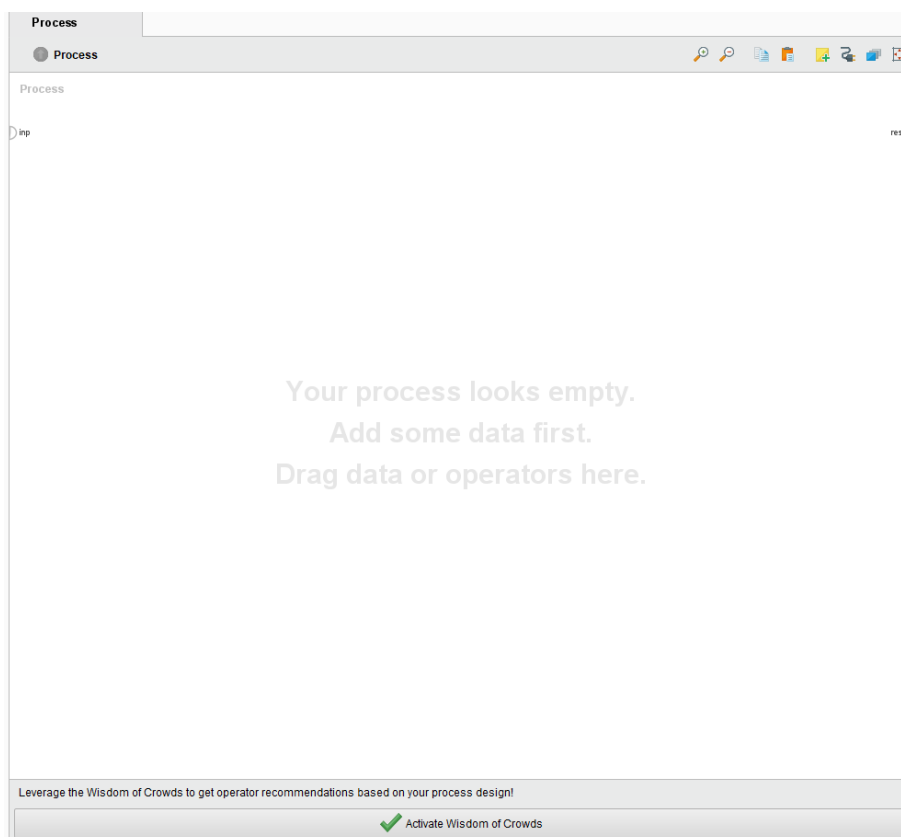
## 2.1.2 Περιγραφή περιβάλλοντος

Μετά την είσοδο στην εφαρμογή, παρουσιάζεται το περιβάλλον εργασίας χρήστη, το οποίο αποτελείται από τις 5 βασικές οθόνες και μία βασική εργαλειοθήκη. (εικόνα 2.5)



Εικόνα 2.5: Αρχική οθόνη – περιβάλλον εργασίας RapidMiner.

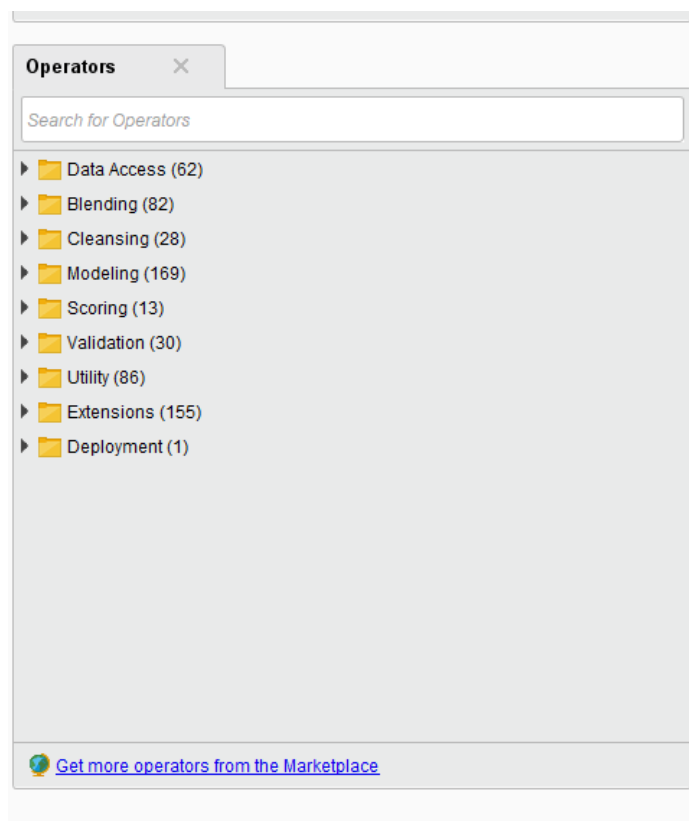
Στο κέντρο της οθόνης βρίσκεται ο πίνακας διεργασιών (processes). Στο σημείο αυτό γίνεται η μεταφορά και η απόθεση (drag and drop) των «χειριστών» (operators), στους οποίους θα γίνει αναφορά στην συνέχεια. Η τοποθέτηση των χειριστών και η σύνδεση τους με την είσοδο (inp) και την έξοδο (res), όπως φαίνεται στην εικόνα 2.6, δημιουργούν την λεγόμενη διεργασία.



Εικόνα 2.6: Η εισαγωγή νέας διεργασίας.

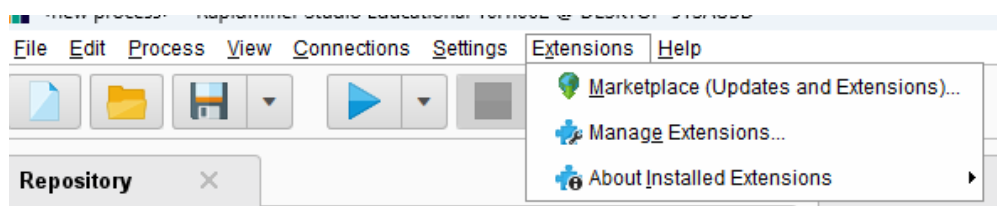
Εν συνεχεία, στην οθόνη παρατηρείται κάτω αριστερά ο πίνακας των «χειριστών» (operators), οι οποίοι αποτελούν τα δομικά στοιχεία μιας διεργασίας, τα οποία είναι αναγκαίο να συνδεθούν μεταξύ τους. Στην εικόνα 2.7, εμφανίζονται τα διάφορα είδη των χειριστών όπως είναι οι “Data Access”, “Blendng”, “Cleansing” κτλ. Κάθε επιλογή έχει μια πληθώρα χειριστών, οι οποίοι είναι απαραίτητοι στις εκάστοτε διεργασίες που δημιουργούνται. Για να εμφανιστεί ένας χειριστής στην επιφάνεια εργασίας αρκεί ο χρήστης να επιλέξει ένα από τα διαθέσιμα είδη ή να αναζητήσει τον επιθυμητό χειριστή και να κάνει μεταφορά και απόθεση στην κεντρική οθόνη.





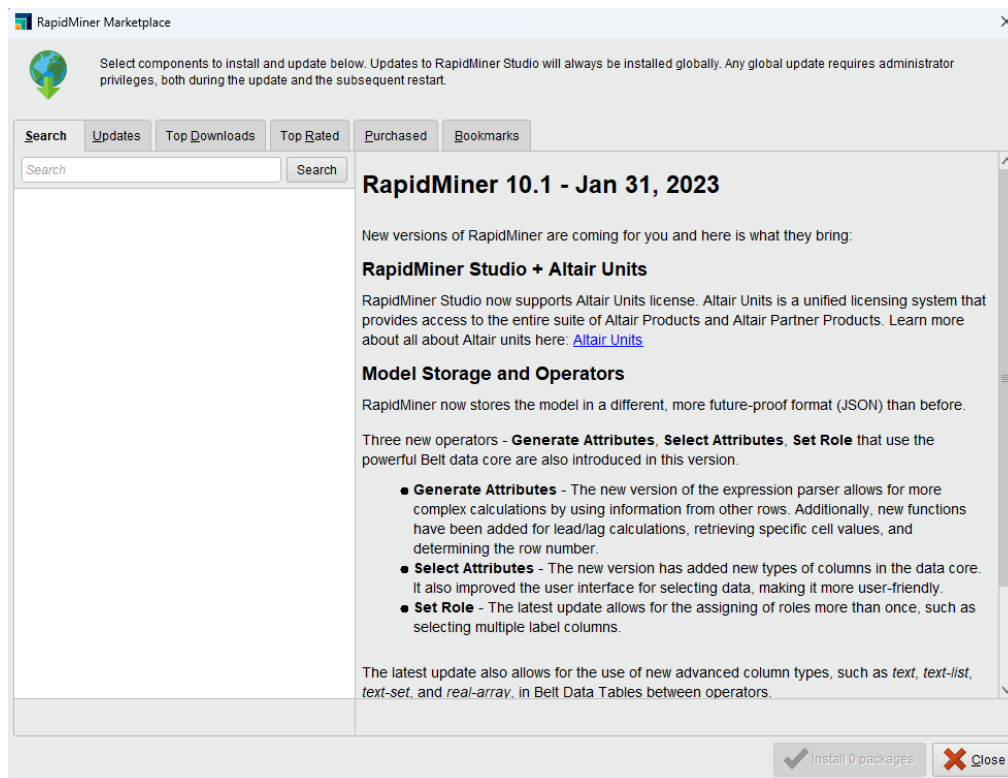
Εικόνα 2.7: Τα είδη των χειριστών.

Όπως φαίνεται στην εικόνα 2.7 ένας από τους χειριστές είναι οι επεκτάσεις (Extensions), οι οποίες είναι διαθέσιμες στο RapidMiner Marketplace. Στην εργαλειοθήκη, πατώντας “Extensions” (εικόνα 2.8) ανοίγει το menu και επιλέγοντας το “Marketplace (Updates and Extensions)” εμφανίζονται όλα οι διαθέσιμες επεκτάσεις.



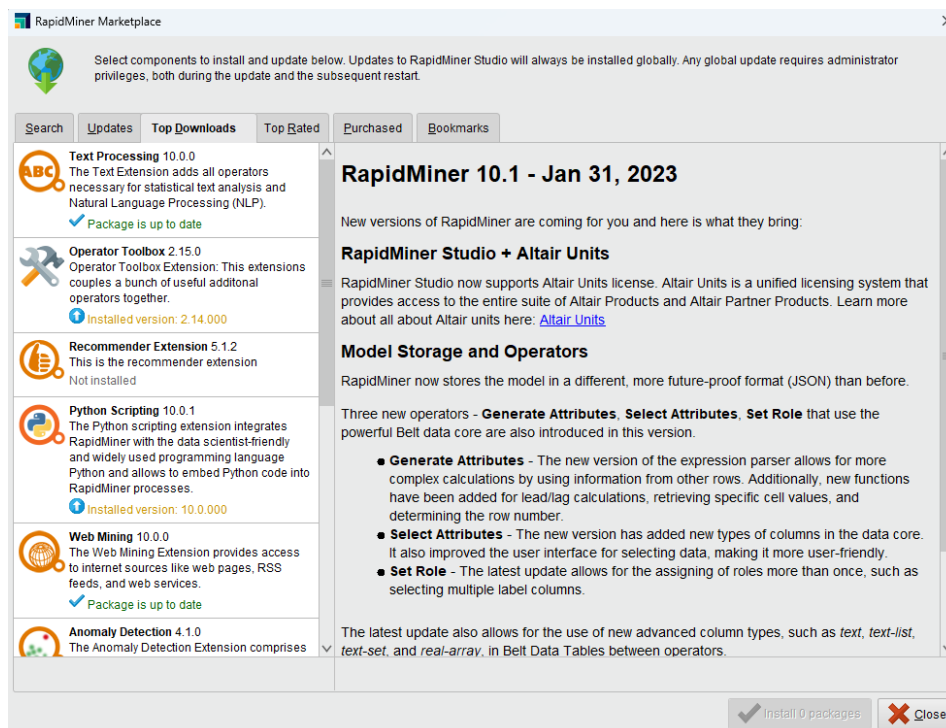
Εικόνα 2.8: Η επιλογή των επεκτάσεων (Extensions) από την εργαλειοθήκη.

Αφού πατηθεί η προαναφερθείσα επιλογή, εμφανίζεται στην οθόνη το ακόλουθο παράθυρο (εικόνα 2.9). Ο χρήστης έχει την δυνατότητα είτε να κάνει αναζήτηση την επέκταση που χρειάζεται είτε να την αναζητήσει στην επιλογή “Top Downloads” (εικόνα 2.10) όπου εμφανίζονται όλες οι διαθέσιμες επεκτάσεις του RapidMiner.



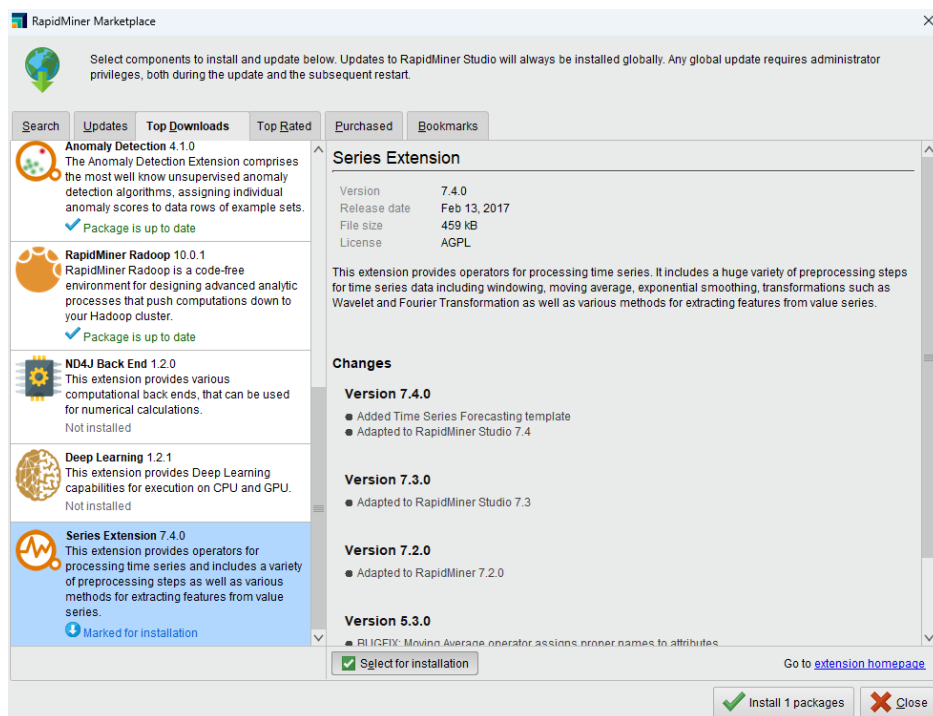
Εικόνα 2.9: Άνοιγμα παραθύρου Marketplace του RapidMiner.

Οι πιο δημοφιλείς και αναγκαίες επεκτάσεις είναι οι Text Processing, Web Mining, Python/R integration, Anomaly Detection, Series extension και τέλος RapidMiner Radoop.



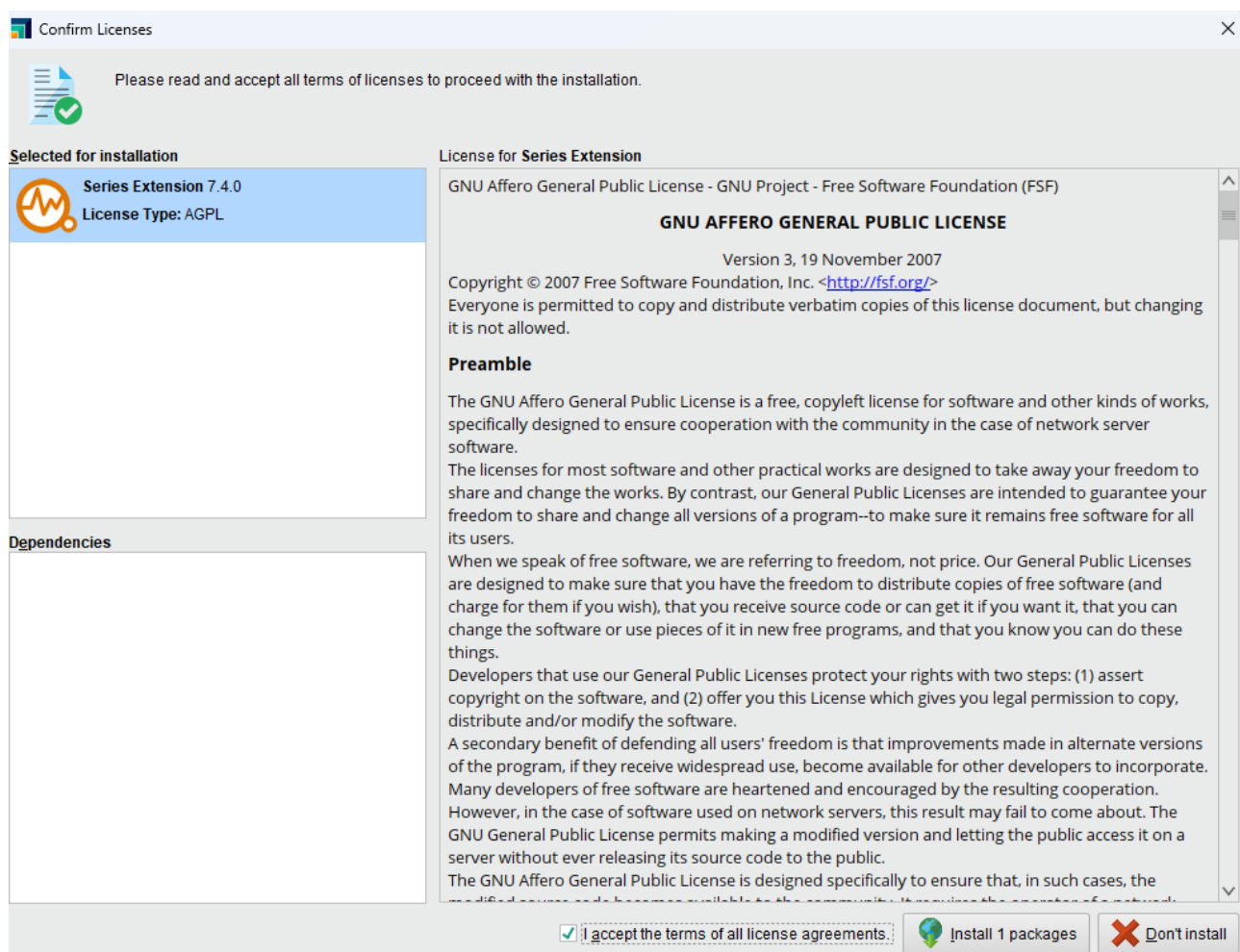
Εικόνα 2.10: Η λίστα με τις δημοφιλέστερες επεκτάσεις.

Στην περίπτωση που ο χρήστης επιθυμεί να εγκαταστήσει ένα πακέτο, αρκεί να αποδεχτεί τους όρους της συμφωνίας άδειας χρήσης. (εικόνα 2.12)



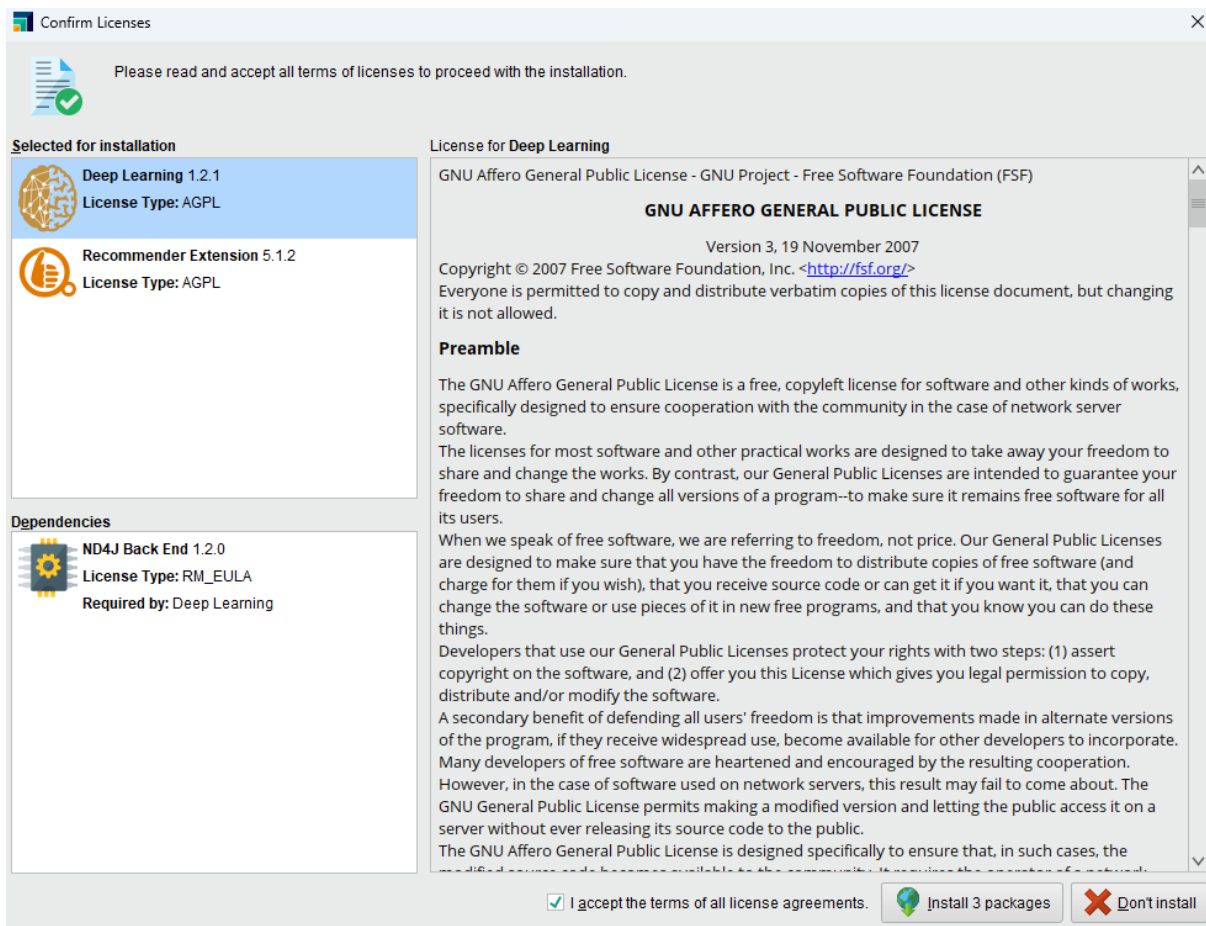
Εικόνα 2.11: Επιλογή πακέτου επέκτασης για εγκατάσταση.

Αντίθετα, αν επιθυμεί να εγκαταστήσει παραπάνω από μία επέκταση, ακολουθεί την παραπάνω διαδικασία με την εξαίρεση ότι δεν πρέπει να πατήσει την επιλογή "Install 1 packages" αλλά να διαλέξει και τις υπόλοιπες επιθυμητές επεκτάσεις. Ύστερα, επιλέγει "Install 2 packages" όπου αναγράφεται και ο συνολικός αριθμός των πακέτων που πρόκειται να εγκατασταθούν.



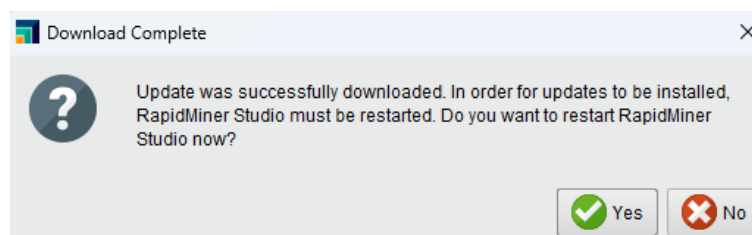
Εικόνα 2.12: Επιβεβαίωση εγκατάστασης του πακέτου της επέκτασης.

Στην εικόνα 2.13, εκτός από τα δύο επιλεγμένα πακέτα, εμφανίζεται και ένα τρίτο, το οποίο είναι αναγκαίο να εγκατασταθεί καθώς ένα από τα παραπάνω εξαρτάται από αυτό. Έτσι, γίνεται αποδοχή των όρων και μετά η επιλογή για την εκκίνηση της διαδικασίας της εγκατάστασης (Install 3 packages).



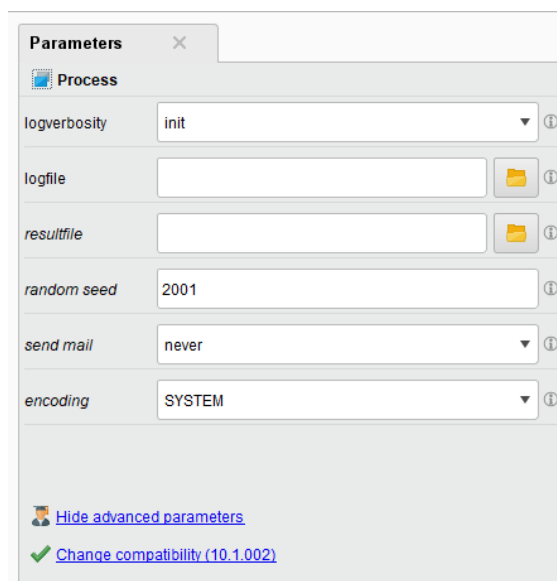
Εικόνα 2.13: Εγκατάσταση παραπάνω από ένα πακέτο επέκτασης.

Αφού ολοκληρωθεί η προαναφερθείσα διαδικασία, κρίνεται απαραίτητη η επανεκκίνηση του προγράμματος ώστε να εμφανιστούν οι εγκατεστημένες επεκτάσεις και να είναι διαθέσιμες προς χρήση. (εικόνα 2.14)



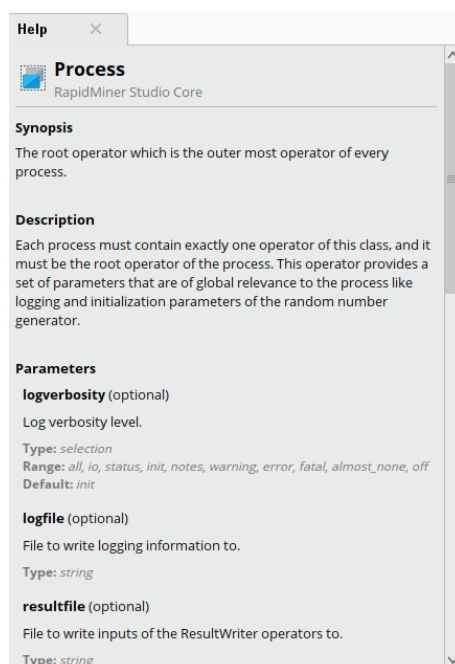
Εικόνα 2.14: Εμφάνιση παραθύρου επανεκκίνησης της εφαρμογής.

Συνεχίζοντας την περιγραφή της αρχικής οθόνης της εφαρμογής, πάνω δεξιά εντοπίζεται ο πίνακας των παραμέτρων (Parameters). Σε αυτό το σημείο ο χρήστης μπορεί να ρυθμίσει τις παραμέτρους του εκάστοτε χειριστή, οι οποίες εμφανίζονται πατώντας πάνω του. (εικόνα 2.15)



Εικόνα 2.15: Ο πίνακας των παραμέτρων.

Τέλος, κάτω αριστερά στην οθόνη συναντάται ο πίνακας με την ονομασία “Help” (εικόνα 2.16). Πατώντας πάνω σε έναν χειριστή, ο χρήστης έχει την δυνατότητα να τον κατανοήσει καλύτερα καθώς στο RapidMiner εξηγείται αναλυτικά ο τρόπος λειτουργίας του, η απαιτούμενη είσοδος και έξοδος και οι παράμετροί του. Επιπρόσθετα, στην πλειονότητα των χειριστών παρέχεται μία φροντιστηριακή διεργασία (Tutorial process), όπου η εφαρμογή έχει έτοιμες διεργασίες με συνδεδεμένους χειριστές για να γίνει καλύτερη κατανόηση του εκάστοτε χειριστή.



Εικόνα 2.16: Το παράθυρο παροχής βοήθειας.

### 2.1.3 Ανάλυση κατηγοριών χειριστών (operators)

Στην εικόνα 2.7 διακρίνονται εννέα (9) κατηγορίες χειριστών, οι οποίες εξηγούνται ακολούθως.

Data Access: Οι χειριστές που περιλαμβάνονται εδώ σχετίζονται με την πρόσβαση, την φόρτωση και την αποθήκευση δεδομένων από και προς διάφορες πηγές δεδομένων.

Blending: Το σύστημα παρέχει την δυνατότητα να συνδυάζει πολλά μοντέλα μηχανικής μάθησης ή αλγορίθμους για να δημιουργήσει ένα μέτα-μοντέλο που συνενώνει τα αποτελέσματά τους.

Cleansing: Οι χειριστές καθαρισμού χρησιμοποιούνται για τον καθαρισμό, τη διόρθωση και την προ-επεξεργασία των δεδομένων πριν από την ανάλυση.

Modeling: Σε αυτή την κατηγορία, οι τελεστές επιλέγουν, εκπαιδεύουν και αξιολογούν μοντέλα ή αλγορίθμους μηχανικής μάθησης.

Scoring: Το εκπαιδευμένο μοντέλο χρησιμοποιείται για την πρόβλεψη τιμών για νέα δεδομένα ή να αξιολογήσουν την απόδοση του μοντέλου.

Validation: Οι χειριστές επικύρωσης ελέγχουν την ακρίβεια και την αξιοπιστία των μοντέλων μηχανικής μάθησης μέσω διάφορων τεχνικών.

Utility: Επιτρέπεται στον χρήστη να προσαρμόσει την λειτουργία του RapidMiner με προσαρμοσμένα σενάρια και επεκτάσεις.

Extensions: Η προσθήκη νέων χειριστών και λειτουργιών στο RapidMiner, επεκτείνει τις δυνατότητες του για εξειδικευμένες ανάγκες.

Deployment: Εξάγονται μοντέλα και αποτελέσματα μηχανικής μάθησης για χρήση σε παραγωγικά συστήματα ή εφαρμογές.

## 2.2 Προ-επεξεργασία κειμένων

Στην παρούσα διπλωματική εργασία καθίσταται αναγκαία η διαδικασία της προ-επεξεργασίας των κειμένων που θα αναλυθούν. Η προ-επεξεργασία κειμένου είναι η προετοιμασία του με σκοπό να συμμετάσχει σε εξόρυξη και ανάλυση των πληροφοριών που εμπεριέχει. Το κείμενο «καθαρίζεται» από άχρηστες πληροφορίες και διαμορφώνεται κατάλληλα ώστε να γίνει κατανοητό και κατάλληλο για ανάλυση από τους αλγορίθμους εξόρυξης πληροφοριών, παίρνει δηλαδή εκείνη τη μορφή που είναι κατάλληλη ώστε να αποτελέσει «δεδομένα εισόδου» (input

data) για τον αντίστοιχο αλγόριθμο. Η προ-επεξεργασία πραγματοποιείται κι αυτή με την βοήθεια του εργαλείου RapidMiner.

Στην συνέχεια παρουσιάζονται 3 από τις πιο βασικές διαδικασίες προ-επεξεργασίας κειμένου εντός του Rapidminer. Χρησιμοποιείται ως παράδειγμα το ακόλουθο κείμενο που επιλέχθηκε τυχαία από το διαδίκτυο και είναι άσχετο με το θέμα της εργασίας:

*“Burning fossil fuels, cutting down forests and farming livestock are increasingly influencing the climate and the earth’s temperature.*

*This adds enormous amounts of greenhouse gases to those naturally occurring in the atmosphere, increasing the greenhouse effect and global warming.*

*2011-2020 was the warmest decade recorded, with global average temperature reaching 1.1°C above pre-industrial levels in 2019. Human-induced global warming is presently increasing at a rate of 0.2°C per decade.*

*An increase of 2°C compared to the temperature in pre-industrial times is associated with serious negative impacts on to the natural environment and human health and wellbeing, including a much higher risk that dangerous and possibly catastrophic changes in the global environment will occur.*

*For this reason, the international community has recognized the need to keep warming well below 2°C and pursue efforts to limit it to 1.5°C.*

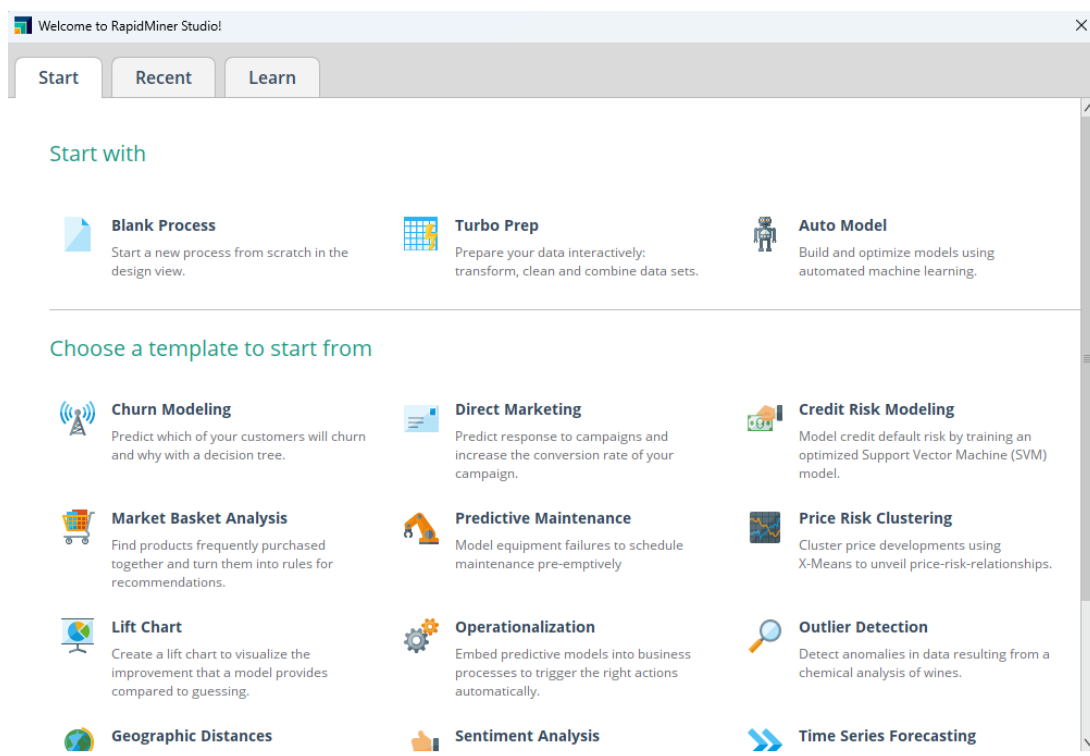
*CO<sub>2</sub> produced by human activities is the largest contributor to global warming. By 2020, its concentration in the atmosphere had risen to 48% above its pre-industrial level (before 1750).*

*Other greenhouse gases are emitted by human activities in smaller quantities. Methane is a more powerful greenhouse gas than CO<sub>2</sub>, but has a shorter atmospheric lifetime. Nitrous oxide, like CO<sub>2</sub>, is a long-lived greenhouse gas that accumulates in the atmosphere over decades to centuries. Non-greenhouse gas pollutants, including aerosols like soot, have different warming and cooling effects and are also associated with other issues such as poor air quality.*

*Natural causes, such as changes in solar radiation or volcanic activity are estimated to have contributed less than plus or minus 0.1°C to total warming between 1890 and 2010.”*

Αρχικά, είναι αναγκαίο να περιγραφεί ο τρόπος δημιουργίας μίας νέας διεργασίας. Όπως φαίνεται στην εικόνα 2.17, κατά το άνοιγμα της εφαρμογής εμφανίζεται το παράθυρο καλωσορίσματος της εφαρμογής, στο οποίο ο χρήστης μπορεί να επιλέξει με τι θέλει να ασχοληθεί.





Εικόνα 2.17: Παράθυρο έναρξης της εφαρμογής RapidMiner.

Πατώντας την επιλογή “Blank Process”, προβάλλεται στην οθόνη η επιφάνεια εργασίας. Έτσι, ο χρήστης διαλέγοντας από το πλαίσιο των Operators, όποιον χειριστή θέλει, δημιουργεί αυτόματα μία διεργασία.

Στην προκειμένη περίπτωση, οι τρεις χειριστές προ-επεξεργασίας κειμένου, που θα χρησιμοποιηθούν, είναι ο Stem (Snowball), ο Tokenize και ο Filter Stopwords (English), οι οποίοι βρίσκονται στην κατηγορία “Extensions” των χειριστών και πιο συγκεκριμένα στην υποκατηγορία “Text Processing”.

- Χειριστής “Tokenize”

Ο συγκεκριμένος χειριστής διαιρεί το προς προ-επεξεργασία κείμενο σε μικρότερες ομάδες, οι οποίες ονομάζονται “tokens”. Τα “tokens” μπορούν να είναι λέξεις, αριθμοί, ακρωνύμια ή οποιαδήποτε άλλη μονάδα του κειμένου που έχει νόημα. Έτσι, ένα κείμενο, το οποίο έχει επεξεργαστεί με τον χειριστή “Tokenize”, δομείται καλύτερα σε βασικές μονάδες που μπορούν να αναλυθούν και να επεξεργαστούν σε μεγαλύτερο βαθμό. Είναι σημαντικό να σημειωθεί ότι ο χειριστής “Tokenize” τοποθετείται πρώτος από τους άλλους χειριστές τεχνικής προ-επεξεργασίας.

Για παράδειγμα, η φράση “It weighs 10 kilos.” χωρίζεται στα ακόλουθα “tokens” : “it”, “weights”, “10”, “kilos”, “.”.

- Χειριστής “Stem (Snowball)”

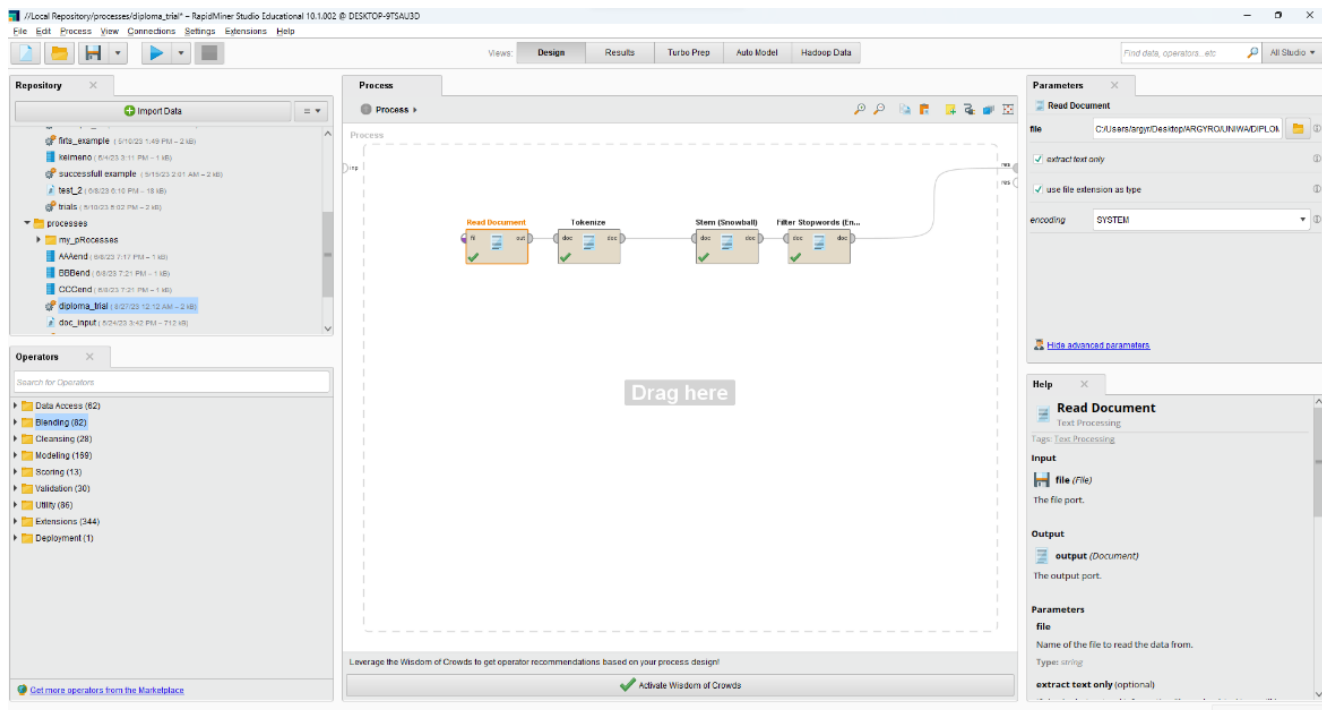
Με την εφαρμογή αυτού του χειριστή, το μέγεθος των λέξεων μειώνεται καθώς αφαιρούνται οι καταλήξεις τους και μένουν μόνο τα «στελέχη». Αυτό θα έχει ως αποτέλεσμα ο αλγόριθμος που θα δημιουργηθεί να αναγνωρίσει τις λέξεις με ίδιο «στέλεχος» ως μία ενιαία ομάδα. Για παράδειγμα, εφαρμόζοντας τον χειριστή “Stem” στις λέξεις “running” και “runs, θα αφαιρέσει τις καταλήξεις και θα κρατήσει το θέμα τους ή «στέλεχος» τους (stem), το οποίο είναι “run”.

- Χειριστής “Filter Stopwords (English)”

Ο χειριστής αυτό αφαιρεί τα λεγόμενα “stop-words” από ένα κείμενο στην αγγλική γλώσσα. Οι λέξεις “stop words” είναι οι κοινές λέξεις “this”, “and”, “for”, “with”, “an” κλπ., οι οποίες δεν προσφέρουν κάποια πληροφορία όσον αφορά το νόημα ή το περιεχόμενο του κειμένου. Η αφαίρεση αυτών των λέξεων, μειώνει σημαντικά το μέγεθος του κειμένου και εξαλείφει κάθε πιθανό θόρυβο που εμποδίζει στην ανάλυση και εξόρυξη πληροφοριών. Ο χειριστής μπορεί να αφαιρέσει τα “stop-words” στις εξής γλώσσες: Αγγλικά, Γερμανικά, Γαλλικά, Τσέχικα και Αραβικά.

Σύμφωνα με το παράδειγμα του χειριστή “Tokenize”, η λέξη “it” είναι μία stop word λέξη και αφαιρείται με την εφαρμογή του χειριστή “Filter Stopwords (English)”.

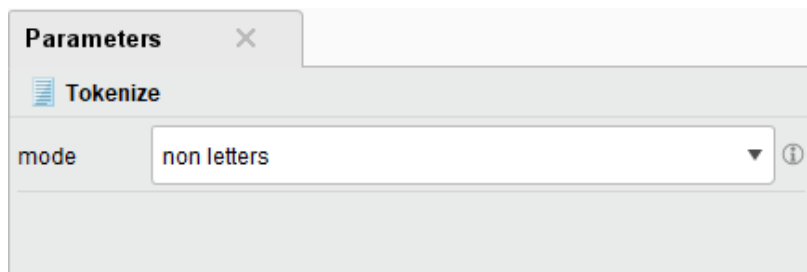
Στο κείμενο που έχει παρατεθεί παραπάνω, θα εφαρμοστούν αυτοί οι τρεις βασικοί χειριστές. Δημιουργώντας μία νέα διεργασία στο RapidMiner και μετά από αναζήτηση στον πίνακα “Operators”, τοποθετούνται οι προαναφερθέντες χειριστές στην επιφάνεια εργασίας. Εκτός από αυτούς τους τρεις χειριστές είναι αναγκαίο να προστεθεί και ο “Read Document”, που ανήκει στην κατηγορία χειριστών που αναφέρθηκε προηγουμένως και έχει ως σκοπό την ανάγνωση, δηλαδή την εισαγωγή, του εν λόγω κειμένου. Στην εικόνα 2.18 φαίνεται η προαναφερθείσα διεργασία.



Εικόνα 2.18: Η δημιουργία διεργασίας για προ-επεξεργασία κειμένου.

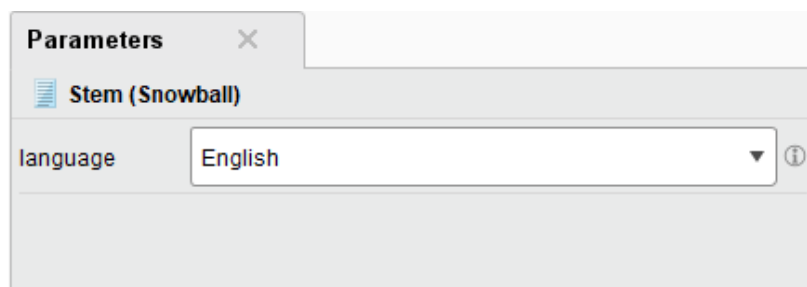
Όπως αναφέρθηκε ο χειριστής “Tokenize” τοποθετήθηκε πρώτος από τους άλλους χειριστές προ-επεξεργασίας κειμένου. Στον πίνακα “Parameters” παρατηρούνται οι παράμετροι του χειριστή “Read Document”, όπου φαίνεται από ποιο αρχείο και από ποιον φάκελο του υπολογιστή έχει ανακτηθεί το κείμενο καθώς έχει επιλεγθεί να αγνοηθούν και να απορριφθούν οι ετικέτες “xml” και “html”. Επίσης, έχει επιλεγθεί ο τύπος των αρχείων να καθορίζεται από τις επεκτάσεις τους.

Στις παραμέτρους του χειριστή “Tokenize” έχει αφαιρεθεί η προκαθορισμένη επιλογή, η οποία δηλώνει ότι ο διαχωρισμός των λέξεων σε “tokens” θα γίνει με βάση τα κενά και τα σημεία στίξης. (εικόνα 2.19)



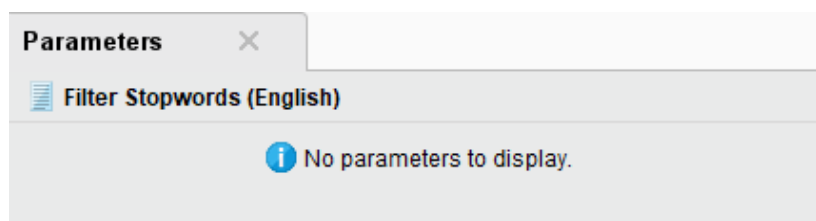
Εικόνα 2.19: Παραμετροποίηση του χειριστή Tokenize

Στις παραμέτρους του χειριστή “Stem (Snowball)” έχει επιλεγεί ως γλώσσα η αγγλική διότι το κείμενο είναι στα αγγλικά. (εικόνα 2.20)



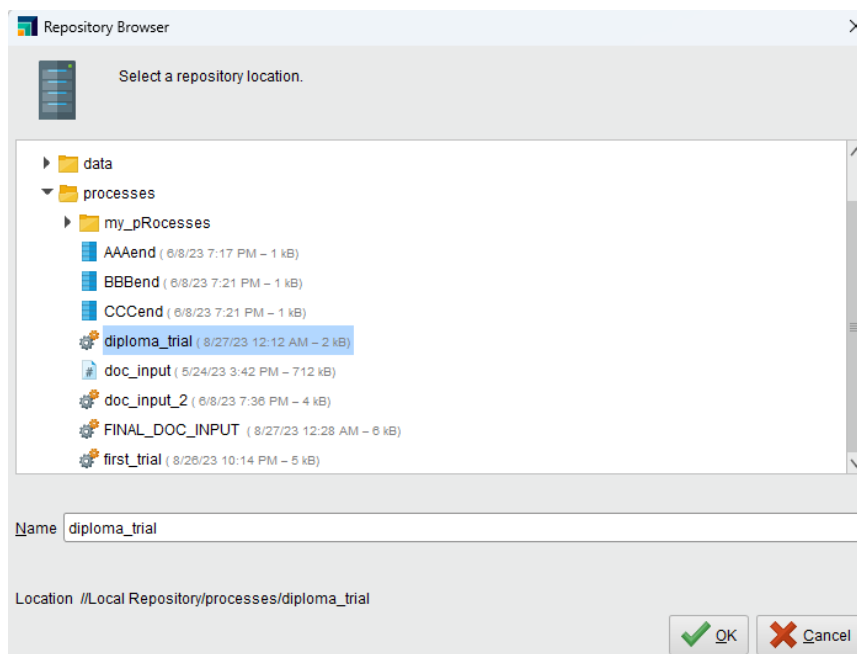
Εικόνα 2.20: Παραμετροποίηση του χειριστή Stem (Snowball).

Ο χειριστής “Filter Stopwords (English)” δεν έχει παραμέτρους. (εικόνα 2.21)



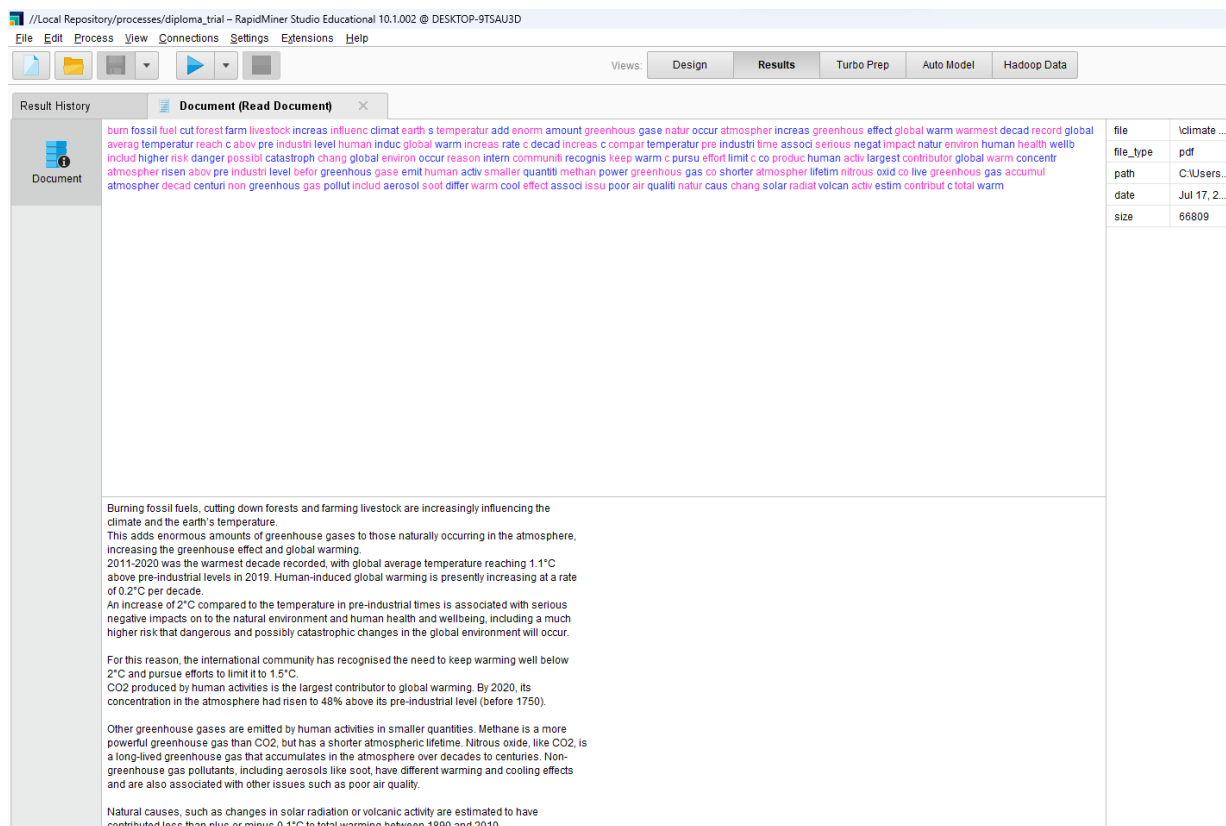
Εικόνα 2.21: : Παραμετροποίηση του χειριστή Filter Stopwords (English).

Ολοκληρώνοντας και την ρύθμιση των παραμέτρων, είναι απαραίτητο η διεργασία να αποθηκευτεί. Ονομάστηκε “diploma\_trial” και αποθηκεύτηκε στον αποθηκευτικό χώρο της εφαρμογής και συγκεκριμένα στον φάκελο “my\_pRocesses”. (εικόνα 2.22)



Εικόνα 2.22: Αποθήκευση διεργασίας με τίτλο “diploma\_trial”

Υστερα η διεργασία εκτελείται είτε πατώντας το κουμπί της εκτέλεσης που βρίσκεται στην εργαλειοθήκη είτε το F11. Στην εικόνα 2.23 παρουσιάζονται τα αποτελέσματα της διεργασίας.



Εικόνα 2.23: Αποτελέσματα διεργασίας.

Η οθόνη των αποτελεσμάτων (Results) χωρίζεται σε δύο μέρη. Στο κάτω μέρος εμφανίζεται το κείμενο πριν την προ-επεξεργασία. Στο πάνω μέρος διαπιστώνεται η εφαρμογή των τριών χειριστών. Ο χειριστής “Tokenize” έχει χωρίσει το κείμενο σε μικρότερες ομάδες, καθώς η καθεμία λέξη είναι ξεχωριστή. Ο χειριστής “Stem (Snowball)” έχει αφαιρέσει τις καταλήξεις από όλες τις λέξεις και έχει διατηρήσει τα «στελέχη» τους, όπως είναι το «στελέχος» “increas”. Τέλος, ο χειριστής “Filter Stopwords (English)” έχει αφαιρέσει όλες τις λέξεις που δεν είναι σημαντικές για το περιεχόμενο ή για το νόημα του κειμένου. Για παράδειγμα στην πρώτη πρόταση, στο προ-επεξεργασμένο κείμενο δεν εμφανίζονται οι “stop words” λέξεις “and”, “are” και “the”.

## 2.3 Αλγόριθμος Latent Dirichlet Allocation (LDA) για εξαγωγή θέματος (topic extraction)

### 2.3.1 Η εξαγωγή θέματος (topic extraction)

Τα τελευταία χρόνια, εν μέρει και εξαιτίας της πανδημίας Covid-19, έχει παρατηρηθεί η ραγδαία αύξηση στο διαθέσιμο πλήθος των κειμένων βιοϊατρικού ενδιαφέροντος, έτσι ώστε να καθίσταται δύσκολη η διαλογή, κατηγοριοποίηση και εντέλει ανάγνωσή τους για την άντληση των σημαντικών πληροφοριών που εμπεριέχουν. Για αυτόν τον λόγο, η εξόρυξη κειμένου (text mining) και κατά συνέπεια η εξαγωγή θέματος έρχονται στο προσκήνιο, με σκοπό την διευκόλυνση της ανάλυσης αυτής της πληθώρας κειμένων. Τα βαθιά νευρωνικά δίκτυα (Deep Neural Networks, DNN), των οποίων η εκπαίδευση πραγματοποιείται με την βοήθεια της μηχανικής μάθησης, αποτελούν μία από τις πιο διαδεδομένες και εξελιγμένες τεχνικές εξόρυξης κειμένου. Η συγκεκριμένη τεχνική ονομάζεται **επιβλεπόμενη τεχνική (supervised learning)** διότι οι προϋπάρχουσες ετικέτες (labels) των ήδη χαρακτηρισμένων (από άνθρωπο-ειδικό) εκπαιδευτικών δειγμάτων (training set), διευκολύνουν την εκπαίδευση του αλγορίθμου ώστε να μπορεί στη συνέχεια να κατατάσσει και «άγνωστα», μη χαρακτηρισμένα (unlabelled) δεδομένα σε ομάδες, με υψηλά ποσοστά επιτυχίας. Όμως, στον βιοϊατρικό κλάδο αυτή η τεχνική καθίσταται δύσκολο να εφαρμοστεί εξαιτίας της ποικιλομορφίας των βιοϊατρικών όρων αλλά και του μεγάλου όγκου τους, που κάνει δύσκολη την δημιουργία χαρακτηρισμένων data sets. Για αυτό τον λόγο, οι ερευνητές ξεκίνησαν να χρησιμοποιούν τρεις διαφορετικές τεχνικές μηχανικής μάθησης, οι οποίες είναι:

- η μη επιβλεπόμενη (unsupervised learning),
- η αυτό-επιβλεπόμενη (self-supervised) και
- η ημι-επιβλεπόμενη τεχνική (semi-supervised).

Ως μη επιβλεπόμενη τεχνική ορίζεται η τεχνική μηχανικής μάθησης κατά την οποία δεν είναι απαραίτητο να υπάρχει διαθέσιμο ένα χαρακτηρισμένο σύνολο δεδομένων για εκπαίδευση. Αναρρείται δηλαδή η ανάγκη για χειροκίνητη τοποθέτηση ετικετών (labels) στα δεδομένα, κάτι το οποίο αποτελούσε το κύριο «κόστος» σε (ανθρώπινους) πόρους. Τα δεδομένα χρησιμοποιούνται από τον αλγόριθμο χωρίς να προηγηθεί φάση εκπαίδευσής του και τα κατηγοριοποιεί προσπαθώντας να βρει μόνος του ομοιότητες και διαφορές μεταξύ τους. Για να ποσοτικοποιήσει το στοιχείο ομοιότητα/διαφορά, χρησιμοποιεί διάφορα μέτρα (μαθηματικούς τύπους) ομοιότητας. Η τεχνική αυτή θεωρείται χρήσιμη και ευέλικτη λόγω της εξερεύνησης πολλών μονοπατιών συγχρόνως.

Η ενδιάμεση λύση είναι η αυτό-επιβλεπόμενη τεχνική, που είναι παρόμοια με την επιβλεπόμενη αλλά με την διαφορά ότι οι ετικέτες δημιουργούνται όχι από ανθρώπινη προεπεξεργασία, αλλά αυτοματοποιημένα, από τον ίδιο τον αλγόριθμο, με βάση το περιεχόμενο που εμπεριέχεται στο εκάστοτε σύνολο δεδομένων (εν προκειμένω, κείμενο), όπως ο ίδιος ο αλγόριθμος το εξάγει.

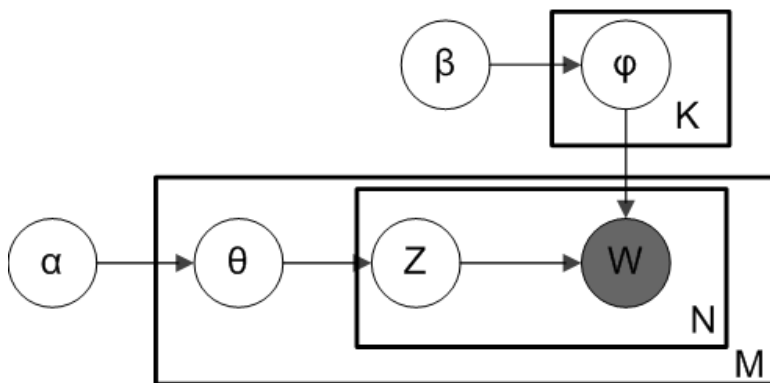
Τέλος, στην ημι-επιβλεπόμενη τεχνική, το νευρωνικό δίκτυο εκπαιδεύεται από δεδομένα με ή χωρίς ετικέτα. (4)

Η **μοντελοποίηση θεμάτων (topic modeling)** αποτελεί μία μη-επιβλεπόμενη τεχνική μηχανικής μάθησης, κατά την διαδικασία της οποίας τα κείμενα ομαδοποιούνται αυτοματοποιημένα, από τον ίδιο τον αλγόριθμο, με βάση τις λέξεις που εμπεριέχουν, τη συχνότητα χρήσης τους και τα συμφραζόμενά τους. Τα κείμενα που πρόκειται να κατηγοριοποιηθούν μπορούν να αντιστοιχίζονται σε ένα ή παραπάνω θέματα (topics) καθώς ο συνδυασμός των λέξεων από τις οποίες απαρτίζονται υπάρχει η πιθανότητα να ανήκουν σε μία ή περισσότερες ομάδες. Έτσι, η μοντελοποίηση παρέχει την δυνατότητα της αυτόματης οργάνωσης των κειμένων, της εύκολης αναζήτησης και της καλύτερης κατανόησης τους. Μία βέλτιστη μέθοδος μοντελοποίησης θεμάτων είναι ο αλγόριθμος Latent Dirichlet Allocation (LDA).

Ο αλγόριθμος **Latent Dirichlet Allocation (LDA)** αναπτύχθηκε το 2000 από ερευνητές με σκοπό να μελετήσουν την γενετική του πληθυσμού. Αργότερα, το 2003, οι David Blei, Andrew Ng και Michael Jordan τον εφάρμοσαν στον τομέα της μηχανικής μάθησης. Ο αλγόριθμος LDA είναι το πιο γνωστό Bayes μοντέλο που βασίζεται στον πιθανοτικό υπολογισμό. Σύμφωνα με το μοντέλο αυτό, κάθε κείμενο αποτελείται από ένα μείγμα θεμάτων και τα θέματα εμφανίζονται ως πιθανές κατανομές λέξεων. Η ανακάλυψη των θεμάτων και οι λέξεις που σχετίζονται με αυτά, γίνεται με πιθανοτική κατανομή Dirichlet. Πραγματοποιείται η δειγματοληψία πιθανοτήτων σε έναν πιθανοτικό χώρο, όπου το άθροισμα των τιμών των πιθανοτήτων ισούται με 1. Οι δειγματοληπτημένες τιμές αυτές κατατάσσουν τις πιθανότητες σε  $K$  διακριτές κατηγορίες. Η κατανομή Dirichlet με διάσταση  $K$  περιλαμβάνει  $K$  παραμέτρους και χρησιμοποιείται για να αναπαραστήσει τον βαθμό αβεβαιότητας ως κατανομή πιθανότητας. (5)

Στην εικόνα 2.24 παρουσιάζεται γραφικά η αναπαράσταση του μοντέλου του αλγορίθμου. Όπως φαίνεται στην εικόνα 2.24, η παράμετρος  $\mathbf{M}$  καθορίζει τον αριθμό των κειμένων και η παράμετρος  $\mathbf{N}$  αναφέρεται στον αριθμό των λέξεων που εμπεριέχονται σε ένα

δοσμένο κείμενο. Οι παράμετροι Dirichlet  $\alpha$  και  $\beta$  χρησιμοποιούνται ως παράμετροι σε *corpus*<sup>2</sup> επίπεδο και εκφράζουν αντίστοιχα την κατανομή θέματος ανά κείμενο και την κατανομή λέξης ανά θέμα. Η παράμετρος  $\Theta$  δηλώνει την κατανομή θέματος για έγγραφο  $i$ . Οι μεταβλητές  $\mathbf{z}$  και  $\mathbf{w}$ , οι οποίες βρίσκονται στο επίπεδο της λέξης, υποδηλώνουν το θέμα για την λέξη  $j$  στο έγγραφο  $i$  και την λέξη  $j$ , αντίστοιχα. Τέλος, η μεταβλητή  $\phi$  είναι αναπαράσταση της κατανομής πιθανοτήτων των λέξεων για ένα συγκεκριμένο θέμα  $k$ .



Εικόνα 2.24: Γραφική αναπαράσταση του αλγορίθμου LDA.

Σύμφωνα με το μοντέλο του αλγορίθμου LDA, τα έγγραφα θεωρούνται συνδυασμοί κρυμμένων θεμάτων, όπου κάθε θέμα αντιπροσωπεύεται από τον τρόπο που κατανέμονται οι λέξεις. Μέσα σε ένα σώμα (*corpus*), η διαδικασία αναπαράστασης του εκάστοτε εγγράφου που μπορεί να μοντελοποιηθεί από τον LDA, παρουσιάζεται ακολούθως:

- Σχεδιασμός της εκάστοτε παραμέτρου θέματος  $\beta \sim \text{Dirichlet}(\phi)$ , όπου  $k \in \{1 \dots K\}$
- Για κάθε κείμενο :
  1. Επιλογή διανομής θέματος  $\theta \sim \text{Dirichlet}(\alpha)$
  2. Για καθεμία από τις  $N$  λέξεις  $\mathbf{w}$  :
    - a) Επιλογή ενός θέματος  $\mathbf{z} \sim \text{Πολυωνυμική}(\theta)$
    - b) Επιλογή μίας λέξης  $\mathbf{w} \sim \text{Πολυωνυμική}(\beta)$

Η πιθανότητα συνόλου εγγράφων ( $D$ ) υπολογίζεται ως το γινόμενο των πιθανοτήτων των μεμονωμένων εγγράφων και η κατανομή πιθανότητας για ένα μεμονωμένο έγγραφο υπολογίζεται ως το άθροισμα των πιθανοτήτων του σε όλα τα θέματα  $\mathbf{z}$  και την ολοκλήρωση πάνω στην παράμετρο Dirichlet ( $\Theta$ ).

<sup>2</sup> Η έννοια “corpus”, η οποία μεταφράζεται ως «σώμα», υποδηλώνει το σύνολο των εγγράφων ή των κειμένων που πρόκειται να αναλυθούν από τον αλγόριθμο LDA.



$$P(\mathbf{W}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{i=1}^K P(\varphi_i; \beta) \prod_{j=1}^M P(\theta_j; \alpha) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{Z_{j,t}})$$

Εικόνα 2.25: Υπολογισμός πιθανότητας συνόλου εγγράφων

Ερμηνεύοντας την παραπάνω εξίσωση (εικόνα 2.25) με πιο απλό τρόπο, ισχύει ότι για την εκάστοτε λέξη σε κάθε έγγραφο και για κάθε θέμα, υπολογίζεται η πιθανότητα  $P(\text{θέμα} | \text{έγγραφο})$ , η οποία δηλώνει το ποσοστό των λέξεων σε ένα έγγραφο, οι οποίες συνδέονται άμεσα με ένα θέμα. Εν συνεχεία, η πιθανότητα  $P(\text{λέξη} | \text{θέμα})$  αναφέρεται στην αναλογία της αντιστοίχισης ενός θέματος σε ένα σώμα, το οποίο εμπεριέχει την συγκεκριμένη λέξη. Τέλος, ξανά πραγματοποιείται η διαδικασία της ανάθεσης μιας λέξης σε ένα καινούργιο θέμα, όπου αυτό προκύπτει από τον υπολογισμό της πιθανότητας  $P(\text{θέμα} | \text{έγγραφο}) * P(\text{λέξη} | \text{θέμα})$ . Έτσι, προκύπτει ένα καινούργιο θέμα, το οποίο παρήγαγε αυτή την λέξη.

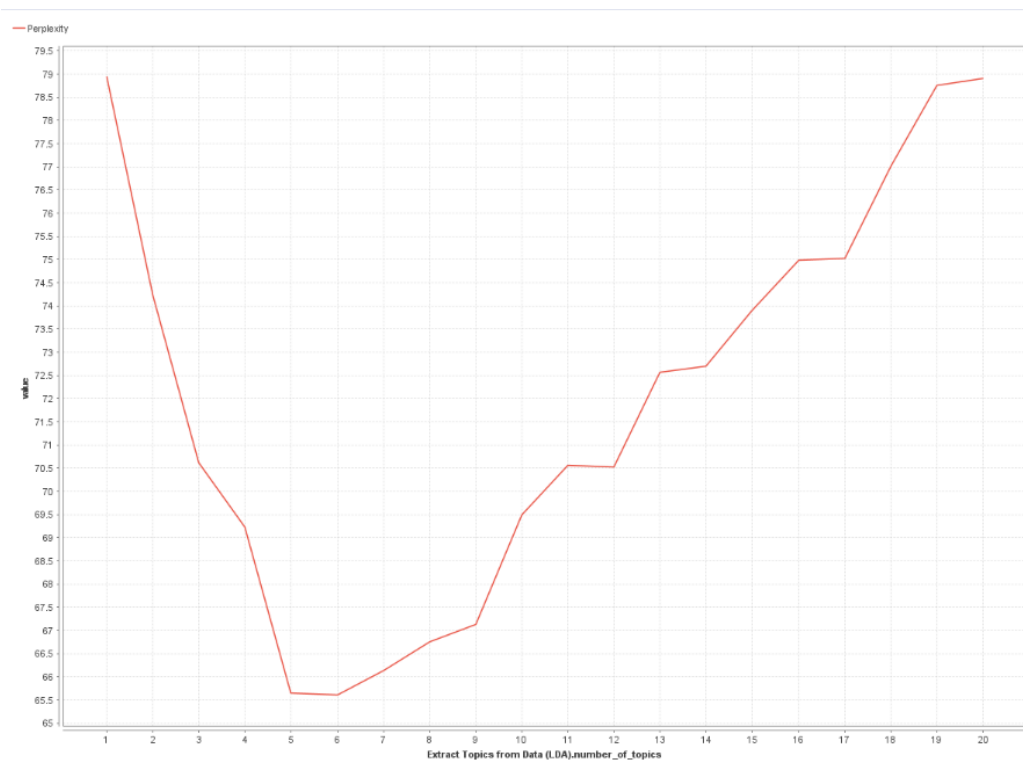
Επαναλαμβάνοντας τους παραπάνω υπολογισμούς πιθανοτήτων πολλαπλές φορές, προκύπτει μία -σχεδόν- σταθερή κατάσταση όπου η αντιστοίχιση θέματος-κειμένου θεωρείται κατάλληλη. Παράλληλα, είναι εφικτό να ανακαλυφθούν οι λέξεις οι οποίες έχουν την υψηλότερη πιθανότητα να αντιστοιχηθούν σε ένα θέμα. (6)

Ένα μείζον χαρακτηριστικό στο οποίο πρέπει να γίνει αναφορά, είναι ο προτιμητέος (βέλτιστος) αριθμός θεμάτων, με τον οποίο ένα «σώμα» έχει ορθά διαχωρισμένα τα έγγραφα του σε ομάδες. Η επιλογή του (βέλτιστου) αριθμού πραγματοποιείται βάσει των χαρακτηριστικών και της έκτασης του «σώματος» των κειμένων. Για παράδειγμα, όσο εκτενέστερο είναι το «σώμα», τόσο υψηλότερος θα είναι αυτός ο αριθμός, με την προϋπόθεση ότι υπάρχει ποικιλομορφία εγγράφων και δεν προστίθενται παραπλήσια έγγραφα στο υπάρχον «σώμα». Αυτό έχει ως σκοπό την αξιοποίηση της κρυμμένης πληροφορίας και την αξιολόγηση του αλγορίθμου. Η διαδικασία επιλογής του βέλτιστου αριθμού θεμάτων (number of topics) βασίζεται στον υπολογισμό της παραμέτρου Perplexity, η οποία υπολογίζεται με τον ακόλουθο τύπο:

$$Perplexity = e^{-\frac{\log \text{likelihood}}{\text{number of tokens}}}$$

Με βάσει τον αριθμητικό υπολογισμό της παραμέτρου Perplexity δοκιμάζοντας διαδοχικά αυξανόμενες τιμές πλήθους θεμάτων, έστω  $k$ , μέσα σε ένα εύρος τιμών που ορίζουμε εμπειρικά, δημιουργείται μία γραφική παράσταση  $Perplexity(x)$ , στην οποία ο άξονας  $y$  είναι οι τιμές Perplexity για κάθε  $k$  και ο άξονας  $x$  είναι οι διαδοχικές τιμές του πλήθους θεμάτων  $k$ . Στο παράδειγμα της εικόνας που ακολουθεί, η εν λόγω καμπύλη έχει χαραχθεί για κάποιο

σύνολο κειμένων και για εύρος τιμών  $k = 1 - 20$ . Παρατηρείται ότι η καμπύλη έχει ελάχιστο για  $k = 6$  και αυτό αποτελεί το βέλτιστο πλήθος θεμάτων για το συγκεκριμένο σώμα κειμένων. Εδώ αξίζει να σημειωθεί ότι ενδέχεται ένα σώμα κειμένων να μην έχει ένα τέτοιο καθαρό ελάχιστο, τουλάχιστον όχι μέσα στο εύρος τιμών του  $k$  που ορίστηκε εμπειρικά, ή να έχει περισσότερα από ένα τοπικά ελάχιστα. Στις περιπτώσεις αυτές οι ερευνητές ορίζουν το  $k$ , είτε εμπειρικά ώστε να συνάδει με τη φύση του προβλήματος, είτε με βάση άλλα κριτήρια όπως ένα μικρό σχετικά  $k$  που εξυπηρετεί την ταχύτητα της εφαρμογής.



Εικόνα 2.26 : Παράδειγμα γραφικής παράστασης για εύρεση βέλτιστου αριθμού θεμάτων. [Πηγή: <https://community.rapidminer.com/discussion/57377/whats-the-best-way-to-determine-the-number-of-topics-in-the-extract-topics-from-data-lda-operator>]

Ακόμη ένα βασικό χαρακτηριστικό του αλγορίθμου είναι ο όρος “Term Frequency-Inverse Document Frequency” (*TF-IDF*) και χρησιμοποιείται για να μετρήσει το βάρος μιας λέξης βάσει των φορών που εμφανίζεται μέσα σε ένα κείμενο. Πιο συγκεκριμένα, ελέγχει αν η συγκεκριμένη λέξη-κλειδί είναι σχετική του «σώματος» κειμένων μέσα στο οποίο εμπεριέχεται. Ο όρος Term Frequency αναφέρεται στην αναλογία του αριθμού μιας συγκεκριμένης λέξης της πρότασης σε σχέση με το σύνολο των λέξεων της ίδιας πρότασης. Ο όρος Inverse Document Frequency χαρακτηρίζει τον λογάριθμο της αναλογίας του συνόλου των γραμμών προς τις γραμμές που εμφανίζεται η συγκεκριμένη λέξη σε ένα κείμενο. Το

γινόμενο αυτών των δύο λόγων έχει ως αποτέλεσμα τον  $TF-IDF$ . Οι τιμές που λαμβάνει έχουν εύρος από 0 έως 1. Αν η τιμή  $TF-IDF$  πλησιάζει την τιμή 1, ο όρος έχει μεγαλύτερο βάρος δηλαδή συναντάται σπανιότερα μέσα στο κείμενο, ενώ αντίθετα όσο πλησιάζει η τιμή  $TF-IDF$  προς το 0, ο όρος έχει μικρότερο βάρος και άρα είναι πιο «κοινός».

Υπολογιστικά ισχύει ότι:

$$TF-IDF = TF * IDF,$$

όπου

- $TF$  = (φορές που εμφανίζεται ο όρος στο κείμενο/ συνολικό αριθμό των όρων του κειμένου)
- $IDF = \log \frac{N}{n}$ ,  $N$  συνολικός αριθμός γραμμών και  $n$  αριθμός των γραμμών που έχουν τον συγκεκριμένο όρο (7)

Κάθε αλγόριθμος μηχανικής μάθησης, συμπεριλαμβανομένου και του LDA, έχει κάποιους περιορισμούς που πρέπει να ληφθούν υπόψη. Παραδείγματα περιορισμών είναι τα εξής:

1. δεν μπορεί να εξαχθεί αυτοματοποιημένα ο αριθμός των κρυμμένων θεμάτων, άρα πρέπει να δοθεί στον αλγόριθμο ως είσοδος,
2. η κατανομή Dirichlet δεν μπορεί να ανακαλύψει συσχετίσεις μεταξύ των λέξεων ή των θεμάτων,
3. η φύση του αλγορίθμου είναι στατική καθώς δεν απεικονίζει την πρόοδο των θεμάτων χρονικά, και τέλος
4. η απλουστευτική προσέγγιση, όπως εκφράζεται και με τον όρο «σακούλι λέξεων»<sup>3</sup> (bag-of-words), όπου όλες οι λέξεις θεωρούνται σημασιολογικά 'ίσες' χωρίς να λαμβάνεται υπόψη η σειρά ή η θέση και ο συντακτικός ρόλος τους.

Όμως, οι περιορισμοί αυτοί δεν αποτρέπουν τους ερευνητές από τη χρήση του αλγορίθμου LDA. Η αποφυγή χρήσης ενός υπέρογκου «σώματος» κειμένων και η ορθή παραμετροποίηση οδηγούν στην επιτυχημένη εφαρμογή του αλγορίθμου για μοντελοποίηση θεμάτων καθώς και στην κατανόηση των αποτελεσμάτων του.

---

<sup>3</sup> Σακούλι λέξεων (bag-of-words): μία απλή αναπαράσταση κειμένου που χρησιμοποιείται τόσο στην φυσική επεξεργασία γλώσσας (NLP) όσο και σε μεθόδους εξόρυξης κειμένου (π.χ. LDA). Η συντακτική δομή και η σειρά των λέξεων αγνοούνται αλλά διατηρείται η πολλαπλότητα, δηλαδή την συχνότητα εμφάνισης της εκάστοτε λέξης.

### 2.3.2 Εξήγηση του αλγορίθμου με ένα παράδειγμα

Σε αυτή την ενότητα της διπλωματικής εργασίας, θα δοθεί ένα παράδειγμα εφαρμογής του αλγορίθμου LDA σε ένα τεχνητό, μη βιοϊατρικό πρόβλημα της καθημερινότητας, με σκοπό την κατανόηση του τρόπου λειτουργίας του.

Έστω οι ακόλουθες προτάσεις:

- Pasta and rice belong to the carbohydrate category.
- China and India belong to Asia.
- China was one of the world's first civilizations.
- There are simple and complex carbohydrates.
- Carbohydrates are often used in the preparation of traditional Chinese foods

Για να εφαρμοστεί ο αλγόριθμος στις προτάσεις, πρέπει πρώτα να πραγματοποιηθεί η διαδικασία της προ-επεξεργασίας. Όπως παρουσιάστηκε αναλυτικά στην ενότητα 2.2 , οι παραπάνω προτάσεις θα γίνουν ως εξής :

- “past” “ric” “belong” “carbohydrat” “categor””
- “Chin” “Ind” “belong” “Asia”
- “Chin” “world” “first” “civiliz”
- “simpl” “complex” “carbohydrat”
- “carbohydrat” “prepar” “tradition” “chin” “food”

Σύμφωνα με την διαδικασία της προ-επεξεργασίας, παρατηρείται ότι από τις προτάσεις αφαιρέθηκαν οι καταλήξεις των λέξεων και οι “stop words” και οι προτάσεις έχουν πάρει την μορφή των “tokens”. Έτσι γίνεται, πλέον, εφικτό να εφαρμοστεί θεωρητικά ο αλγόριθμος. Διακρίνονται δύο θέματα, τα οποία θα ονομαστούν θέμα 1<sup>ο</sup> και θέμα 2<sup>ο</sup> .

- Η πρώτη και η τέταρτη πρόταση ανήκουν στο θέμα 1<sup>ο</sup> καθώς εμπεριέχουν την λέξη “carbohydrat”.
- Η δεύτερη και η τρίτη πρόταση ανήκουν στο θέμα 2<sup>ο</sup> καθώς εμπεριέχουν την λέξη “Chin”.
- Τέλος, παρατηρείται ότι η Πέμπτη πρόταση αποτελείται τόσο από την λέξη του πρώτου θέματος όσο και από του δευτέρου. Άρα ισχύει ότι ανήκει και στα δύο θέματα, όχι όμως ισάξια, δηλαδή αντιπροσωπεύει κατά 70% το θέμα 1<sup>ο</sup> και κατά 30% το θέμα 2<sup>ο</sup>.<sup>4</sup>

---

<sup>4</sup> Τα ποσοστά δεν είναι αντιπροσωπευτικά αλλά χρησιμοποιήθηκαν για την κατανόηση του αλγορίθμου.

Συμπερασματικά, προκύπτει ότι ναι μεν μία πρόταση ή ένα έγγραφο μπορεί να ανήκει σε ένα μόνο θέμα αλλά υπάρχουν προτάσεις ή έγγραφα τα οποία αποτελούνται από μία μίξη θεμάτων. Τα ποσοστά αντικατοπτρίζουν την σχετικότητα της πρότασης ή του εγγράφου με τα εν λόγω θέματα: όσο μεγαλύτερο είναι το ποσοστό, τόσο πιο αντιπροσωπευτικό του όλου κειμένου είναι το θέμα (topic) που έχει ανακαλυφθεί/εξαχθεί.

## **ΚΕΦΑΛΑΙΟ 3: Βιοϊατρική Μηχανική και ο ρόλος των δημοσιεύσεων**

Το τρίτο κεφάλαιο της διπλωματικής εργασίας ασχολείται με τη σημασία και το ρόλο που παίζει η βιοϊατρική μηχανική στις δημοσιεύσεις που έγιναν κατά την περίοδο του Covid-19. Αυτό το στοιχείο συνδυάζεται με το ερευνητικό ερώτημα της εργασίας και οδηγεί στον πειραματικό σχεδιασμό. Εδώ τέλος παρουσιάζεται περιληπτικά και το πείραμα που θα πραγματοποιηθεί στο επόμενο, τέταρτο κεφάλαιο.

### **3.1 Γενική εισαγωγή περί Βιοϊατρικής Μηχανικής**

Το βασικό ερώτημα που θα έπρεπε να τεθεί είναι «τι ακριβώς είναι η βιοϊατρική μηχανική;» Ένας από τους (πολλούς) ορισμούς που έχουν δοθεί αναφέρει ότι ο κλάδος αφορά τις αρχές της μηχανικής (engineering) όπως αυτές βρίσκουν εφαρμογή στον τομέα της βιολογίας και της ιατρικής. Με την αυξανόμενη ανάγκη για καινοτόμες λύσεις στον τομέα της υγείας, η βιοϊατρική μηχανική παίζει έναν ουσιαστικό ρόλο στην βελτίωση της ιατρικής πρακτικής και στην παροχή καλύτερης φροντίδας στους ασθενείς. Ασχολείται με μία ευρεία γκάμα προκλήσεων, όπως προηγμένων ιατρικών συσκευών και εξοπλισμού, μέχρι τη σχεδίαση εξειδικευμένων λογισμικών για την ανάλυση ιατρικών εικόνων καθώς και την προσομοίωση χειρουργικών επεμβάσεων. Για παράδειγμα, ένα από τους κύριους τομείς εφαρμογής της βιοϊατρικής μηχανικής είναι η χειρουργική με ρομπότ, τα οποία εκτελούν πολύπλοκες επεμβάσεις με υψηλή ακρίβεια. Επιπλέον, η συνεχής εξέλιξη της τεχνολογίας και η συνεργασία -μεταξύ επιστημών και ιατρών- απέφερε θετικά αποτελέσματα, όπως είναι η ανάπτυξη ανοσοθεραπειών, φαρμάκων και θεραπειών που στοχεύουν στην καταπολέμηση των νόσων.

Με τον τομέα της βιοϊατρικής μηχανικής, οι επιστήμονες ασχολήθηκαν για πρώτη φορά γύρω στο 1940, ενώ την δεκαετία του 1960 εισήχθη σε ακαδημαϊκά ιδρύματα όπως το University of Virginia, Case Western Reserve University κ.ά. Αργότερα, το 1970, το πρώτο ουσιώδες πείραμα, που επιτεύχθηκε με την βοήθεια της βιοϊατρικής μηχανικής, είναι η δημιουργία γενετικής μίξης μορίων DNA, τα οποία προέκυψαν από την ένωση διαφορετικών ειδών DNA και τα εισήγαγαν σε ένα κύτταρο-οικοδεσπότη. Αυτό είχε ως αποτέλεσμα, οι επιστήμονες να παράγουν νέους γενετικούς συνδυασμούς σε ένα κύτταρο-οικοδεσπότη. Η μεγάλη ανατροπή στον τομέα αυτόν έγινε τα τέλη του 20<sup>ου</sup> αιώνα, όπου οι επιστήμονες έθεσαν

ως στόχο να αποτυπώσουν την πλήρη αλληλουχία του ανθρώπινου DNA, το οποίο κατόρθωσαν με την αναγνώριση όλων των βασικών ζευγαριών των γονιδίων του ανθρώπινου γονιδιώματος.

Οι βιοϊατρικοί μηχανικοί, έχοντας κάνει κάποια βασικά πρώτα βήματα σε αυτόν τον τομέα, συνέχισαν να συμβάλλουν στην εξέλιξη του. Ένα θεμελιώδες πεδίο αποτελεί η τεχνολογία υλικών, που αφορά την ανάπτυξη και επιλογή των κατάλληλων εξαρτημάτων με σκοπό την κατασκευή ιατρικών συσκευών και εμφυτευμάτων. Τα υλικά αυτά παρήχθησαν με τον κατάλληλο τρόπο ώστε να είναι ανтимικροβιακά και βιοσυμβατά. Ο σχεδιασμός και η κατασκευή μηχανικών μερών συσκευών και εξοπλισμού υγείας, όπως είναι τα πρόσθετα μέλη, είναι ένας ακόμη βασικός τομέας της βιοϊατρικής μηχανικής. Εν συνεχεία, όσον αφορά τον τομέα της ηλεκτρονικής, αναπτύχθηκαν ηλεκτρονικά συστήματα, με τα οποία λαμβάνονται, καταγράφονται και επεξεργάζονται ιατρικά σήματα, κάποια από τα οποία είναι το καρδιογράφημα και η ακτινογραφία. Με την χρήση υπολογιστών και εξειδικευμένων λογισμικών, τα λαμβάνοντα σήματα επεξεργάζονται και αναλύονται, βοηθώντας στην διάγνωση και στην παρακολούθηση των ασθενών. Παράλληλα, οι ανεπτυγμένοι αλγόριθμοι προσφέρουν την δυνατότητα εξαγωγής συμπερασμάτων από ιατρικά δεδομένα καθώς ανακαλύπτονται νέες, επεξεργάσιμες πληροφορίες.

### **3.2 Ο ρόλος των δημοσιεύσεων στην προαγωγή της έρευνας**

Ο ρόλος των δημοσιεύσεων στην ερευνητική κοινότητα είναι πρωταρχικός και αναπόσπαστος. Αποτελούν την γέφυρα μεταξύ των επιστημονικών ανακαλύψεων και του κοινού, διαδίδοντας τις νέες γνώσεις, ανοίγοντας νέους δρόμους στην έρευνα και συμβάλλοντας στην ανάπτυξη της επιστήμης. Το θεμέλιο της επιστημονικής προόδου αποτελούν οι δημοσιεύσεις, βάσει των οποίων πραγματοποιείται η ανταλλαγή ιδεών ενώ δημιουργούνται και πρωτοποριακές τεχνολογίες.

Οι δημοσιεύσεις ή τα επιστημονικά άρθρα αποτελούν ένα κρίσιμο μέσο για την ανάδειξη και διάδοση της έρευνας και των αποτελεσμάτων της. Για να μπορέσει να δημοσιευτεί μία έρευνα, πρώτα κρίνεται απαραίτητο να αξιολογηθεί από άλλους επιστήμονες και εμπειρογνώμονες. Η διαδικασία της αξιολόγησης βεβαιώνει την ποιότητα και την εγκυρότητα των αποτελεσμάτων. Οι ερευνητές, μέσω των δημοσιεύσεων, κοινοποιούν τα αποτελέσματα των ερευνών τους τόσο στην επιστημονική κοινότητα όσο και στο κοινό ώστε να διαδοθεί η γνώση και η καινοτομία. Επιπλέον, παρατηρείται πρόοδος και εξέλιξη στον επιστημονικό τομέα καθώς οι νέες ιδέες και οι ανακαλύψεις αποτελούν την υποδομή για περαιτέρω έρευνα.

Οι δημοσιεύσεις των ερευνών απέκτησαν μεγαλύτερη σημασία κατά την εποχή του Covid-19, λόγω του ότι η πανδημία τις επηρέασε άμεσα και σημειώθηκε ένας αυξημένος όγκος πληροφοριών και μελετών που αφορούσαν τον Covid-19. Πολλοί ερευνητές συνεργάστηκαν και έστρεψαν την προσοχή τους την έρευνα για τον Covid-19 με σκοπό την ανεύρεση θεραπειών και εμβολίων, ενώ συγχρόνως έκαναν επιδημιολογικές αναλύσεις, παρέχοντας κρίσιμες πληροφορίες. Ο ρόλος των δημοσιεύσεων έπαιξε σημαντικό ρόλο στην διάχυση της νέας πληροφορίας, το οποίο δημιούργησε σύγχυση στην επιστημονική κοινότητα σχετικά με την αξιοπιστία των επιστημονικών άρθρων. Καθοριστικός παράγοντας στην ανίχνευση των πληροφοριών ήταν η εξόρυξη βιοϊατρικών κειμένων (biomedical text mining). Η εξόρυξη των δεδομένων, η κατηγοριοποίηση και η επεξεργασία τους αποτέλεσε την βάση για την καλύτερη και ταχύτερη κατανόηση του ιού καθώς και για την ανακάλυψη μίας πιθανής θεραπείας και τον σχεδιασμό των εμβολίων αντιμετώπισής του.

Στην παρούσα διπλωματική εργασία, επίκεντρο είναι οι ερευνητικές δημοσιεύσεις σχετικά με την νόσο Alzheimer. Οι πρόσφατες δημοσιεύσεις παρέχουν βασικές πληροφορίες για την κατανόηση των θεμελιωδών εννοιών σχετικά με τη νόσο και την αντιμετώπισή της.

- *Evolution of Alzheimer's disease research from a health-tech perspective: Insights from text mining (Dominic D. Martinelli, 2022)*

Περίληψη: Χιλιάδες επιστημονικές δημοσιεύσεις έχουν ασχοληθεί με τον κίνδυνο που προκαλεί η νόσος Alzheimer, ξεπερνώντας τα όρια των επιστημονικών πεδίων που προσφέρουν νέες εισηγήσεις σχετικά με την νόσο και τις θεραπευτικές επιλογές. Η εξόρυξη κειμένου βοηθάει στην διευκόλυνση της υπέρογκης ποσότητας ερευνών. Η εξόρυξη της πρόσφατης γνώσης για τους φυσικούς μηχανισμούς και τις χημικές ουσίες της νόσου, έγινε από 17.286 περιλήψεις των τελευταίων 5 ετών, τα οποία αναλύθηκαν, αξιολογήθηκε η ομοιότητά τους και χρησιμοποιήθηκε ο αλγόριθμος LDA. Η ταξινόμηση των περιλήψεων έγιναν με τους τρόπους που αναφέρονται στο άρθρο και παρουσιάζεται τα αποτελέσματα του αλγορίθμου LDA.

- *Community challenges in biomedical text mining over 10 years: success, failure and the future (Chung-Chi, Zhiyong Lu, 2015)*

Περίληψη: Το άρθρο αυτό εστιάζει στον τομέα της επεξεργασίας της φυσικής γλώσσας για την βιοϊατρική έρευνα. Εξετάζονται οι διάφορες αξιολογήσεις που πραγματοποιήθηκαν από το 2002 έως το 2014 και παρουσιάζονται οι προκλήσεις που προέκυψαν από την έρευνα της επεξεργασίας φυσικής γλώσσας και των βιοϊατρικών



εφαρμογών. Ο συνδυασμός αυτών των δύο τομέων αναλύεται και γίνεται συζήτηση γύρω από του περιορισμούς και τις δυσκολίες που ενδέχεται να υπάρχουν.

- *Neural network-based approaches for biomedical relation classification: A review (Yijia Zhang, Hongfei Lin, Zhihao Yang, Jian Wang, Yuanyan Sun, Bo Xu, Zhenhuan Zhao, 2019)*

Περίληψη: Σε αυτό το άρθρο παρουσιάζεται η επιρροή των νευρωνικών δικτύων στην κατηγοριοποίηση των βιοϊατρικών σχέσεων. Δύο είδη νευρωνικών δικτύων (CNN, RNNs) παρουσιάζονται και εφαρμόζονται στα διαθέσιμα «σώματα» ενώ το άρθρο καταλήγει με τα αποτελέσματα των προσεγγίσεων των ειδών των νευρωνικών δικτύων και των προκλήσεων που αντιμετώπισαν.

- *BioBERT: a pre-trained biomedical language representation model for biomedical text mining (Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang, 2019)*

Περίληψη: Γίνεται παρουσίαση του BioBERT, το οποίο είναι ένα προ-εκπαιδευμένο μοντέλο για μεγάλα βιοϊατρικά «σώματα». Η διαδικασία του πειράματος και τα αποτελέσματα που παρατίθενται, αποδεικνύουν την καταλληλότητα του μοντέλου.

- *Unsupervised and self-supervised deep learning approaches for biomedical text mining (Mohamed Nadif, Francois Role, 2021)*

Περίληψη: Το άρθρο αναφέρεται στην αυτοματοποιημένη εξόρυξη κειμένου που προσφέρουν οι τεχνικές των βαθιών νευρωνικών δικτύων. Συγκεκριμένα, αναλύει δύο τεχνικές, την unsupervised και self-supervised και τις συγκρίνει με την supervised τεχνική.

- *A systematic review of text mining approaches applied to various application areas in the biomedical domain (Sudha Cheerkoot-Jalim, Kavi Kumar Khedo, 2020)*

Περίληψη: Σε αυτή την δημοσίευση παρουσιάζονται τα αποτελέσματα της συστηματικής βιβλιογραφικής ανασκόπησης σχετικά με την βιοϊατρική εξόρυξη κειμένου. Παρουσιάζονται διαφορετικές προσεγγίσεις εξόρυξης κειμένου για τα διαφορετικά πεδία της βιοϊατρικής.

- *A Review on Biomedical Mining (Rachakonda Venkatesh, Kosaraju Chaintanya, Thulasi Bikku, Radhika Paturi, 2019)*

Περίληψη: Το άρθρο εξηγεί την σημασία της εξόρυξης δεδομένων από μεγάλα αποθετήρια εγγράφων και της κατηγοριοποίησης τους με σκοπό να ανακαλύψουν τα μοτίβα γονιδίων-ασθενειών.

- *SSL: Inferring disease-related genes using Sentence Structure and Literature data (Jeongwoo Kim, Won Gi Choi, Jungrim Kim, Sanghyum Park, 2017)*

Περίληψη: Στο άρθρο προτείνεται η μέθοδος SSL, η οποία χρησιμοποιεί την δομή της πρότασης και τα λογοτεχνικά δεδομένα για να ανιχνεύσει τις συσχετίσεις μεταξύ ασθενειών και γονιδίων. Αφού αναλυθεί ο τρόπος λειτουργίας της μεθόδου, γίνεται η εφαρμογή της σε πέντε διαφορετικές ασθένειες και τα αποτελέσματα που προκύπτουν, συγκρίνονται με άλλες μεθόδους.

### 3.3 Το ερευνητικό ερώτημα της διπλωματικής εργασίας

Το ερώτημα που τέθηκε σε αυτή την διπλωματική εργασία είναι το εξής:

- Είναι αντιπροσωπευτικές οι περιλήψεις (abstracts) των δημοσιεύσεων βιοϊατρικών ερευνών στο θέμα της νόσου Alzheimer; Ένας ερευνητής μπορεί να αντλήσει όλες τις σημαντικές πληροφορίες για την διαλογή και κατηγοριοποίηση μίας (νέας) δημοσίευσης, προς περαιτέρω μελέτη, κάνοντας ανάγνωση μόνο της περίληψης ή είναι αναγκαίο να διαβάσει ολόκληρο το κείμενο;

Ας σημειωθεί εδώ ότι η ύπαρξη μίας περίληψης (abstract) στην αρχή κάθε επιστημονικής δημοσίευσης προσπαθεί να καλύψει ακριβώς αυτή την ανάγκη. Η προσπάθεια της διπλωματικής εργασίας είναι να επιβεβαιώσει και πειραματικά (μέσα από την αξιοποίηση των δυνατοτήτων της τεχνολογίας text mining) την ορθή λειτουργία των abstracts, δηλαδή ότι εκπληρώνουν το ρόλο τους,.

Σύμφωνα με την ανάλυση που προηγήθηκε, η παρούσα διπλωματική εργασία εξετάζει πειραματικά αυτό το ερώτημα με βάση επιστημονικά κείμενα βιοϊατρικών ερευνών. Συγκεκριμένα, οι δημοσιεύσεις αφορούν την νόσο Alzheimer, μία χρόνια νευρολογική διαταραχή που προκαλεί σοβαρές νοητικές δυσλειτουργίες σε πολλούς ανθρώπους. Η βιοϊατρική έρευνα αποτελεί σημαντικό βοήθημα για την καλύτερη κατανόηση και την αντιμετώπιση της νόσου Alzheimer. Η ανάλυση των βιοϊατρικών κειμένων παρέχει πολύτιμες πληροφορίες για την πάθηση, αλλά η πολυπλοκότητα των δεδομένων απαιτεί εξειδικευμένες τεχνικές. Για την αντιμετώπιση αυτού του προβλήματος υλοποιήθηκε το παρακάτω πείραμα.

Αναπτύχθηκε στο περιβάλλον RapidMiner ένα πείραμα εξαγωγής θέματος (topic extraction) αφενός από επιστημονικά άρθρα σχετικά με τη νόσο Alzheimer, και αφετέρου από τις περιλήψεις (abstracts) των ίδιων άρθρων. Στο τέλος συγκρίθηκαν τα αποτελέσματα για να εξαχθούν συμπεράσματα.

Αφού αναπτύχθηκε η αρχιτεκτονική του πειράματος (σύνολο διεργασιών και διασύνδεσή τους, είσοδοι και έξοδοι) εντός του περιβάλλοντος RapidMiner, και πριν χρησιμοποιηθεί με τα βιοϊατρικά δεδομένα, πραγματοποιήθηκε ως πιλοτική αξιολόγηση μία απλή πειραματική διαδικασία με βάση μη βιοϊατρικά κείμενα, για επιβεβαίωση της ορθής λειτουργίας της εφαρμογής. Η demo αυτή διαδικασία έγινε με κείμενα αντλημένα από το Internet, τα οποία ανήκουν σε 3 εντελώς διαφορετικές θεματολογίες. Η καθεμία από τις 3 ομάδες αποτελείται από 5 κείμενα, έκτασης μίας σελίδας περίπου το καθένα. Η εκτέλεση της διεργασίας πάνω στο εν λόγω τεχνητό σώμα κειμένων και τα αποτελέσματά της, επιβεβαιώνουν ότι το μοντέλο που έχει δημιουργηθεί λειτουργεί σωστά καθώς θα γίνει σύγκριση μεταξύ της αυτόματης κατηγοριοποίησης των εγγράφων και της χειροκίνητης, η οποία έχει γίνει από την γράφουσα.

Εν συνεχεία, στο μοντέλο αυτό θα εισαχθούν τα βιοϊατρικά κείμενα που έχουν αντληθεί από την ιστοσελίδα PubMed και είναι ήδη ομαδοποιημένα στις κατηγορίες

1. Controlled Clinical Trial,
2. Clinical Trial I και
3. Clinical Trial II.

Στο πειραματικό μέρος, που θα παρουσιαστεί αναλυτικά στο ακόλουθο κεφάλαιο 4, αφού εισαχθούν με τον κατάλληλο χειριστή στο εργαλείο RapidMiner τα βιοϊατρικά κείμενα, ακολουθεί η προ-επεξεργασία τους, και καταλήγουν έτσι στην εφαρμογή του αλγορίθμου LDA για εξαγωγή των θεμάτων (topic extraction). Η διαδικασία επαναλαμβάνεται για τις περιλήψεις των άρθρων, και τέλος συγκρίνονται τα αποτελέσματα.

Τα αποτελέσματα αναμένεται να βοηθήσουν στην πλήρη αξιολόγηση του μοντέλου. Με βάση αυτήν την αξιολόγηση, θα απαντηθεί και το βασικό ερώτημα της διπλωματικής εργασίας, το οποίο ευελπιστούμε ότι θα διευκολύνει τους ερευνητές, επιστήμονες και μελετητές βιοϊατρικών κειμένων, εξοικονομώντας πόρους από την ανάγνωση πληθώρας βιοϊατρικών ερευνών αλλά και φέρνοντας στην επιφάνεια και επισημαίνοντάς τους «κρυμμένες» θεματολογίες των άρθρων, με σκοπό την βέλτιστη κατανόηση της νόσου Alzheimer καθώς και των τρόπων αντιμετώπισής της.

## ΚΕΦΑΛΑΙΟ 4: Πειραματικό μέρος και αποτελέσματα

Στο παρόν κεφάλαιο της διπλωματικής εργασίας, που αποτελεί και το κύριο πειραματικό μέρος, αρχικά παρουσιάζεται η σχεδίαση και ανάπτυξη της σύνθετης, δομημένης διεργασίας (process) για την διεκπεραίωση του πειράματος. Στην συνέχεια, η διεργασία αυτή δοκιμάζεται και αξιολογείται πιλοτικά σε ένα demo πείραμα με μη βιοϊατρικά κείμενα, το οποίο θα επιβεβαιώσει την ορθή λειτουργία της. Ακολούθως σχεδιάζονται και εκτελούνται δύο πειράματα, το πρώτο με είσοδο το σώμα των περιλήψεων βιοϊατρικών κειμένων και το δεύτερο με είσοδο τα πλήρη κείμενα καθανά. Τέλος συγκρίνονται τα αποτελέσματα, από τα οποία προκύπτει η απάντηση του ερωτήματος που τέθηκε στην *Ενότητα 3.3*.

### 4.1 Ανάπτυξη της δομημένης διεργασίας στο RapidMiner

Για την εκτέλεση του πειράματος, σχεδιάστηκε και προσομοιώθηκε στο RapidMiner από μηδενική βάση μία διεργασία (process) εξόρυξης κειμένου και εξαγωγής θεμάτων (topic extraction) δομημένη σε διαδοχικά επίπεδα και με επαναληπτική μορφή ώστε αυτοματοποιημένα να χειρίζεται το σώμα των κειμένων, όπως συνοπτικά δόθηκε στην *Ενότητα 2.2*.

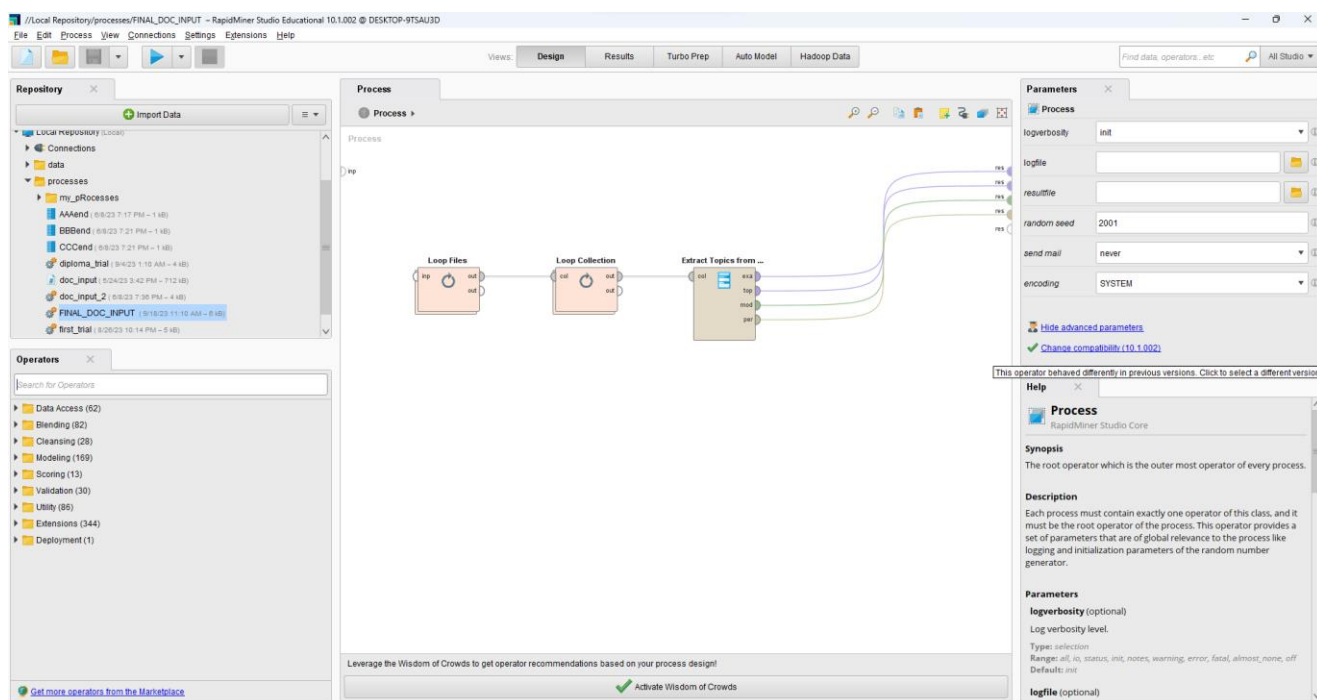
Ανοίγοντας την εφαρμογή RapidMiner, γίνεται η επιλογή “Blank Process” και εμφανίζεται το κενό παράθυρο της επιφάνειας εργασίας. Η διεργασία αποτελείται από δύο επίπεδα, εκ των οποίων στο πρώτο είναι σχεδιασμένη η βασική διεργασία και στο δεύτερο κάποιοι βοηθητικοί χειριστές.

Ξεκινώντας από το δεύτερο επίπεδο, οι χειριστές που εμφανίζονται εκεί είναι ο “**Read Document**”, ο “**Tokenize**”, ο “**Stem(Snowball)**”, και τέλος ο “**Filter Stopwords (English)**”. Ο χειριστής “Read Document” ανήκει στην κατηγορία χειριστών “Extensions” και συγκεκριμένα στην “Text Processing” και η λειτουργία του αφορά την ανάγνωση ενός κειμένου από το πρόγραμμα RapidMiner. Οι χειριστές “Tokenize”, “Stem(Snowball)” και “Filter Stopwords (English)” ανήκουν ομοίως στην ίδια κατηγορία “Text Processing” και ειδικότερα στις υποκατηγορίες “Tokenization”, “Stemming” και “Filtering” αντίστοιχα. Η λειτουργία του εκάστοτε χειριστή παρουσιάζεται στην *ενότητα 2.2* με τις κατάλληλες παραμετροποιήσεις.

Στο πρώτο επίπεδο, ο πρώτος χειριστής που συναντάται είναι ο “**Loop Files**”. Ανήκει στην κατηγορία “Utility” → “Process Control” → “Loops”. Ο χειριστής αυτός είναι ένας βρόχος και εκτελεί την υπό-διεργασία, η οποία είναι η λειτουργία του “Read Document”, πάνω

στα επιλεγμένα αρχεία. Συνεχίζοντας, ο επόμενος χειριστής είναι ο **“Loop Collection”** και ανάγεται στην ίδια κατηγορία με τον προηγούμενο. Επαναλαμβάνει μία διαδικασία πάνω σε μία συλλογή αντικειμένων και η υπό-διεργασία εκτελείται μία φορά για κάθε αντικείμενο αυτής της συλλογής. Οι χειριστές **“Tokenize”**, **“Stem(Snowball)”** και **“Filter Stopwords (English)”** τοποθετούνται στο εσωτερικό του **“Loop Collection”**, καθώς απαιτείται η προεπεξεργασία κάθε κειμένου που ανήκουν στα εισαγόμενα αρχεία. Τέλος, είναι τοποθετημένος ο χειριστής **“Extract Topics from Documents (LDA)”** της κατηγορίας **“Text Processing”**. Όπως γίνεται αντιληπτό, είναι ο χειριστής που θα εξαγάγει τα κρυμμένα θέματα των κειμένων και θα τα κατατάσσει σε ομάδες.

Η ολοκληρωμένη διεργασία παρουσιάζεται στην εικόνα 4.1 και έχει αποθηκευτεί στο τοπικό Repository processes του RapidMiner με τίτλο **“FINAL\_DOC\_INPUT”**.

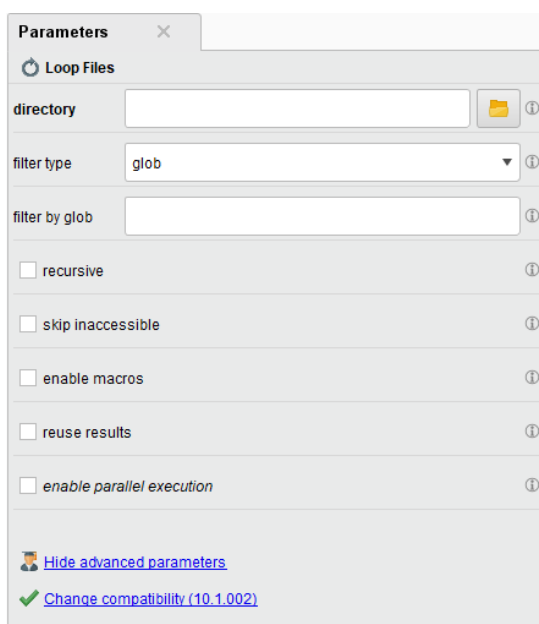


Εικόνα 4.1: Η δομημένη διεργασία στο περιβάλλον RapidMiner.

Για την εμφάνιση αποτελεσμάτων, πρέπει να συνδεθούν οι έξοδοι **“exa”**, **“top”**, **“mod”** και **“per”** του **“Extract Topics from Documents (LDA)”** με τις εξόδους **“res”** του προγράμματος. Αναλυτικότερα, η έξοδος **“exa”** θα δημιουργήσει έναν πίνακα με 3 στήλες, οι οποίες ονομάζονται **“documentId”**, **“TopicId”** και **“confidence(Topic\_x)”**. Η πρώτη στήλη αριθμεί τα κείμενα, στην δεύτερη στήλη αναγράφεται σε ποιο θέμα ανήκει το εκάστοτε κείμενο και στην τρίτη στήλη εμφανίζονται τα ποσοστά συσχέτισης των κειμένων με όλα τα θέματα. Η έξοδος **“top”** θα εμφανίσει στην καρτέλα των αποτελεσμάτων τις 3 πιο συχνές λέξεις που

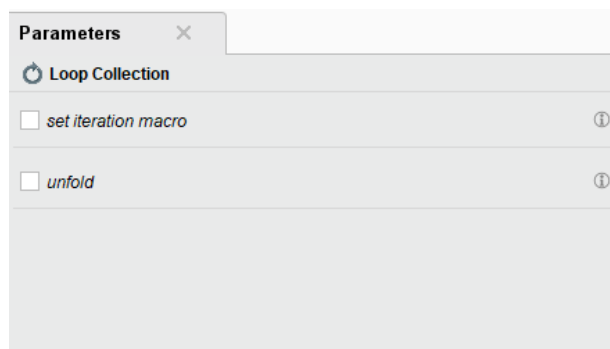
αναδύθηκαν από κάθε θέμα καθώς και τα βάρη τους. Η έξοδος “mod” θα παρουσιάσει το μοντέλο των θεμάτων που δημιουργήθηκε μαζί με τις τιμές των παραμέτρων  $\alpha$  και  $\beta$ , που αναφέρονται στην ενότητα 2.3.1 και παρουσιάζει ένα εκπαιδευμένο μοντέλο που μπορεί να εφαρμοστεί σε μία νέα συλλογή. Τέλος, η έξοδος “per” δίνει την τιμή Loglikelihood, από την οποία αξιολογείται το μοντέλο.

Οι παράμετροι του χειριστή “Loop Files” δεν αλλάζουν, εκτός από τον καθορισμό του καταλόγου, δηλαδή από ποιον φάκελο του υπολογιστή θα αντληθούν τα υπό εξέταση αρχεία. (Εικόνα 4.2)



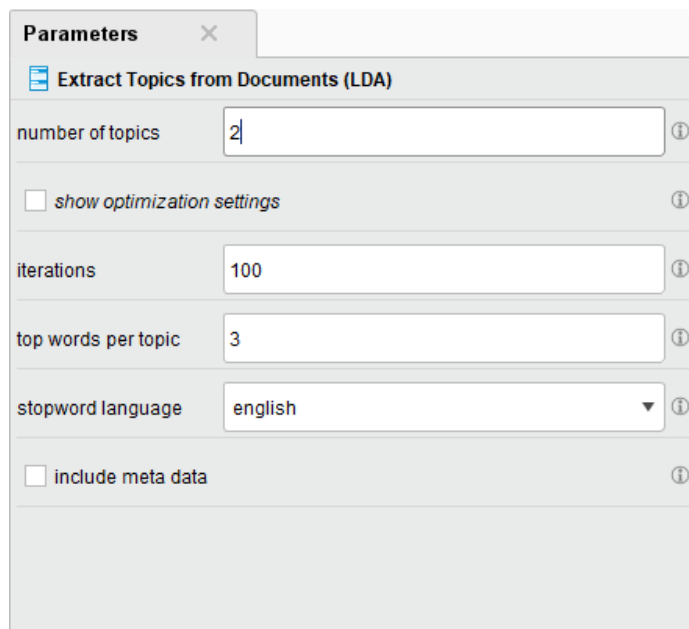
Εικόνα 4.2: Παραμετροποίηση του χειριστή “Loop Files”.

Συνεχίζοντας την διαδικασία της παραμετροποίησης, στον χειριστή “Loop Collection” αφήνονται οι default παράμετροι. (Εικόνα 4.3)



Εικόνα 4.3: Παραμετροποίηση του χειριστή “Loop Collection”.

Όπως φαίνεται στην εικόνα 4.4, οι παράμετροι του χειριστή “Extract Topics from Documents (LDA)” μεταβάλλονται. Αρχικά, ορίζεται ένας αριθμός θεμάτων (number of topics), βάσει του οποίου θα διαχωριστούν -αρχικά- τα κείμενα σε 2 θέματα και κάθε θέμα θα αποτελείται από 3 λέξεις (top words per topic). Οι επαναλήψεις (iterations) που θα εκτελεστούν για τον διαχωρισμό αυτό, ανέρχονται στις 100.



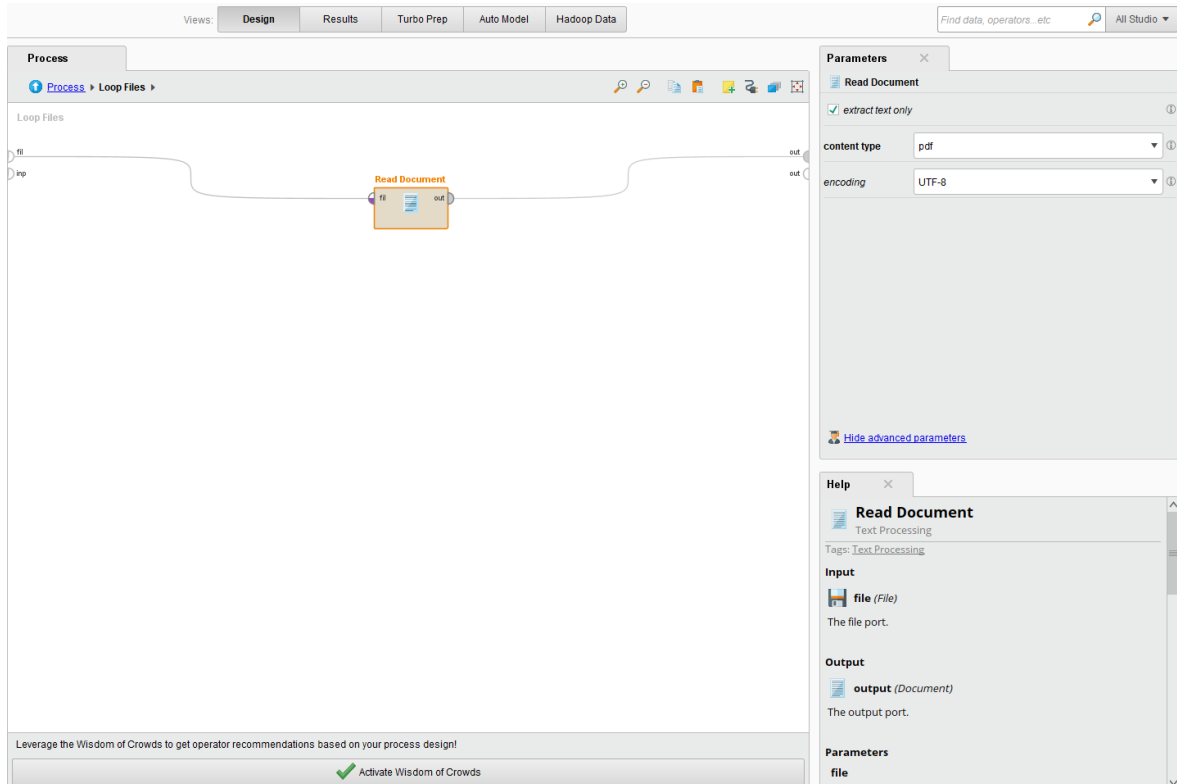
Εικόνα 4.4: Παραμετροποίηση του χειριστή “Extract Topics from Documents (LDA)”.

Το δεύτερο επίπεδο, όπως αναφέρθηκε παραπάνω, αποτελείται από τον χειριστή “Read Document” που εμπεριέχεται στον χειριστή “Loop Files” (εικόνα 4.5) και από τους χειριστές “Tokenize”, “Stem(Snowball)” και “Filter Stopwords (English)” που βρίσκονται στο εσωτερικό του βρόχου “Loop Collection” (εικόνα 4.6).

Στην εικόνα 4.5, στο πάνω δεξιά πλαίσιο εμφανίζονται οι παράμετροι του χειριστή “Read Document”, οι οποίες παραμένουν ως έχουν. Για να εκτελεστεί η λειτουργία του βρόχου, πρέπει να συνδεθεί η είσοδος “fil” του “Read Document” με την εσωτερική είσοδο “fil” του “Loop Files” όπως πρέπει να γίνει και αντίστοιχα με την έξοδο “out” του “Read Document”, η οποία πρέπει να συνδεθεί με την εσωτερική έξοδο “out” του “Loop Files”.

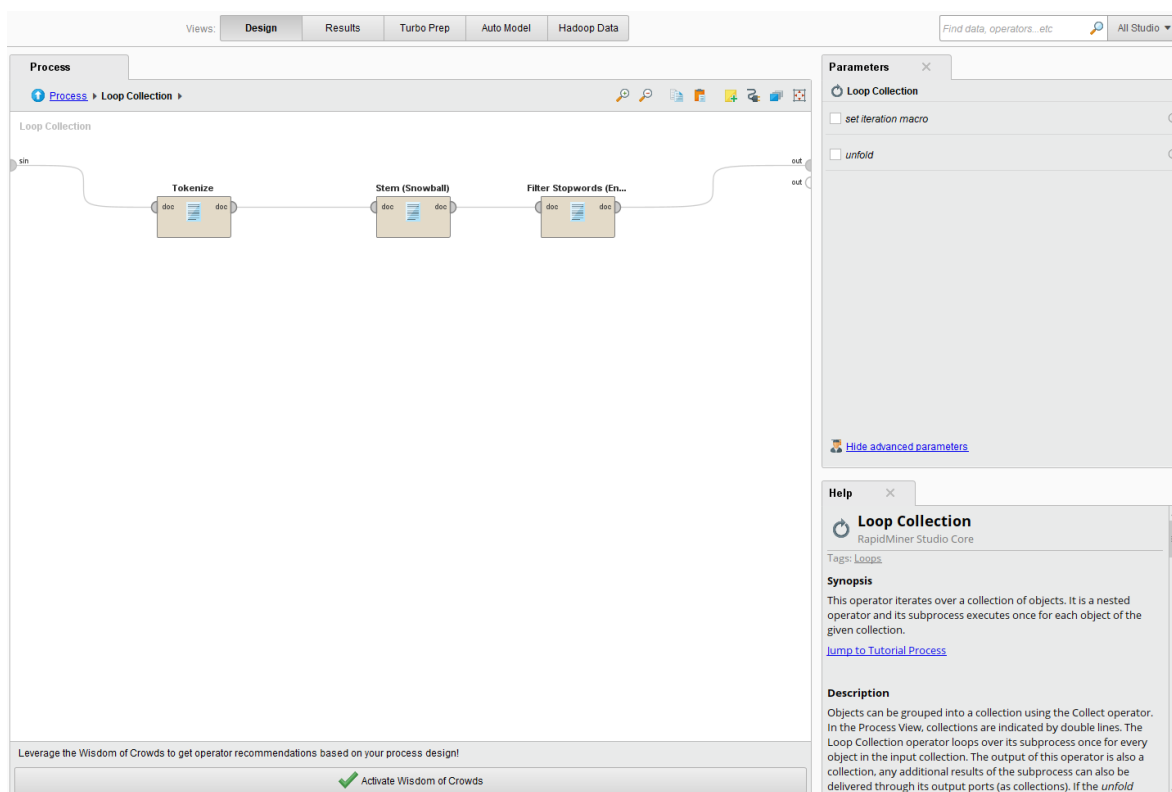
Οι παράμετροι των χειριστών “Tokenize”, “Stem(Snowball)” και “Filter Stopwords (English)” έχουν δοθεί στην ενότητα 2.2. Υπενθυμίζεται ότι ο χειριστής “Tokenize” πρέπει να τοποθετηθεί πρώτος στην σειρά των χειριστών προ-επεξεργασίας, ώστε να διευκολυνθεί η συνέχιση της διαδικασίας της προ-επεξεργασίας. Αυτό φαίνεται και στην εικόνα 4.6, όπου αρχικά πρέπει να γίνει σύνδεση της εσωτερικής εισόδου “sin” του “Loop Collection” με την

είσοδο “doc” του “Tokenize. Αφού συνδεθούν και οι χειριστές μεταξύ τους, ενώνεται η έξοδος “doc” του “Filter Stopwords (English)” με την εσωτερική έξοδο “out” του “Loop Collection” ώστε να κλείσει ο βρόχος.



Εικόνα 4.5: Ο χειριστής “Read Document” ως υπό-διεργασία του χειριστή “Loop Files”.





Εικόνα 4.6: Οι χειριστές “Tokenize”, “Stem (Snowball)” και “Filter Stopwords (English)” ως υπό-διεργασία του χειριστή “Loop Collection”.

## 4.2 Πιλοτική αξιολόγηση σε τεχνητό σύνολο μη βιοϊατρικών κειμένων

Ως πιλοτική αξιολόγηση, πραγματοποιήθηκε η δοκιμαστική εφαρμογή της διεργασίας πάνω σε τεχνητό σύνολο μη βιοϊατρικών κειμένων, για να επαληθευτεί η ορθότητά της. Τα κείμενα που θα εισαχθούν στο πρόγραμμα έχουν αντληθεί από το Internet, είναι όλα στην αγγλική γλώσσα, και τα θέματα στα οποία τα κείμενα αυτά χωρίζονται αφορούν στις τυχαία επιλεγμένες από τη γράφουσα θεματολογίες (α) κλιματική αλλαγή, (β) ηλεκτρικό αυτοκίνητο και (γ) φύτευση λουλουδιών. Σημειώνεται ότι είναι διακριτές αλλά όχι εντελώς άσχετες μεταξύ τους θεματολογίες.

Στον φάκελο του υπολογιστή DEMO\_extra trial, αποθηκεύτηκαν 5 κείμενα από το καθένα από τα 3 ανωτέρω θέματα (15 κείμενα συνολικά). Ο χειριστής γνωρίζει εξ αρχής ότι τα 5 πρώτα ανήκουν στην κατηγορία της κλιματικής αλλαγής, τα 5 ακόλουθα στην κατηγορία του ηλεκτρικού αυτοκινήτου και τα 5 τελευταία ανήκουν στην κατηγορία της φύτευσης λουλουδιών. (εικόνα 4.7)

The screenshot shows a Windows File Explorer window with the address bar set to 'ARGYRO > UNIWA > DIPLOMA > DEMO\_extra trial'. The left sidebar shows the navigation pane with 'Κεντρική' selected. The main pane displays a list of files with columns for 'Όνομα', 'Ημερομηνία τροποποι...', 'Τύπος', and 'Μέγεθος'. The files are categorized into 'climate change', 'electric car', and 'flower planting'.

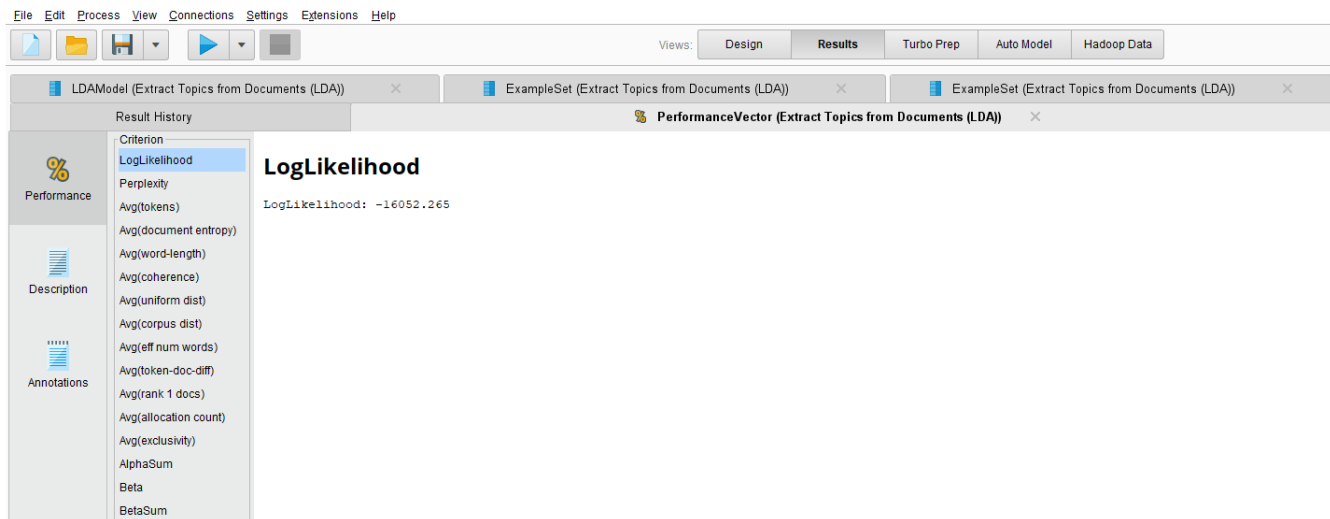
Όνομα	Ημερομηνία τροποποι...	Τύπος	Μέγεθος
climate change 1	17/7/2023 3:18 μμ	Adobe Acrobat D...	66 KB
climate change 2	17/7/2023 3:19 μμ	Adobe Acrobat D...	66 KB
climate change 3	18/7/2023 10:53 πμ	Adobe Acrobat D...	53 KB
climate change 4	18/7/2023 10:55 πμ	Adobe Acrobat D...	65 KB
climate change 5	18/7/2023 10:57 πμ	Adobe Acrobat D...	64 KB
electric car 1	17/7/2023 3:09 μμ	Adobe Acrobat D...	79 KB
electric car 2	17/7/2023 3:11 μμ	Adobe Acrobat D...	45 KB
electric car 3	17/7/2023 3:11 μμ	Adobe Acrobat D...	66 KB
electric car 4	17/7/2023 3:14 μμ	Adobe Acrobat D...	57 KB
electric car 5	17/7/2023 3:16 μμ	Adobe Acrobat D...	64 KB
flower planting 1	17/7/2023 2:41 μμ	Adobe Acrobat D...	123 KB
flower planting 2	17/7/2023 3:01 μμ	Adobe Acrobat D...	54 KB
flower planting 3	17/7/2023 3:03 μμ	Adobe Acrobat D...	46 KB
flower planting 4	17/7/2023 3:05 μμ	Adobe Acrobat D...	56 KB
flower planting 5	17/7/2023 3:07 μμ	Adobe Acrobat D...	53 KB

Εικόνα 4.7: Τα κατηγοριοποιημένα κείμενα στον προσωπικό υπολογιστή.

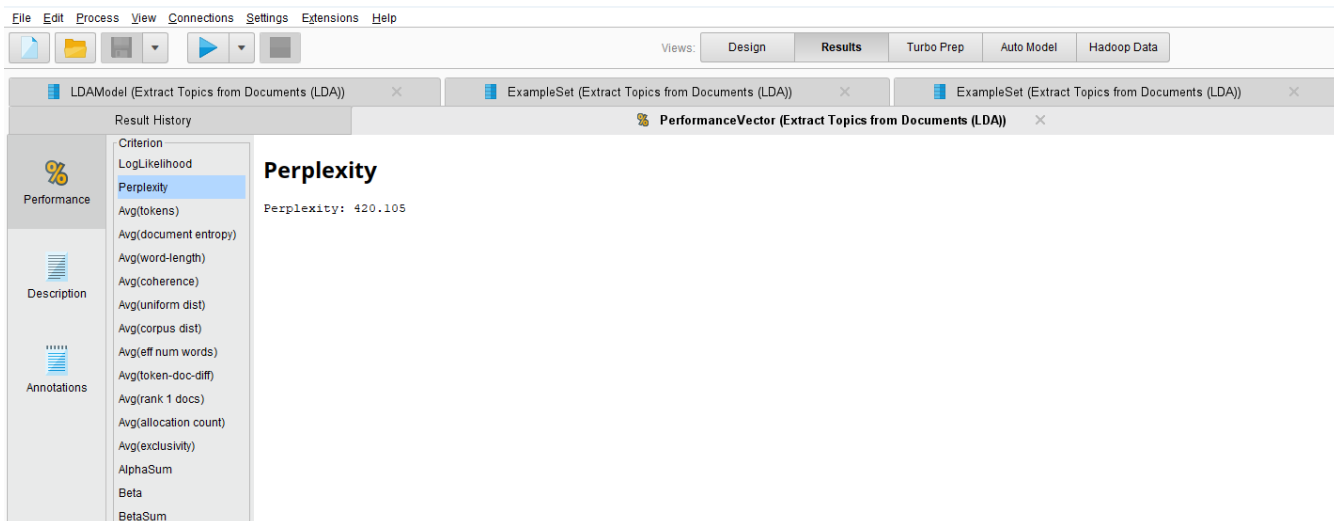
Η ορθότητα της διαδικασίας θα επαληθευτεί δίνοντας διαδοχικά έναν από τρεις διαφορετικούς αριθμούς ( $k = 2$ ,  $k = 3$ ,  $k = 4$ ) ως πλήθος των ζητούμενων «κρυμμένων» θεμάτων ως είσοδο στον χειριστή “Extract Topics from Documents (LDA)” και ζητώντας να εξάγει τα ανάλογα θέματα. Τα αποτελέσματα συγκρίνονται με τη χειροκίνητη κατηγοριοποίηση των κειμένων από τη γράφουσα. Η διεργασία θεωρείται ότι λειτουργεί σωστά όταν τα αποτελέσματα αυτόματης μέσω LDA και χειροκίνητης διεργασίας είναι παρόμοια. [8-22]

#### 4.2.1 Πιλοτική αξιολόγηση, $k = 2$ θέματα

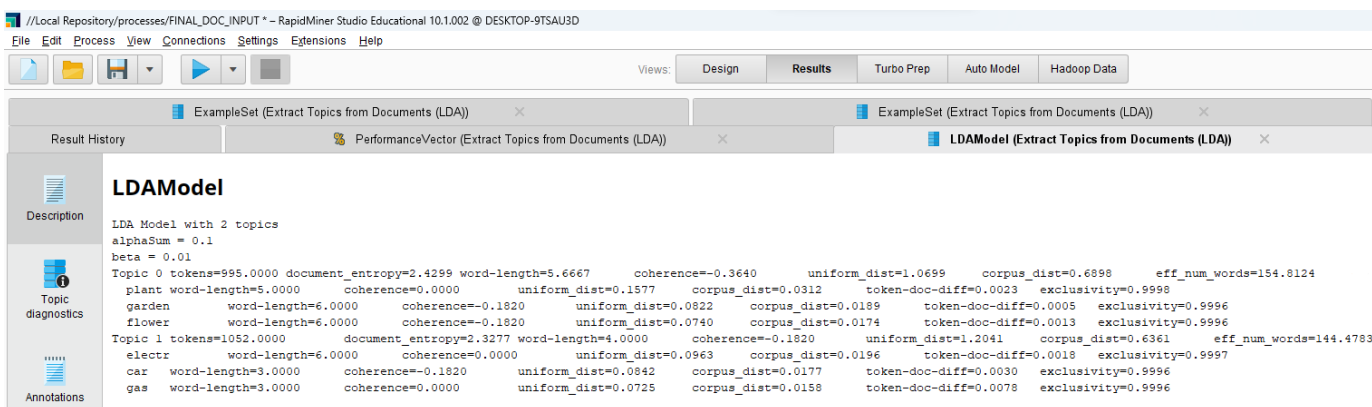
Αρχικά, το πείραμα θα ξεκινήσει ορίζοντας ως  $k = 2$  το πλήθος των ζητούμενων «κρυμμένων» θεμάτων προς ανακάλυψη. Τα αποτελέσματα που προκύπτουν μετά την εκτέλεση της διεργασίας είναι τα ακόλουθα.



Εικόνα 4.8: Η τιμή Loglikelihood για k = 2 θέματα.



Εικόνα 4.9: Η τιμή perplexity για k = 2 θέματα.



Εικόνα 4.10: Το μοντέλο LDA για k = 2 θέματα.

Row No.	topicid	word	weight
1	0	plant	43
2	0	garden	26
3	0	flower	24
4	1	electr	31
5	1	car	28
6	1	gas	25

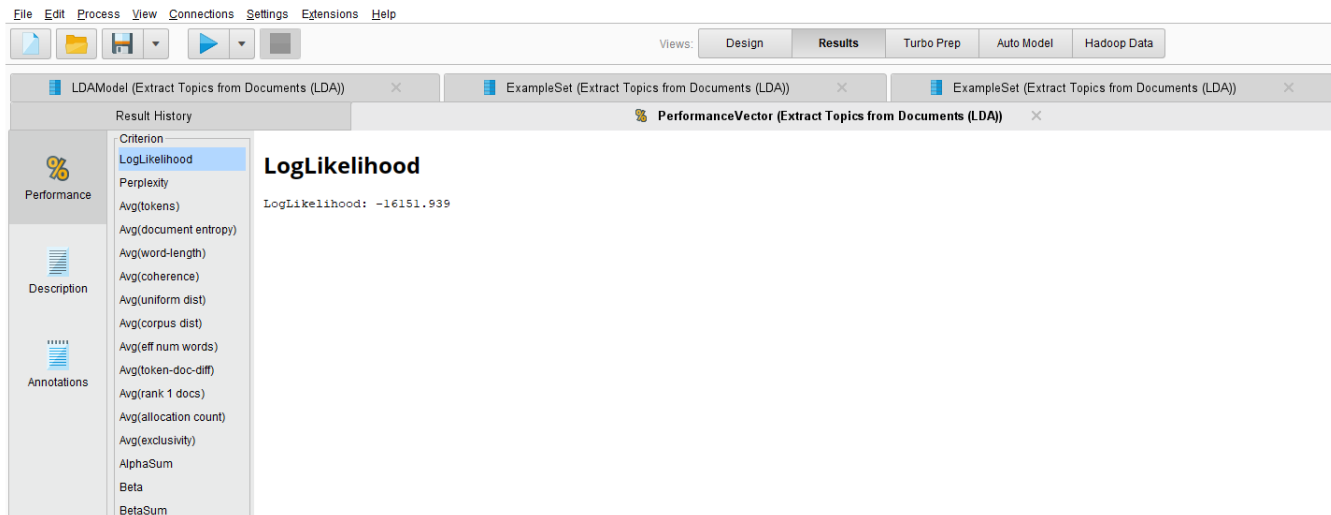
Εικόνα 4.11: Οι 3 πιο συχνές λέξεις και τα βάρη τους για  $k = 2$  θέματα.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_1	0.327	0.673	Burning fossi...
2	1	Topic_1	0.195	0.805	The Sustaina...
3	2	Topic_1	0.222	0.778	Global climat...
4	3	Topic_1	0.322	0.678	Climate chan...
5	4	Topic_1	0.266	0.734	We know why...
6	5	Topic_1	0.043	0.957	While electric...
7	6	Topic_1	0.184	0.816	Many drivers ...
8	7	Topic_1	0.270	0.730	You're at the r...
9	8	Topic_1	0.140	0.860	There are ple...
10	9	Topic_1	0.337	0.663	In 2008 I bou...
11	10	Topic_0	1.000	0.000	March is an e...
12	11	Topic_0	0.960	0.040	Weeding and...
13	12	Topic_0	1.000	0.000	As a beginnin...
14	13	Topic_0	1.000	0.000	Flower garde...
15	14	Topic_0	0.960	0.040	Successful b...

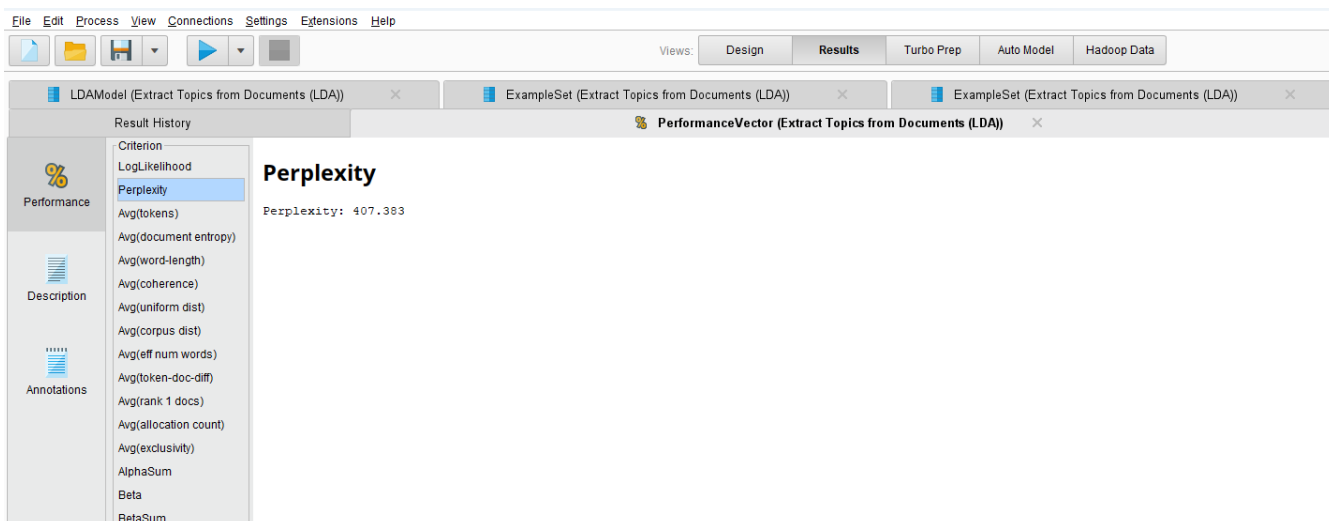
Εικόνα 4.12: Η κατηγοριοποίηση των κειμένων σε  $k = 2$  θέματα.

#### 4.2.2 Πιλοτική αξιολόγηση, $k = 3$ θέματα

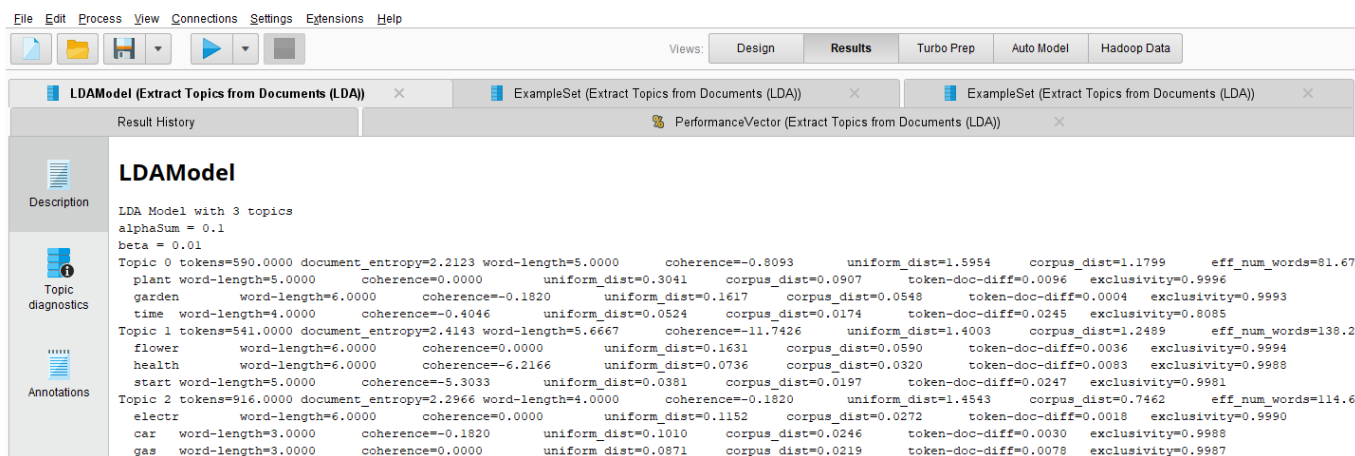
Αλλάζοντας τον αριθμό των θεμάτων σε  $k = 3$ , εξάγονται τα παρακάτω αποτελέσματα.



Εικόνα 4.13: Η τιμή Loglikelihood για k = 3 θέματα.



Εικόνα 4.14: Η τιμή perplexity για k = 3 θέματα.



Εικόνα 4.15: Το μοντέλο LDA για k = 3 θέματα.

Row No.	topicid	word	weight
1	0	plant	43
2	0	garden	26
3	0	time	11
4	1	flower	24
5	1	health	13
6	1	start	8
7	2	electr	31
8	2	car	28
9	2	gas	25

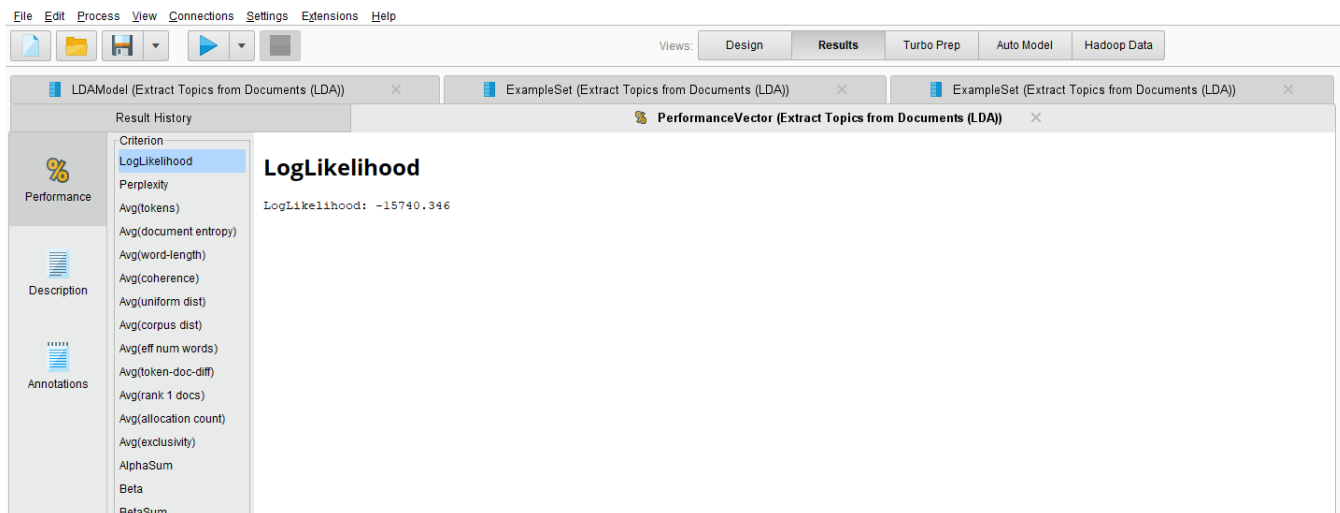
Εικόνα 4.16: Οι 3 πιο συχνές λέξεις και τα βάρη τους για  $k = 3$  θέματα.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_2	0.427	0.096	0.476	Burning fossi...
2	1	Topic_2	0.000	0.292	0.708	The Sustaina...
3	2	Topic_2	0.206	0.082	0.712	Global climat...
4	3	Topic_2	0.057	0.407	0.536	Climate chan...
5	4	Topic_2	0.426	0.000	0.574	We know why...
6	5	Topic_2	0.024	0.000	0.976	While electric...
7	6	Topic_2	0.045	0.292	0.664	Many drivers ...
8	7	Topic_2	0.000	0.376	0.623	You're at the r...
9	8	Topic_2	0.000	0.079	0.921	There are ple...
10	9	Topic_2	0.319	0.124	0.556	In 2008 I bou...
11	10	Topic_0	0.514	0.486	0.000	March is an e...
12	11	Topic_0	0.572	0.390	0.038	Weeding and...
13	12	Topic_0	0.713	0.287	0.000	As a beginnin...
14	13	Topic_1	0.420	0.580	0.000	Flower garde...
15	14	Topic_1	0.484	0.516	0.000	Successful b...

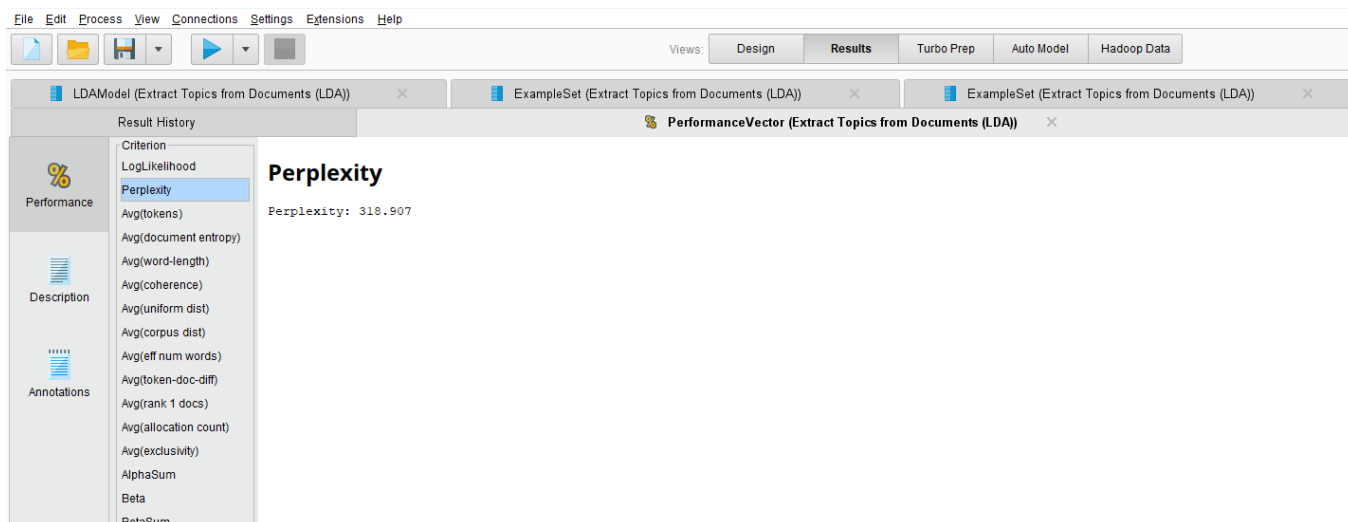
Εικόνα 4.17: Η κατηγοριοποίηση των κειμένων σε  $k = 3$  θέματα.

### 4.2.3 Πιλοτική αξιολόγηση, $k = 4$ θέματα

Τέλος, ορίζεται ο αριθμός θεμάτων σε  $k = 4$  και τα αποτελέσματα παρουσιάζονται παρακάτω.

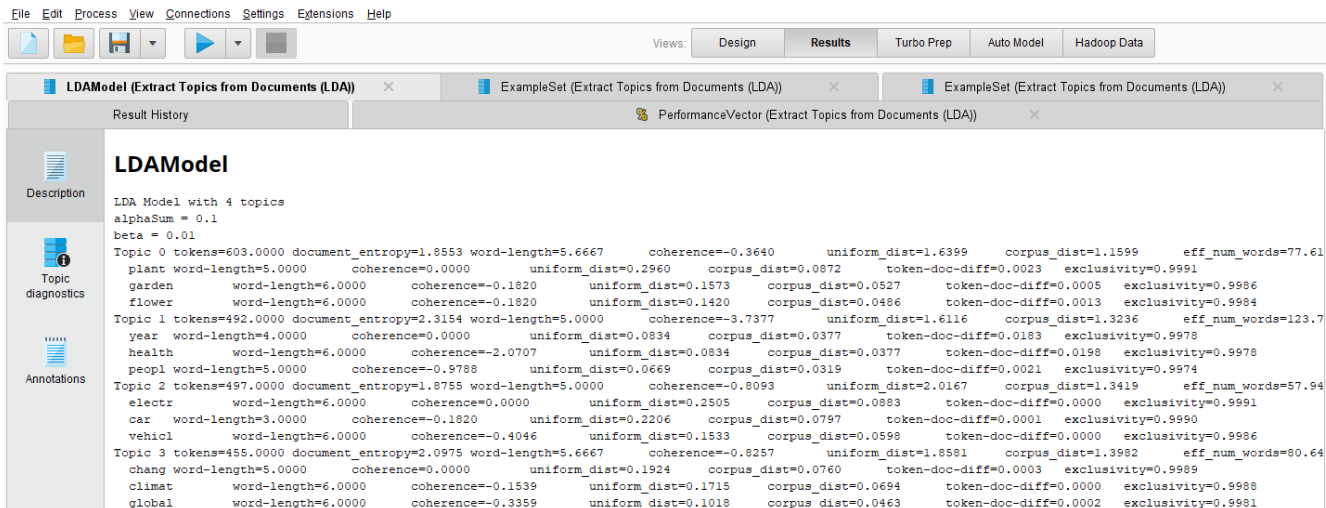


Εικόνα 4.18: Η τιμή Loglikelihood για  $k = 4$  θέματα.

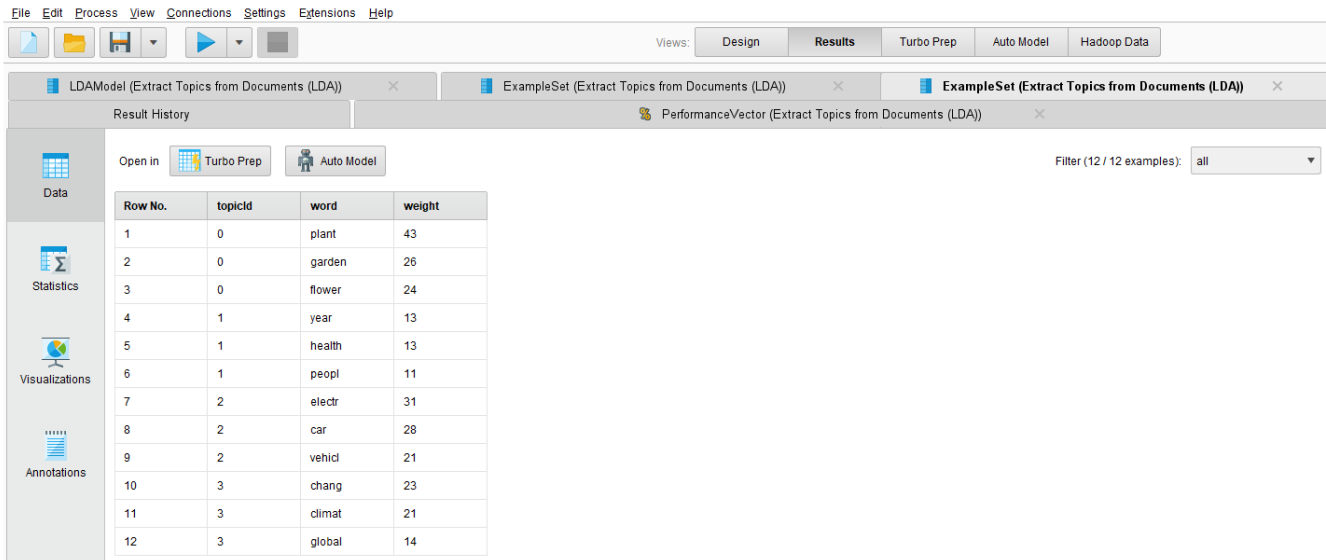


Εικόνα 4.19: Η τιμή perplexity για  $k = 4$  θέματα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»



Εικόνα 4.20: Το μοντέλο LDA για k = 4 θέματα.



Εικόνα 4.21: Οι 3 πιο συχνές λέξεις και τα βάρη τους για k = 4 θέματα.



Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_3	0.000	0.310	0.078	0.612	Burning fossi...
2	1	Topic_1	0.000	0.487	0.328	0.185	The Sustaina...
3	2	Topic_3	0.043	0.123	0.006	0.829	Global climat...
4	3	Topic_1	0.000	0.544	0.000	0.455	Climate chan...
5	4	Topic_1	0.000	0.593	0.018	0.389	We know why...
6	5	Topic_2	0.000	0.000	0.718	0.282	While electric...
7	6	Topic_2	0.169	0.114	0.718	0.000	Many drivers ...
8	7	Topic_2	0.068	0.276	0.656	0.000	You're at the r...
9	8	Topic_2	0.000	0.222	0.721	0.057	There are ple...
10	9	Topic_2	0.176	0.221	0.603	0.000	In 2008 I bou...
11	10	Topic_0	0.821	0.126	0.000	0.053	March is an e...
12	11	Topic_0	0.745	0.188	0.000	0.066	Weeding and...
13	12	Topic_0	0.883	0.000	0.000	0.117	As a beginnin...
14	13	Topic_0	0.842	0.062	0.000	0.096	Flower garde...
15	14	Topic_0	0.796	0.121	0.000	0.082	Successful b...

Εικόνα 4.22: Η κατηγοριοποίηση των κειμένων σε k = 4 θέματα.

#### 4.2.4 Πιλοτική αξιολόγηση, σχολιασμός αποτελεσμάτων

Στον επόμενο Πίνακα 4.1 συνοψίζονται συγκριτικά για τα τρία διαφορετικά πλήθη θεμάτων k = 2, 3, 4 του ανωτέρω παραδείγματος όλες οι παράμετροι του μοντέλου.

Πίνακας 4.1: Οι παράμετροι των μοντέλων για 3 διαφορετικά εμπειρικά k.

Παράμετρος μοντέλου	k = 2	k = 3	k = 4
<b>loglikelihood</b>	-16052.265	-16151.939	-15740.346
<b>perplexity</b>	420.105	407.383	318.907
<b>alphaSum</b>	0.1	0.1	0.1
<b>beta</b>	0.01	0.01	0.01
<b>tokens</b>	<ul style="list-style-type: none"> <li>Topic_0: 995</li> <li>Topic_1: 052</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 590</li> <li>Topic_1: 541</li> <li>Topic_2: 916</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 603</li> <li>Topic_1: 492</li> <li>Topic_2: 497</li> <li>Topic_3: 455</li> </ul>
<b>document_entropy</b>	<ul style="list-style-type: none"> <li>Topic_0: 2.429</li> <li>Topic_1: 2.327</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 2.212</li> <li>Topic_1: 2.414</li> <li>Topic_2: 2.296</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 1.855</li> <li>Topic_1: 2.315</li> <li>Topic_2: 1.875</li> <li>Topic_3: 2.097</li> </ul>
<b>word_length</b>	<ul style="list-style-type: none"> <li>Topic_0: 5.667</li> <li>Topic_1: 4.000</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 5.00</li> <li>Topic_1: 5.66</li> </ul>	<ul style="list-style-type: none"> <li>Topic_0: 5.667</li> <li>Topic_1: 5.000</li> </ul>

		<ul style="list-style-type: none"> <li>• Topic_2: 4.00</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_2: 5.000</li> <li>• Topic_3: 5.667</li> </ul>
<b>coherence</b>	<ul style="list-style-type: none"> <li>• Topic_0: -0.364</li> <li>• Topic_1: -0.182</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: -0.809</li> <li>• Topic_1: -11.742</li> <li>• Topic_2: -0.182</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: -0.364</li> <li>• Topic_1: -3.737</li> <li>• Topic_2: -0.809</li> <li>• Topic_3: -0.825</li> </ul>
<b>uniform_dist</b>	<ul style="list-style-type: none"> <li>• Topic_0: 1.069</li> <li>• Topic_1: 1.204</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 1.595</li> <li>• Topic_1: 1.400</li> <li>• Topic_2: 1.454</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 1.639</li> <li>• Topic_1: 1.611</li> <li>• Topic_2: 2.016</li> <li>• Topic_3: 1.858</li> </ul>
<b>corpus_dist</b>	<ul style="list-style-type: none"> <li>• Topic_0: 0.689</li> <li>• Topic_1: 0.636</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 1.179</li> <li>• Topic_1: 1.248</li> <li>• Topic_2: 0.746</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 1.159</li> <li>• Topic_1: 1.312</li> <li>• Topic_2: 1.341</li> <li>• Topic_3: 1.398</li> </ul>
<b>eff_num_words</b>	<ul style="list-style-type: none"> <li>• Topic_0: 154.812</li> <li>• Topic_1: 144.478</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 81.67</li> <li>• Topic_1: 138.23</li> <li>• Topic_2: 114.68</li> </ul>	<ul style="list-style-type: none"> <li>• Topic_0: 77.61</li> <li>• Topic_1: 123.75</li> <li>• Topic_2: 55.94</li> <li>• Topic_3: 80.64</li> </ul>

Αρχικά συγκρίνουμε τις τιμές των παραμέτρων loglikelihood και perplexity. Η εξέταση αυτού του ζεύγους τιμών, θα οδηγήσει στην απόφαση επιλογής καλύτερου μοντέλου. Γενικά, καθώς το  $k$  πλησιάζει τη βέλτιστη τιμή, η τιμή perplexity μειώνεται ενώ η τιμή Loglikelihood αυξάνεται (βλ. Πίνακα 4.1). Στο βέλτιστο μοντέλο, οι τιμές αυτές θα είναι αντιστρόφως ανάλογες, δηλαδή όσο μεγαλύτερη είναι η τιμή loglikelihood, τόσο μικρότερη θα είναι η τιμή perplexity. Συγκρίνοντας τις εικόνες 4.8-4.9, 4.13-4.14 και 4.18-4.19, διαπιστώνεται ότι το βέλτιστο μοντέλο LDA είναι αυτό των  $k = 4$  θεμάτων, με τιμή Loglikelihood ίση με -15740.346 και με τιμή perplexity ίση με 318.907.

Οι εικόνες 4.10, 4.15 και 4.20 παρουσιάζουν το εκπαιδευμένο μοντέλο. Οι παράμετροι alphaSum και beta αφορούν τις Bayes υπερπαραμέτρους σχετικά με την κατανομή θέματος και την κατανομή λέξης αντίστοιχα και έχουν τις ίδιες τιμές και στους 3 διαφορετικούς αριθμούς θεμάτων που δοκιμάστηκαν εδώ. Η τιμή alphaSum = 0.1 υποδηλώνει τον αριθμό των θεμάτων που αναγνωρίζονται από το μοντέλο. Η χαμηλή τιμή αντιστοιχεί στο γεγονός ότι το σύνολο των κειμένων είναι μικρό και άρα τα αναδυόμενα θέματα συγκεντρώνονται σε αυτό και δεν εμφανίζονται άσχετα θέματα. Η τιμή beta = 0.01 είναι ο έλεγχος συσχέτισης των λέξεων με τα θέματα. Η χαμηλή τιμή της υπερπαραμέτρου δείχνει ότι το μοντέλο LDA έδωσε

περισσότερο βάρος στις σπάνιες λέξεις, κάνοντάς τις πιο σημαντικές για την ανάλυση θεμάτων.

Επιπλέον, εμφανίζονται οι σημαντικές λέξεις και ο συνολικός αριθμός “tokens” σε κάθε μία. Η παράμετρος “document\_entropy” αναφέρεται στον υπολογισμό της εντροπίας των εγγράφων, δηλαδή πόσο εσωτερικά περίπλοκα είναι τα κείμενα. Στους 3 διαφορετικούς αριθμούς θεμάτων, η τιμή της εντροπίας είναι χαμηλή, δηλαδή τα κείμενα δεν είναι ιδιαίτερα εσωτερικά περίπλοκα και συγκεκριμένα οι μικρότερες τιμές εμφανίζονται όταν ο αριθμός των θεμάτων είναι  $k = 4$ .

Η παράμετρος “word-length” σημειώνει στο εκάστοτε θέμα το μέσο μήκος των λέξεων ενώ η παράμετρος “coherence” αξιολογεί την ποιότητα των θεμάτων που έχουν εξαχθεί και αντιστοιχηθεί από το LDA μοντέλο. Οι τιμές της παραμέτρου “coherence” του μοντέλου  $k = 4$  θεμάτων είναι οι μεγαλύτερες σε σχέση με τα άλλα δύο μοντέλα  $k = 2$  και  $k = 3$ , καθώς πλησιάζουν πιο κοντά στο 0 και αυτό υποδηλώνει ότι υπάρχει υψηλή συνάφεια, δηλαδή οι λέξεις μέσα σε κάθε θέμα σχετίζονται στενά μεταξύ τους.

Η τιμή της παραμέτρου “uniform\_dist” δηλώνει την ομοιόμορφη κατανομή των θεμάτων στα έγγραφα, δηλαδή τα θέματα ανατίθενται σε όλα τα κείμενα με παρόμοια συχνότητα. Οι μεγαλύτερες τιμές αυτής της παραμέτρου εμφανίζονται στο μοντέλο 4 θεμάτων, συγκριτικά με τα άλλα δύο. Συνεχίζοντας, η “corpus\_dist” αναφέρεται στην κατανομή των θεμάτων αλλά στο «σώμα» των κειμένων. Η μικρή τιμή της “corpus\_dist” δηλώνει ότι τα θέματα που έχουν ανακαλυφθεί είναι πιο εξειδικευμένα και κατανέμονται πιο συγκεκριμένα στα έγγραφα ενώ η μεγάλη τιμή δείχνει ότι τα θέματα είναι πιο γενικά που είναι περισσότερο ανεξάρτητα από τα έγγραφα. Το μοντέλο των 4 θεμάτων στην πλειονότητά του μικρές τιμές της παραμέτρου “corpus\_dist” άρα τα εξαγόμενα θέματα είναι πιο εξειδικευμένα με τα έγγραφα. Η τελευταία παράμετρος, “eff\_num\_words”, αντιπροσωπεύει τον αποτελεσματικό αριθμό των λέξεων που χρησιμοποιούνται για ανάλυση. Η τιμή της, αν είναι υψηλή ή χαμηλή, δεν μπορεί να συμβάλει από μόνη της στην επιλογή του καλύτερου μοντέλου LDA αλλά σε συνδυασμό με την ισορροπημένη κατανομή θεμάτων στα έγγραφα, πως ανταποκρίνεται στους στόχους του πειράματος και αν διατηρεί τον αποδεκτό υπολογιστικό φόρτο, θα βοηθήσει στην διάκριση του βέλτιστου μοντέλου. Σύμφωνα με τις πληροφορίες που έχουν δοθεί, κανένα από τα 3 μοντέλα δεν έχει απαιτήσει πολλούς πόρους, η απάντηση στο αν ανταποκρίνεται στους στόχους του πειράματος θα δοθεί παρακάτω, ενώ αναφέρθηκε ότι το μοντέλο 4 θεμάτων έχει την καλύτερη κατανομή.

Κάποια από τα αποτελέσματα που προέκυψαν από την εκτέλεση της διεργασίας, εμφανίζονται στις εικόνες 4.11, 4.16 και 4.21 και αναγράφονται οι λέξεις των εκάστοτε θεμάτων και τα βάρη τους, δηλαδή τις φορές που επαναλαμβάνονται σε όλο το «σώμα». Αρχικά, επιβεβαιώνεται ότι έχει γίνει και η διαδικασία της προ-επεξεργασίας και συγκεκριμένα το κομμάτι “Stemming”, όπου γίνεται διαχωρισμός των στελεχών των λέξεων και των καταλήξεων. Τα πιο εμφανή στελέχη είναι “reopl”, “electr”, “vehicl”, “chang” και “climat”. Με βάση τους παραπάνω πίνακες, παρατηρείται ότι όσο αυξάνεται ο αριθμός των θεμάτων οι λέξεις παραμένουν σχεδόν ίδιες ενώ προστίθενται κι άλλες. Αξιοσημείωτο, επίσης, είναι ότι τα βάρη δεν μεταβάλλονται από το ένα μοντέλο στο άλλο. Συγκρίνοντας τα αναδυόμενα θέματα με την εικόνα 4.7, μπορεί να γίνει η επιβεβαίωση ότι και τα τρία μοντέλα έχουν ανακαλύψει τα ζητούμενα θέματα.

Το τελικό συμπέρασμα σχετικά με την επιβεβαίωση της ορθής λειτουργίας της διεργασίας, θα προκύψει από τον συνδυασμό των πληροφοριών που δόθηκαν και των εικόνων 4.12, 4.17 και 4.22.

Διαπιστώνουμε αρχικά ότι η τιμή  $k = 2$  είναι ανεπαρκής διότι διακρίνει σωστά τα 5 κείμενα του 3<sup>ο</sup> θέματος (Flower planting) και δίνει και τις σωστές κύριες λέξεις για να ‘ονοματιστεί’ το topic\_0 (plant, garden flower). Όμως τα  $5+5 = 10$  κείμενα των δύο πρώτων θεμάτων (Climate change, Electric Car) τα ενώνει υποχρεωτικά σε ένα θέμα, το topic\_1, δίνοντας ως κύριες λέξεις ένα μείγμα (electr, car, gas).

Αυξάνοντας το  $k$  σε 3, διαπιστώνουμε ότι δεν λειτούργησε όπως αναμενόταν. Συγκεκριμένα ένωσε τις δύο πρώτες ομάδες (Climate Change, Electric Car) σε ένα topic\_2, με κύριες λέξεις (electric, car, gas) και υποδιαίρεσε τα 5 κείμενα της ομάδας Flower Planting σε δύο topics, το topic\_0 (plant, garden, time) και στο topic\_1 (flower, health, start).

Αυξάνοντας το  $k$  σε 4, διαπιστώνεται ότι η κατανομή των κειμένων σε 4 θέματα είναι η πλησιέστερη προς την χειροκίνητη («ορθή») κατανομή. Με βάση την ορθή κατανομή των 15 κειμένων στα 3 θέματα Climate Change (κείμενα 1-5), Electric Car (κείμενα 6-10), Flower Planting (κείμενα 11-15), όπως φαίνονται στην εικόνα 4.7, αξιολογούμε την εικόνα 4.22. Οι δύο τελευταίες ομάδες Electric Car και Flower Planting, των 5 κειμένων η καθεμία, έχουν κατηγοριοποιηθεί αυτόματα όπως ακριβώς είναι κατηγοριοποιημένες και χειροκίνητα. Αντίθετα, τα 5 πρώτα κείμενα της ομάδας Climate Change, που θα έπρεπε να είναι μαζί, έχουν χωριστεί στα δύο εναπομείναντα θέματα topic\_1 (κύριες λέξεις year, health, people) και topic\_3 (κύριες λέξεις chang, climate, global). Όλες αυτές οι λέξεις αυτές θεωρούνται σχετικές

με το θέμα της κλιματικής αλλαγής, αλλά δεν είναι σωστή η εσωτερική υποδιαίρεση της ομάδας.

Σύμφωνα με τις παραπάνω αναλύσεις και τα αριθμητικά δεδομένα, το μοντέλο των  $k = 4$  θεμάτων θεωρείται το καλύτερο εκ των τριών. Διαισθητικά, το μοντέλο των 3 θεμάτων αναμενόταν να είναι το καταλληλότερο και βάσει αυτού να γινόταν η αξιολόγηση της διεργασίας. Όμως, όπως φαίνεται και στην εικόνα 4.17, το  $k = 3$  δεν έδωσε το αποτέλεσμα το πλησιέστερο στην χειροκίνητη (επιθυμητή) κατηγοριοποίηση των κειμένων, και συνεπώς δεν επιλέγεται.

Συμπερασματικά, με βάση τα αποτελέσματα που δίνει το μοντέλο για  $k = 4$  θέματα, διαπιστώνεται ότι η δομημένη διεργασία λειτουργεί αρκετά ικανοποιητικά και μπορεί να χρησιμοποιηθεί στα ακόλουθα πειράματα.

### 4.3 Εφαρμογή σε περιλήψεις βιοϊατρικών κείμενων και αποτελέσματα

Η επιβεβαίωση της ορθής λειτουργίας της διεργασίας που αναπτύχθηκε, όπως έγινε στην *Ενότητα 4.2*, δίνει το πράσινο φως για την εκτέλεση των βασικών πειραμάτων της διπλωματικής εργασίας. Στην παρούσα ενότητα, αρχικά αντλούνται από τη βάση PubMed τα κείμενα και οι περιλήψεις βιοϊατρικών κειμένων. Στη συνέχεια θα εισαχθούν στην διεργασία του RapidMiner και θα προκύψουν αποτελέσματα σχετικά με 3 διαφορετικά πλήθη θεμάτων (topic numbers), που ορίστηκαν εμπειρικά ως  $k = 2$ ,  $k = 3$  και  $k = 4$ .

#### 4.3.1 Αναζήτηση και άντληση βιοϊατρικών κειμένων

Τα βιοϊατρικά κείμενα αντλήθηκαν από την βάση PubMed ήδη κατηγοριοποιημένα σε 3 κατηγορίες, οι οποίες είναι η Clinical Trial I, Clinical Trial II και Controlled Clinical Trial.

Η αναζήτηση έγινε με βάση τα keywords “Alzheimer OR Alzheimer’s”, με φίλτρο “In TITLE/ABSTRACT only”. Αναζητήθηκαν άρθρα σε περιοδικά, στην αγγλική γλώσσα, δημοσιευμένα από το 2020 μέχρι το 2023. Ο τελευταίος περιορισμός τέθηκε και για πρακτικούς λόγους, για να μην χρειαστεί να χειριστεί το πρόγραμμα έναν υπέρογκο αριθμό άρθρων, αλλά και για ουσιαστικούς (δημοσιεύσεις επηρεασμένες από την πανδημία Covid-19).

Τα σύνολο των αρχικών κειμένων ήταν  $N_0 = 367$ , και σε κάθε κατηγορία είναι:

- Clinical Trial I = 18 άρθρα,
- Clinical Trial II = 38 άρθρα και
- Controlled Clinical Trial = 311 άρθρα,

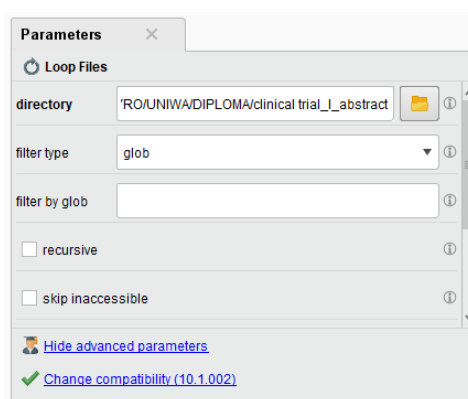
από τα οποία απομονώθηκαν οι περιλήψεις τους. Όπως παρατηρήθηκε, μερικά από τα προαναφερθέντα κείμενα δεν είχαν περιλήψεις, οπότε για να επιτευχθεί η ομαλή λειτουργία του πειράματος και να εξαχθούν στα σωστά αποτελέσματα αφαιρέθηκαν τα κείμενα αυτά από τις κατηγορίες Clinical Trial I και Controlled Clinical Trial και ο τελικός αριθμός των κειμένων περιορίστηκε στα  $N_1 = 362$  συνολικά και ανά κατηγορία, αντίστοιχα:

- Clinical Trial I = 17 άρθρα,
- Clinical Trial II = 38 άρθρα και
- Controlled Clinical Trial = 307 άρθρα,

Παρακάτω, παρατίθενται τα αποτελέσματα από την χρήση του RapidMiner, των οποίων ο σχολιασμός θα γίνει στην ενότητα 4.5.

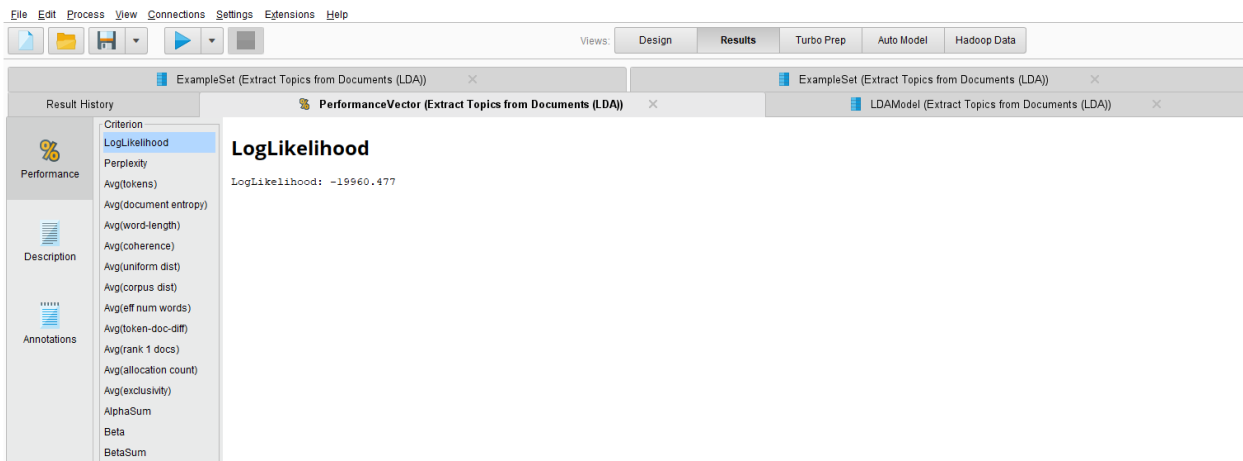
#### 4.3.2 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας άρθρων Clinical Trial I

Στην εικόνα 4.23, στις παραμέτρους του χειριστή “Loop Files” έχει επιλεγθεί ο φάκελος “clinical\_trial\_I\_abstract”, όπου εμπεριέχονται οι περιλήψεις των κειμένων αυτής της κατηγορίας.

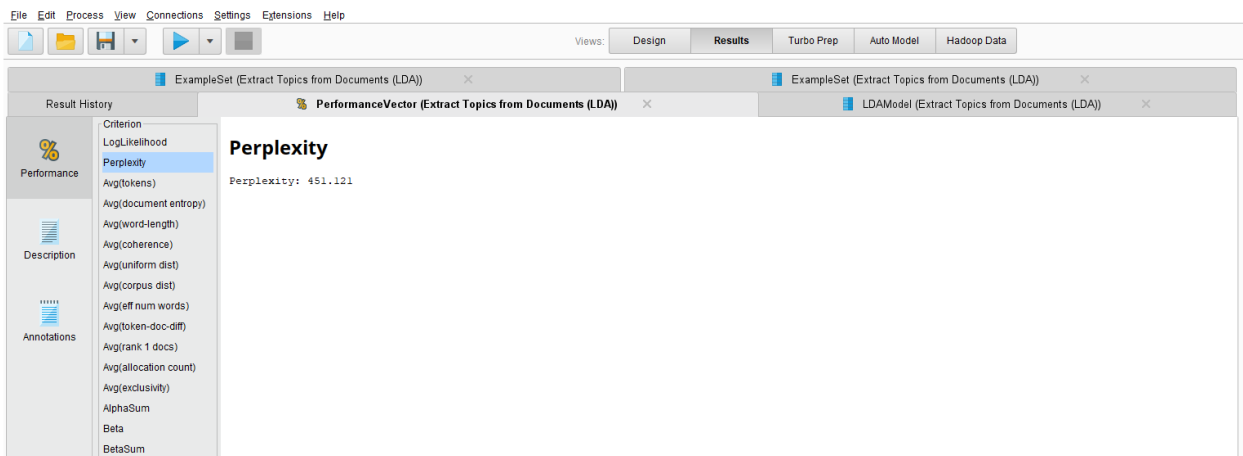


Εικόνα 4.23: Αλλαγή directory για εύρεση αρχείων περιλήψεων των άρθρων της ομάδας Clinical Trial I.

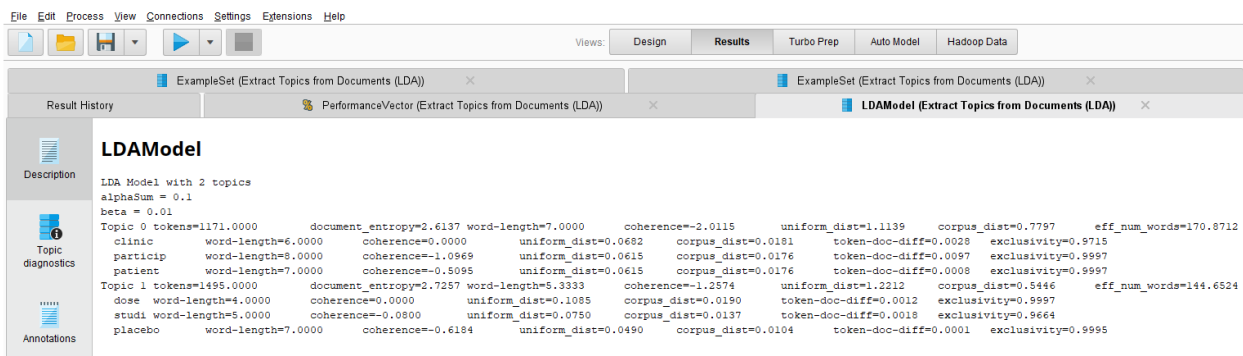
Αποτελέσματα για αριθμό θεμάτων  $k = 2$ :



Εικόνα 4.24: Η τιμή Loglikelihood για  $k = 2$  θέματα, Clinical Trial I - περιλήψεις.



Εικόνα 4.25: Η τιμή Perplexity για  $k = 2$  θέματα, Clinical Trial I – περιλήψεις.



Εικόνα 4.26: Το μοντέλο LDA για  $k = 2$  θέματα, Clinical Trial I – περιλήψεις.

Row No.	topicId	word	weight
1	0	clinic	27
2	0	particip	25
3	0	patient	25
4	1	dose	49
5	1	studi	37
6	1	placebo	27

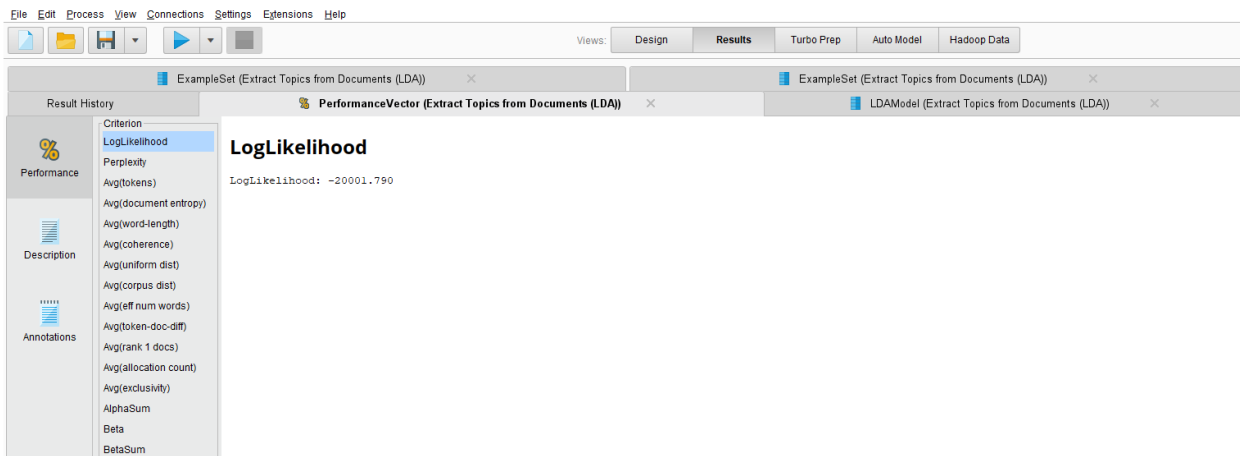
Εικόνα 4.27: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για  $k = 2$  θέματα, Clinical Trial I – περιλήψεις.

Row No.	documentId	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_0	0.841	0.159	TREM2 is a receptor for lipi...
2	1	Topic_0	0.678	0.322	BACKGROUND: Older co...
3	2	Topic_1	0.370	0.630	Research has shown that ...
4	3	Topic_1	0.305	0.695	RATIONALE: The M1M4 pr...
5	4	Topic_0	0.599	0.401	BACKGROUND: Crenezu...
6	5	Topic_1	0.155	0.845	PURPOSE: Safety, tolerabi...
7	6	Topic_1	0.152	0.848	PURPOSE: Ganteneruma...
8	7	Topic_0	0.654	0.346	IMPORTANCE: Individuals ...
9	8	Topic_0	0.723	0.277	OBJECTIVE: PhytoSERM i...
10	9	Topic_1	0.151	0.849	The present study evaluate...
11	10	Topic_0	0.619	0.381	BACKGROUNDS: Alzheim...
12	11	Topic_0	0.778	0.222	Aducanumab is a human i...
13	12	Topic_1	0.383	0.617	RIPK1 is a master regulat...
14	13	Topic_1	0.252	0.748	INTRODUCTION: Brain bio...
15	14	Topic_1	0.109	0.891	BACKGROUND: Fosgoni...
16	15	Topic_1	0.402	0.598	HYPOTHESIS: We hypothe...
17	16	Topic_1	0.163	0.837	Nerve growth factor (NGF) ...

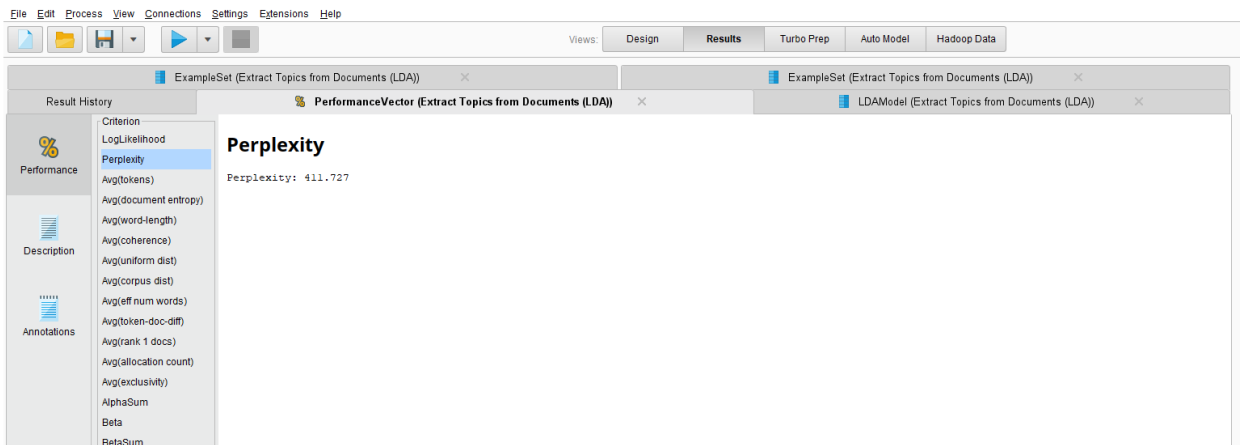
Εικόνα 4.28: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε  $k = 2$  θέματα.



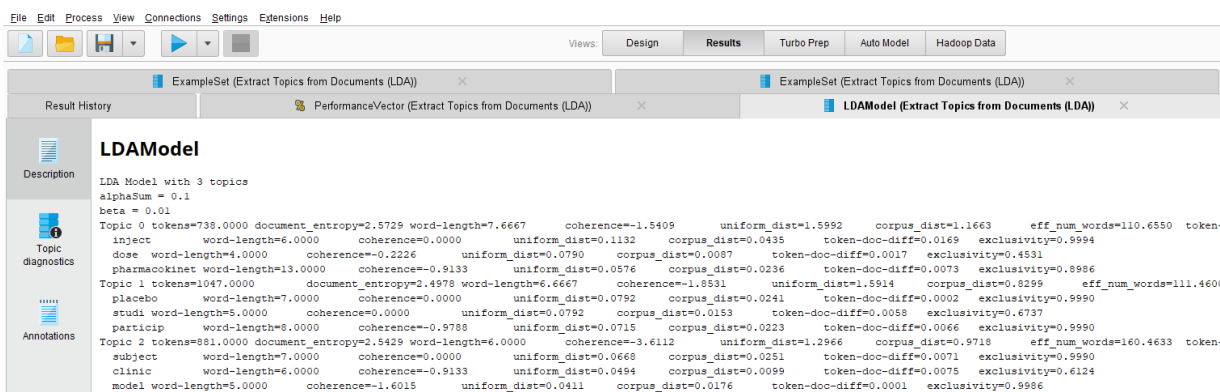
Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :



Εικόνα 4.29: Η τιμή Loglikelihood για  $k = 3$  θέματα, Clinical Trial I – περιλήψεις.



Εικόνα 4.30: Η τιμή Perplexity για  $k = 3$  θέματα, Clinical Trial I – περιλήψεις.



Εικόνα 4.31: Το μοντέλο LDA για  $k = 3$  θέματα, Clinical Trial I – περιλήψεις.

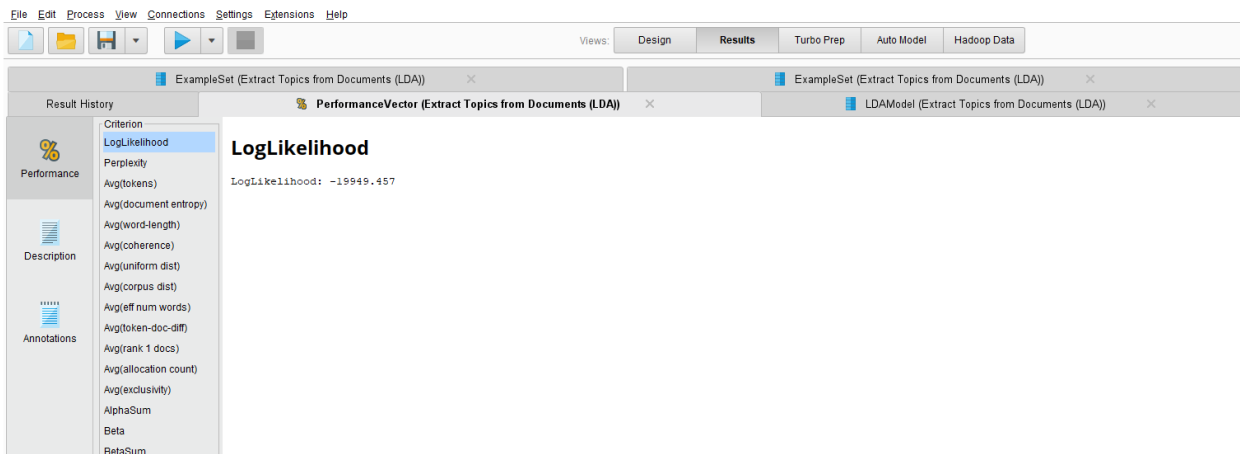
Row No.	topicid	word	weight
1	0	inject	25
2	0	dose	19
3	0	pharmacokinetic	15
4	1	placebo	27
5	1	study	27
6	1	particip	25
7	2	subject	20
8	2	clinic	16
9	2	model	14

Εικόνα 4.32: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 3$  θέματα, Clinical Trial I – περιλήψεις.

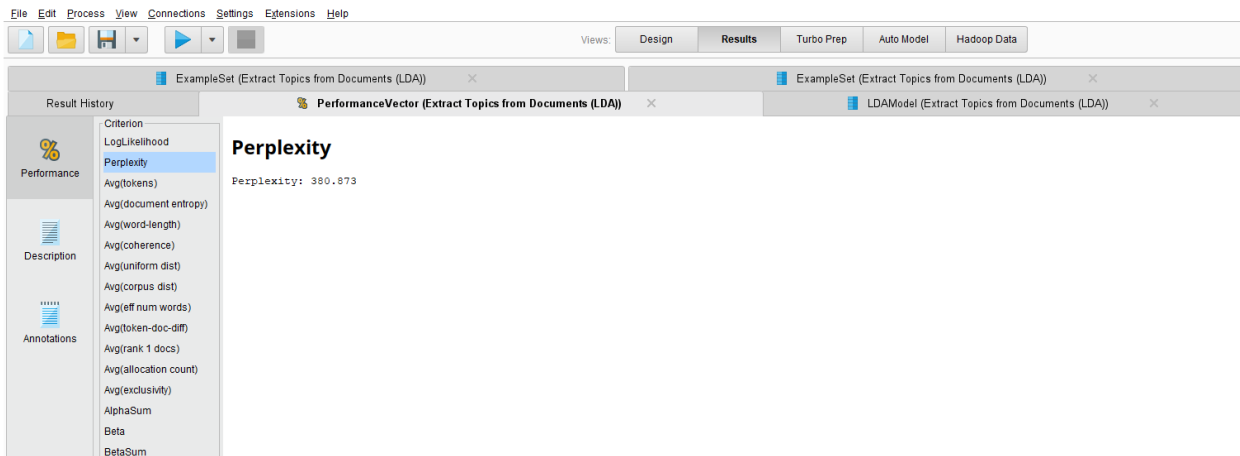
Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_2	0.244	0.092	0.664	TREM2 is a r...
2	1	Topic_2	0.068	0.305	0.627	BACKGROU...
3	2	Topic_1	0.392	0.442	0.166	Research ha...
4	3	Topic_0	0.446	0.113	0.442	RATIONALE: ...
5	4	Topic_1	0.110	0.576	0.314	BACKGROU...
6	5	Topic_2	0.188	0.277	0.535	PURPOSE: S...
7	6	Topic_0	0.695	0.305	0.000	PURPOSE: G...
8	7	Topic_1	0.092	0.850	0.058	IMPORTANC...
9	8	Topic_1	0.310	0.543	0.148	OBJECTIVE: ...
10	9	Topic_2	0.169	0.089	0.742	The present ...
11	10	Topic_1	0.140	0.860	0.000	BACKGROU...
12	11	Topic_2	0.132	0.302	0.566	Aducanumab...
13	12	Topic_0	0.540	0.174	0.286	RIPK1 is a m...
14	13	Topic_0	0.749	0.187	0.064	INTRODUCTI...
15	14	Topic_2	0.170	0.354	0.476	BACKGROU...
16	15	Topic_1	0.040	0.533	0.427	HYPOTHESI...
17	16	Topic_2	0.343	0.056	0.602	Nerve growth ...

Εικόνα 4.33: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε  $k = 3$  θέματα.

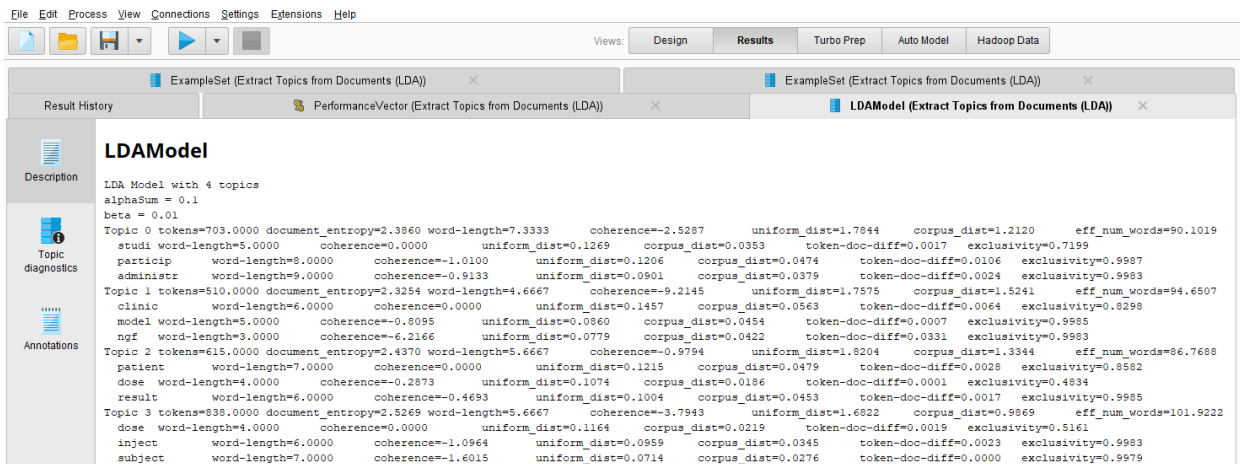
Αποτελέσματα για k = 4 θέματα:



Εικόνα 4.34: Η τιμή Loglikelihood για k = 4 θέματα, Clinical Trial I – περιλήψεις.



Εικόνα 4.35: Η τιμή Perplexity για k = 4 θέματα, Clinical Trial I – περιλήψεις.



Εικόνα 4.36: Το μοντέλο LDA για k = 4 θέματα, Clinical Trial I – περιλήψεις.

Row No.	topicid	word	weight
1	0	studi	26
2	0	particip	25
3	0	administr	20
4	1	clinic	21
5	1	model	14
6	1	ngf	13
7	2	patient	22
8	2	dose	20
9	2	result	19
10	3	dose	29
11	3	inject	25
12	3	subject	20

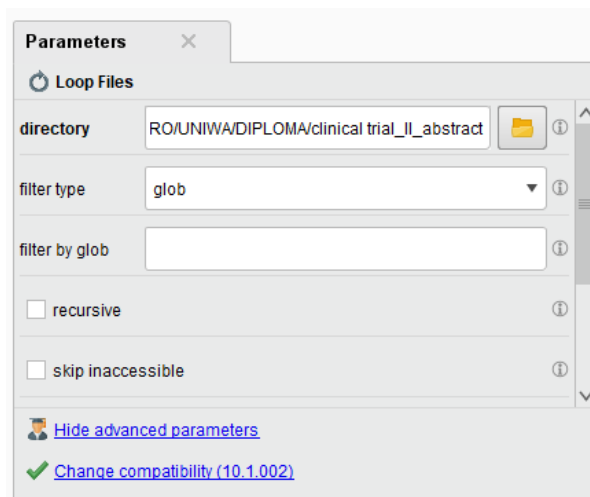
Εικόνα 4.37: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 4$  θέματα, Clinical Trial I – περιλήψεις.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_1	0.416	0.554	0.028	0.001	TREM2 is a r...
2	1	Topic_2	0.001	0.398	0.455	0.147	BACKGROU...
3	2	Topic_3	0.132	0.003	0.298	0.567	Research ha...
4	3	Topic_3	0.000	0.341	0.049	0.609	RATIONALE:...
5	4	Topic_0	0.693	0.049	0.123	0.135	BACKGROU...
6	5	Topic_3	0.447	0.000	0.088	0.464	PURPOSE: S...
7	6	Topic_3	0.199	0.001	0.077	0.723	PURPOSE: G...
8	7	Topic_0	0.544	0.045	0.327	0.084	IMPORTANC...
9	8	Topic_0	0.469	0.249	0.072	0.209	OBJECTIVE:...
10	9	Topic_3	0.150	0.053	0.001	0.796	The present...
11	10	Topic_2	0.302	0.105	0.432	0.161	BACKGROU...
12	11	Topic_3	0.348	0.199	0.000	0.453	Aducanumab...
13	12	Topic_2	0.147	0.134	0.400	0.319	RIPK1 is a m...
14	13	Topic_2	0.095	0.118	0.535	0.252	INTRODUCTL...
15	14	Topic_3	0.104	0.060	0.416	0.420	BACKGROU...
16	15	Topic_1	0.170	0.482	0.337	0.011	HYPOTHESI...
17	16	Topic_1	0.000	0.758	0.154	0.087	Nerve growth...

Εικόνα 4.38: Η κατηγοριοποίηση των περιλήψεων Clinical Trial I σε  $k = 4$  θέματα.

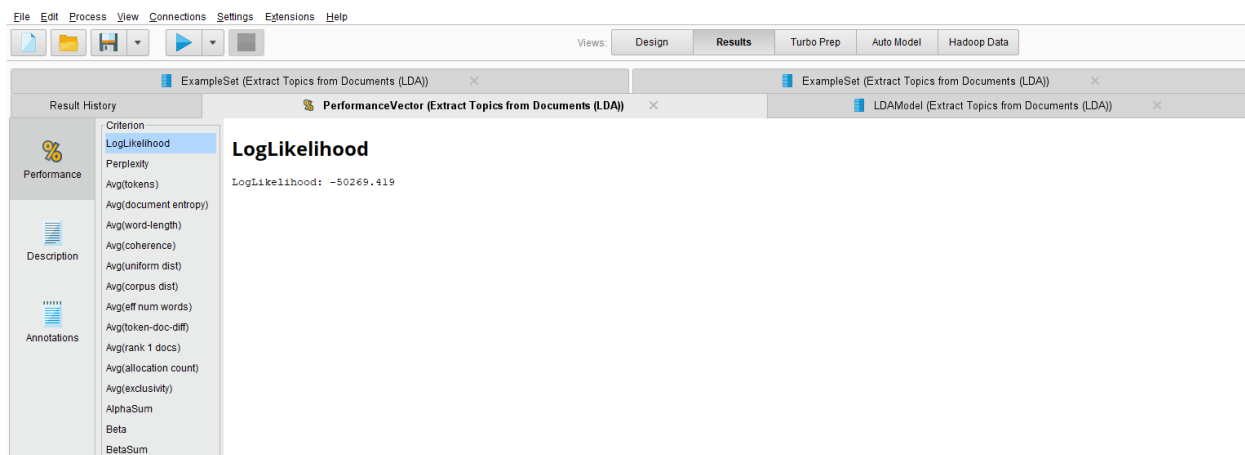
### 4.3.3 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας άρθρων Clinical Trial II

Στην εικόνα 4.39 η παράμετρος directory του “Loop Files” έχει αλλαχθεί και έχει επιλεγεί το αρχείο “clinical\_trial\_II\_abstract” ώστε να αναγνωριστούν τα κείμενα Clinical Trial II.

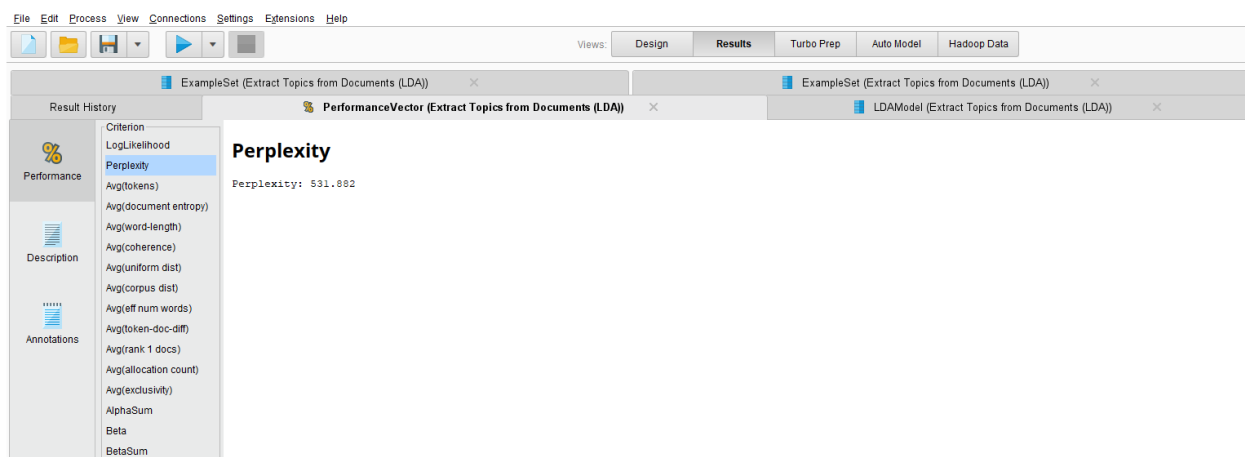


Εικόνα 4.39: Αλλαγή directory για εύρεση αρχείων περιλήψεων των άρθρων της ομάδας Clinical Trial II.

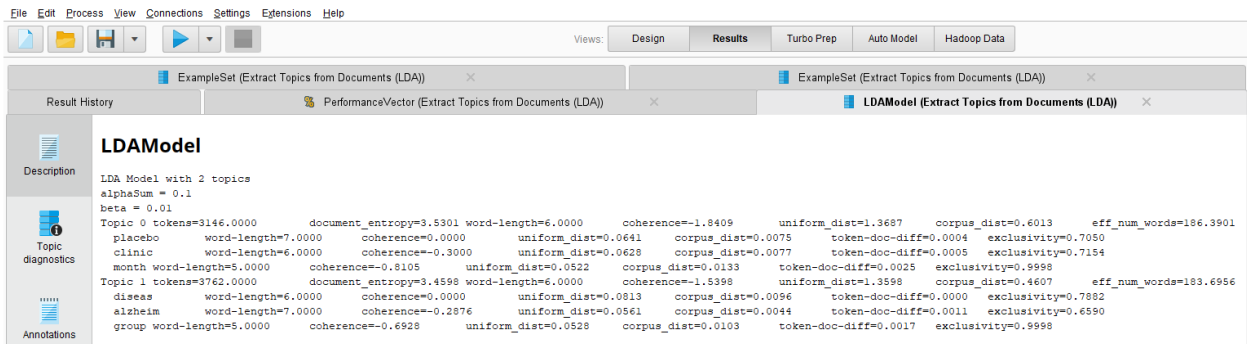
Αποτελέσματα για αριθμό θεμάτων  $k = 2$ :



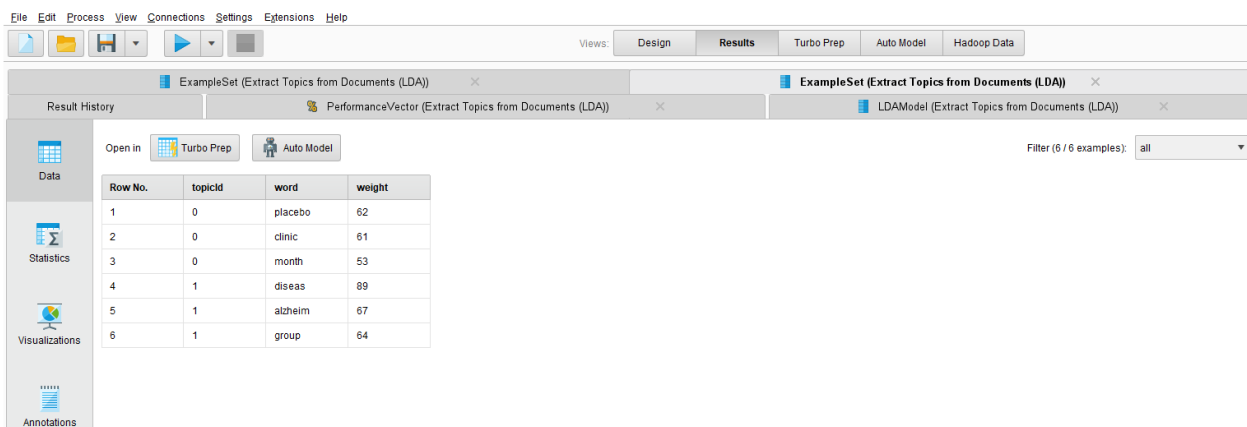
Εικόνα 4.40: Η τιμή Loglikelihood για  $k = 2$  θέματα, Clinical Trial II – περιλήψεις.



Εικόνα 4.41: Η τιμή Perplexity για  $k = 2$  θέματα, Clinical Trial II – περιλήψεις.



Εικόνα 4.42: Το μοντέλο LDA 2 θέματα Clinical Trial II – περιλήψεις.



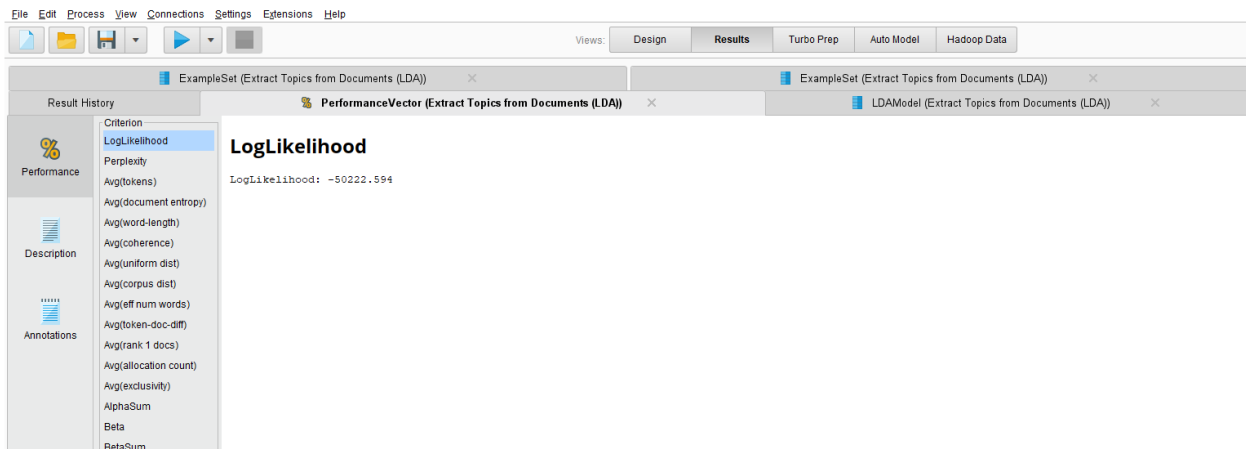
Εικόνα 4.43: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για k=2 θέματα, Clinical Trial II – περιλήψεις.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

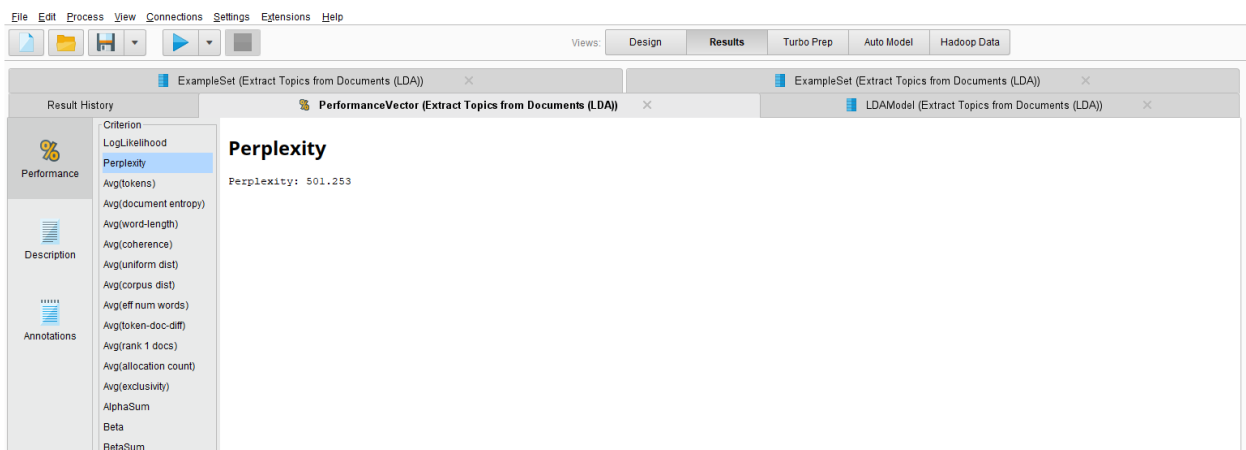
Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_0	0.623	0.377	BACKGROU...
2	1	Topic_1	0.414	0.586	BACKGROU...
3	2	Topic_0	0.833	0.167	BACKGROU...
4	3	Topic_0	0.614	0.386	INTRODUCT...
5	4	Topic_1	0.330	0.670	BACKGROU...
6	5	Topic_1	0.206	0.794	Dysregulatio...
7	6	Topic_1	0.434	0.566	BACKGROU...
8	7	Topic_0	0.815	0.185	Aducanumab...
9	8	Topic_1	0.461	0.539	IMPORTANC...
10	9	Topic_0	0.810	0.190	RIPK1 is a m...
11	10	Topic_1	0.172	0.828	Tau accumul...
12	11	Topic_1	0.129	0.871	Repetitive tra...
13	12	Topic_1	0.325	0.675	BACKGROU...
14	13	Topic_0	0.643	0.357	BACKGROU...
15	14	Topic_0	0.884	0.136	INTRODUCT...
16	15	Topic_1	0.246	0.754	BACKGROU...
17	16	Topic_1	0.246	0.754	BACKGROU...
18	17	Topic_0	0.643	0.357	INTRODUCT...
19	18	Topic_0	0.705	0.295	BACKGROU...
20	19	Topic_1	0.394	0.606	BACKGROU...
21	20	Topic_1	0.138	0.862	Imaging bio...
22	21	Topic_0	0.767	0.233	BACKGROU...
23	22	Topic_0	0.629	0.371	BACKGROU...
24	23	Topic_0	0.657	0.343	BACKGROU...
25	24	Topic_1	0.336	0.664	BACKGROU...
26	25	Topic_1	0.250	0.750	The COVID-1...
27	26	Topic_0	0.889	0.111	The muscari...
28	27	Topic_0	0.902	0.098	BACKGROU...
29	28	Topic_0	0.669	0.331	
30	29	Topic_0	0.625	0.375	OBJECTIVE ...
31	30	Topic_1	0.208	0.792	IMPORTANC...
32	31	Topic_1	0.173	0.827	Background ...
33	32	Topic_0	0.531	0.469	BACKGROU...
34	33	Topic_1	0.148	0.852	The blood-br...
35	34	Topic_1	0.428	0.572	Preclinical st...
36	35	Topic_1	0.162	0.838	IMPORTANC...
37	36	Topic_1	0.320	0.680	IMPORTANC...
38	37	Topic_1	0.463	0.537	The locus co...

Εικόνα 4.44: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε k = 2 θέματα.

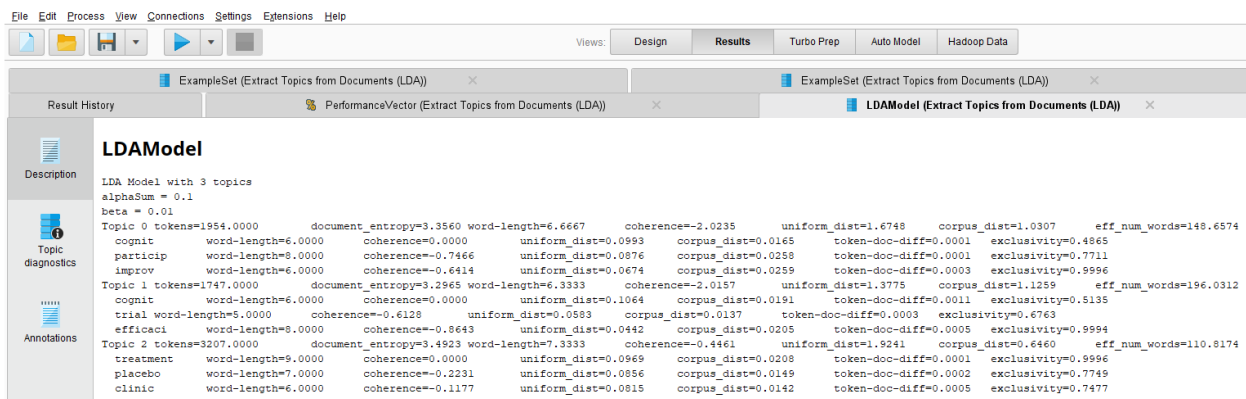
Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :



Εικόνα 4.45: Η τιμή Loglikelihood για  $k = 3$  θέματα, Clinical Trial II – περιλήψεις.



Εικόνα 4.46: Η τιμή Perplexity για  $k = 3$  θέματα, Clinical Trial II – περιλήψεις



Εικόνα 4.47: Το μοντέλο LDA για  $k = 3$  θέματα, Clinical Trial II – περιλήψεις.



«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

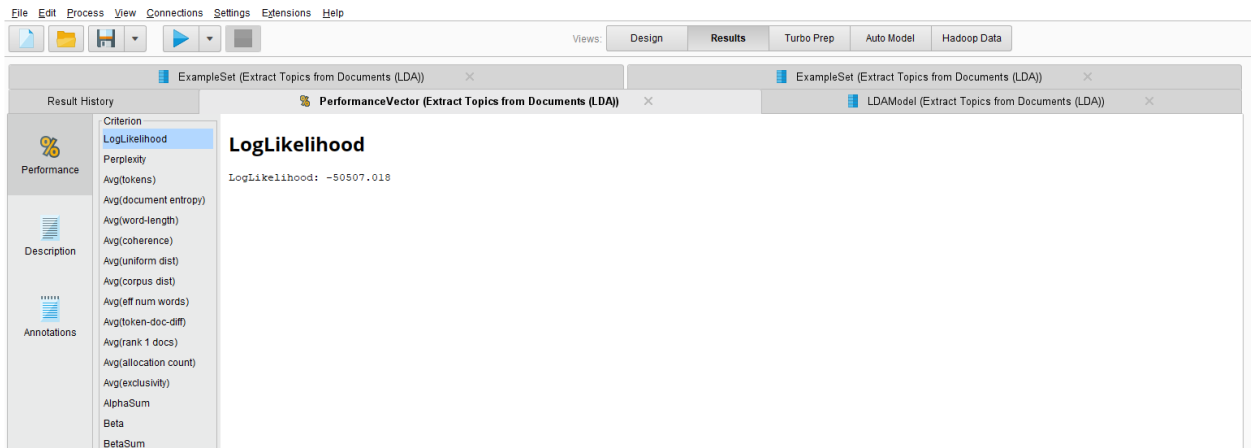
Row No.	topicid	word	weight
1	0	cognit	54
2	0	particip	49
3	0	improv	40
4	1	cognit	51
5	1	trial	32
6	1	efficaci	26
7	2	treatment	87
8	2	placebo	79
9	2	clinic	76

Εικόνα 4.48: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 3$  θέματα, Clinical Trial II – περιλήψεις.

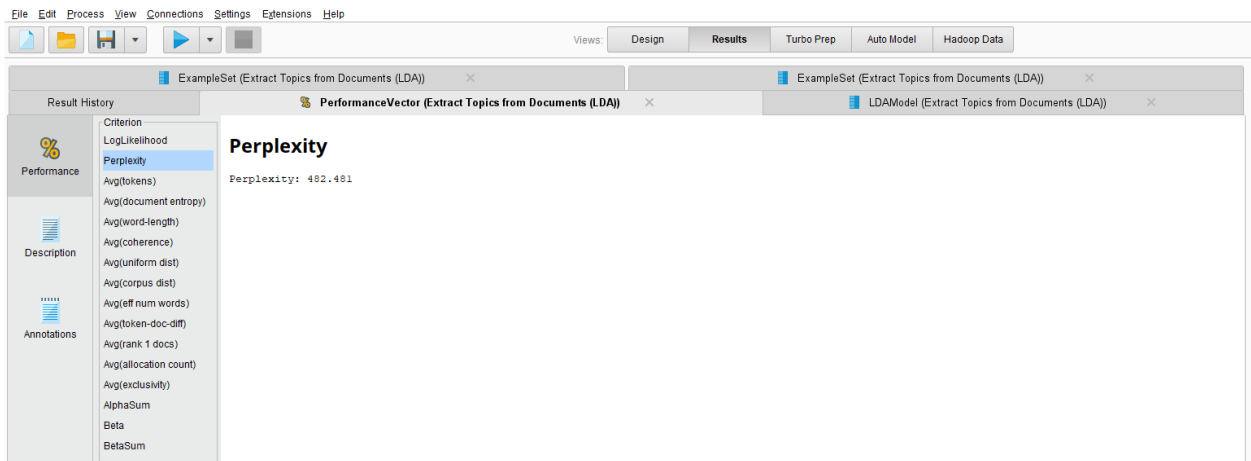
Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_2	0.115	0.053	0.832	BACKGROU...
2	1	Topic_1	0.087	0.463	0.450	BACKGROU...
3	2	Topic_2	0.121	0.181	0.698	BACKGROU...
4	3	Topic_2	0.159	0.000	0.841	INTRODUCT...
5	4	Topic_2	0.250	0.117	0.633	BACKGROU...
6	5	Topic_0	0.635	0.099	0.266	Dysregulato...
7	6	Topic_2	0.210	0.114	0.675	BACKGROU...
8	7	Topic_1	0.133	0.507	0.360	Aducanumab...
9	8	Topic_2	0.276	0.090	0.634	IMPORTANC...
10	9	Topic_1	0.133	0.719	0.148	RIPK1 is a m...
11	10	Topic_2	0.119	0.085	0.797	Tau accumu...
12	11	Topic_2	0.000	0.234	0.766	Repetitive tra...
13	12	Topic_0	0.529	0.102	0.370	BACKGROU...
14	13	Topic_2	0.117	0.394	0.490	BACKGROU...
15	14	Topic_0	0.524	0.000	0.475	INTRODUCT...
16	15	Topic_0	0.577	0.000	0.422	BACKGROU...
17	16	Topic_0	0.577	0.000	0.422	BACKGROU...
18	17	Topic_2	0.157	0.270	0.573	INTRODUCT...
19	18	Topic_1	0.138	0.607	0.255	BACKGROU...
20	19	Topic_2	0.129	0.053	0.818	BACKGROU...
21	20	Topic_0	0.505	0.275	0.220	Imaging bio...
22	21	Topic_1	0.302	0.572	0.127	BACKGROU...
23	22	Topic_2	0.252	0.040	0.708	BACKGROU...
24	23	Topic_1	0.123	0.537	0.340	BACKGROU...
25	24	Topic_1	0.314	0.514	0.172	BACKGROU...
26	25	Topic_0	0.581	0.418	0.000	The COVID-1...
27	26	Topic_2	0.354	0.170	0.476	The muscul...
28	27	Topic_0	0.516	0.139	0.345	BACKGROU...
29	28	Topic_0	0.543	0.133	0.324	BACKGROU...
30	29	Topic_1	0.125	0.722	0.152	OBJECTIVE: ...
31	30	Topic_2	0.431	0.126	0.442	IMPORTANC...
32	31	Topic_1	0.072	0.468	0.460	Background ...
33	32	Topic_2	0.173	0.326	0.501	BACKGROU...
34	33	Topic_2	0.062	0.300	0.638	The blood-br...
35	34	Topic_1	0.121	0.598	0.281	Predimical st...
36	35	Topic_2	0.268	0.221	0.511	IMPORTANC...
37	36	Topic_2	0.308	0.156	0.536	IMPORTANC...
38	37	Topic_0	0.582	0.072	0.346	The locus co...

Εικόνα 4.49: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε  $k = 3$  θέματα.

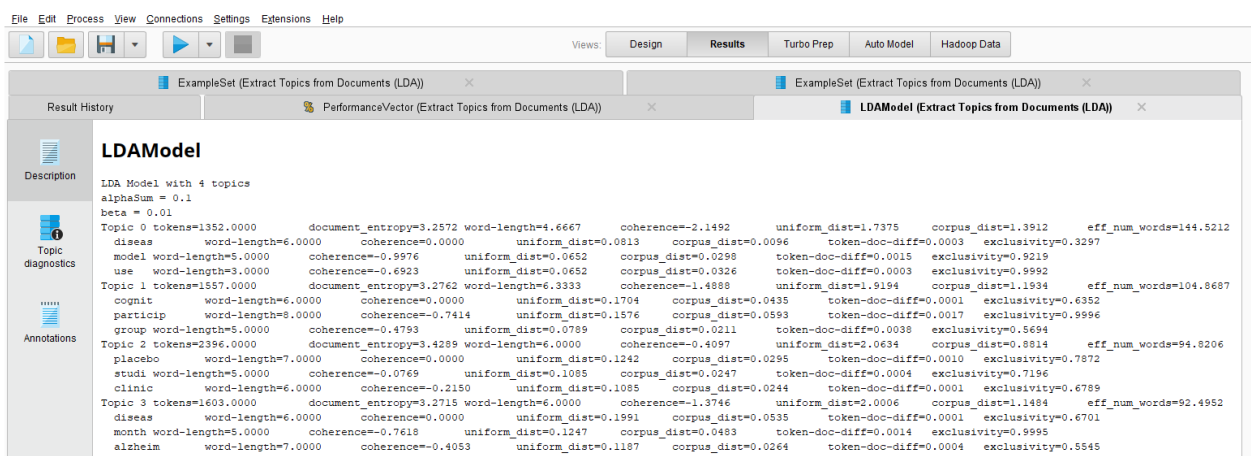
Αποτελέσματα για αριθμό θεμάτων  $k = 4$ :



Εικόνα 4.50: Η τιμή Loglikelihood για  $k = 4$  θέματα, Clinical Trial II – περιλήψεις.



Εικόνα 4.51: Η τιμή Perplexity για  $k = 4$  θέματα, Clinical Trial II – περιλήψεις.



Εικόνα 4.52: Το μοντέλο LDA για  $k = 4$  θέματα, Clinical Trial II – περιλήψεις.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Row No.	topicid	word	weight
1	0	diseas	32
2	0	model	27
3	0	use	27
4	1	cognit	66
5	1	particip	62
6	1	group	36
7	2	placebo	79
8	2	studi	71
9	2	clinic	71
10	3	diseas	77
11	3	month	53
12	3	alzheim	51

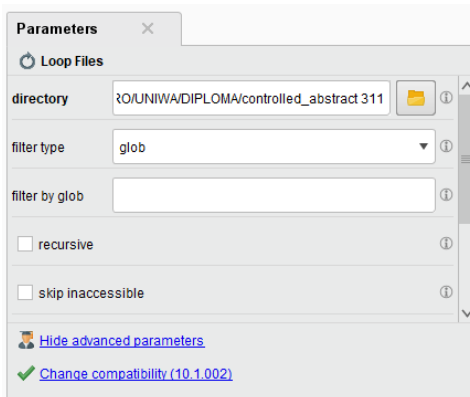
Εικόνα 4.53: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 4$  θέματα, Clinical Trial II – περιλήψεις.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_2	0.208	0.000	0.478	0.314	BACKGROU...
2	1	Topic_3	0.157	0.278	0.196	0.369	BACKGROU...
3	2	Topic_2	0.333	0.000	0.438	0.228	BACKGROU...
4	3	Topic_2	0.000	0.000	0.666	0.333	INTRODUCTL...
5	4	Topic_2	0.063	0.078	0.539	0.320	BACKGROU...
6	5	Topic_3	0.080	0.158	0.161	0.600	Dysregulatio...
7	6	Topic_2	0.030	0.085	0.532	0.353	BACKGROU...
8	7	Topic_0	0.600	0.145	0.255	0.000	Aducanumab...
9	8	Topic_2	0.057	0.351	0.499	0.093	IMPORTANC...
10	9	Topic_0	0.704	0.094	0.085	0.117	RIPK1 is a m...
11	10	Topic_2	0.077	0.165	0.485	0.272	Tau accumul...
12	11	Topic_3	0.039	0.076	0.417	0.469	Repetitive tra...
13	12	Topic_3	0.024	0.205	0.260	0.510	BACKGROU...
14	13	Topic_2	0.197	0.194	0.375	0.234	BACKGROU...
15	14	Topic_2	0.053	0.042	0.596	0.309	INTRODUCTL...
16	15	Topic_2	0.063	0.209	0.682	0.045	BACKGROU...
17	16	Topic_2	0.063	0.209	0.682	0.045	BACKGROU...
18	17	Topic_2	0.131	0.000	0.623	0.245	INTRODUCTL...
19	18	Topic_1	0.266	0.573	0.097	0.064	BACKGROU...
20	19	Topic_2	0.000	0.226	0.594	0.180	BACKGROU...
21	20	Topic_3	0.065	0.008	0.107	0.819	Imaging bio...
22	21	Topic_1	0.113	0.726	0.016	0.145	BACKGROU...
23	22	Topic_2	0.124	0.057	0.713	0.105	BACKGROU...
24	23	Topic_3	0.230	0.329	0.066	0.374	BACKGROU...
25	24	Topic_1	0.115	0.578	0.186	0.121	BACKGROU...
26	25	Topic_0	0.564	0.210	0.226	0.000	The COVID-1...
27	26	Topic_1	0.030	0.465	0.440	0.065	The muscan...
28	27	Topic_2	0.364	0.065	0.548	0.023	BACKGROU...
29	28	Topic_1	0.194	0.669	0.108	0.029	
30	29	Topic_1	0.000	0.666	0.227	0.107	OBJECTIVE: ...
31	30	Topic_1	0.118	0.404	0.144	0.335	IMPORTANC...
32	31	Topic_0	0.567	0.104	0.095	0.234	Background ...
33	32	Topic_0	0.312	0.142	0.300	0.246	BACKGROU...
34	33	Topic_0	0.432	0.097	0.086	0.386	The blood-br...
35	34	Topic_0	0.376	0.271	0.092	0.261	Preclinical st...
36	35	Topic_2	0.321	0.323	0.356	0.000	IMPORTANC...
37	36	Topic_2	0.147	0.249	0.514	0.090	IMPORTANC...
38	37	Topic_3	0.336	0.043	0.257	0.365	The locus co...

Εικόνα 4.54: Η κατηγοριοποίηση των περιλήψεων Clinical Trial II σε  $k = 4$  θέματα.

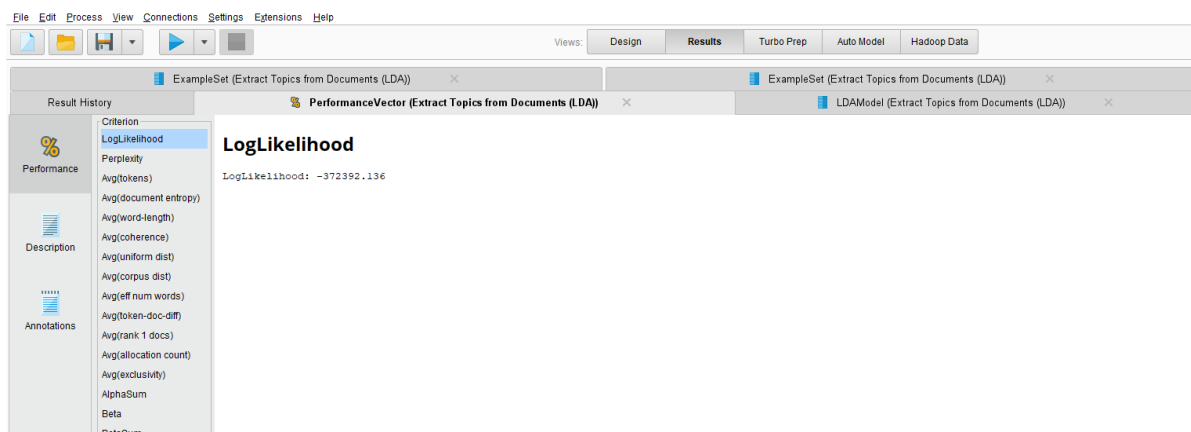
### 4.3.4 Ανάλυση και αποτελέσματα περιλήψεων της ομάδας Controlled Clinical Trial

Γίνεται η αλλαγή του directory του “Loop Files” σε “controlled\_abstract 311” ώστε να εισαχθούν οι περιλήψεις των κειμένων Controlled Clinical Trial.

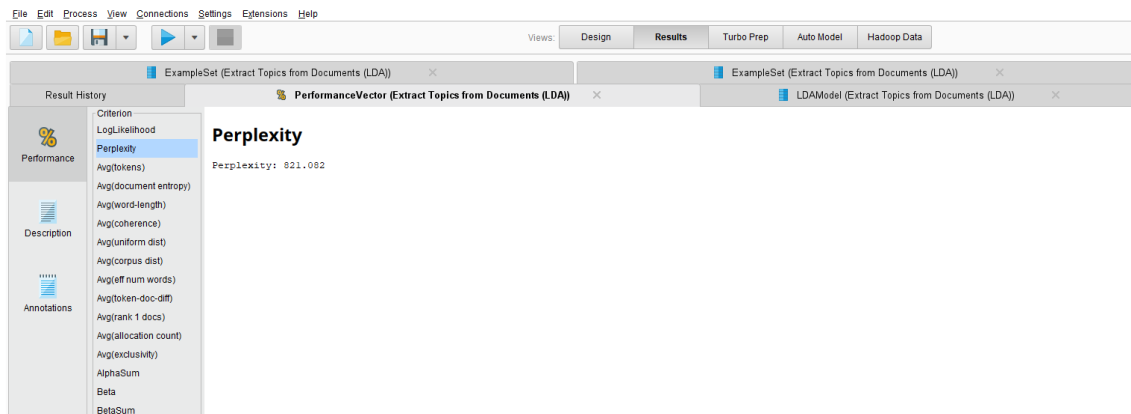


Εικόνα 4.55: Αλλαγή directory για εύρεση αρχείων περιλήψεων της ομάδας Controlled Clinical Trial

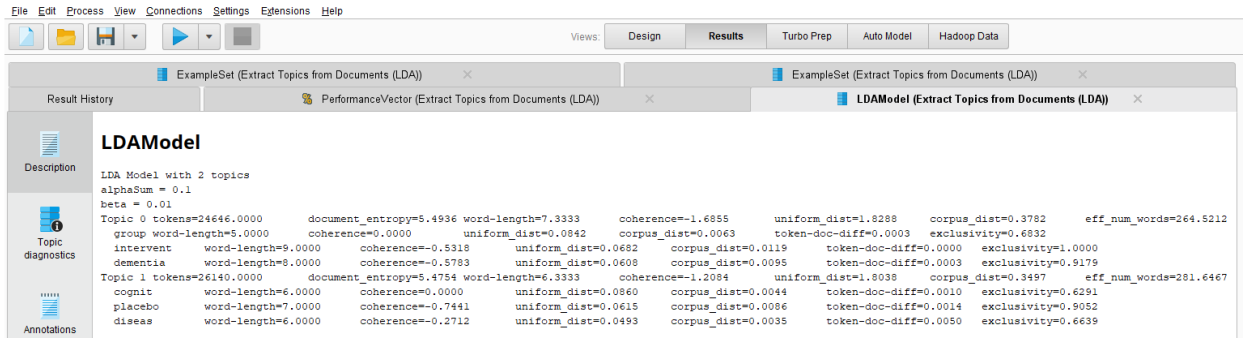
Αποτελέσματα για  $k = 2$  θέματα:



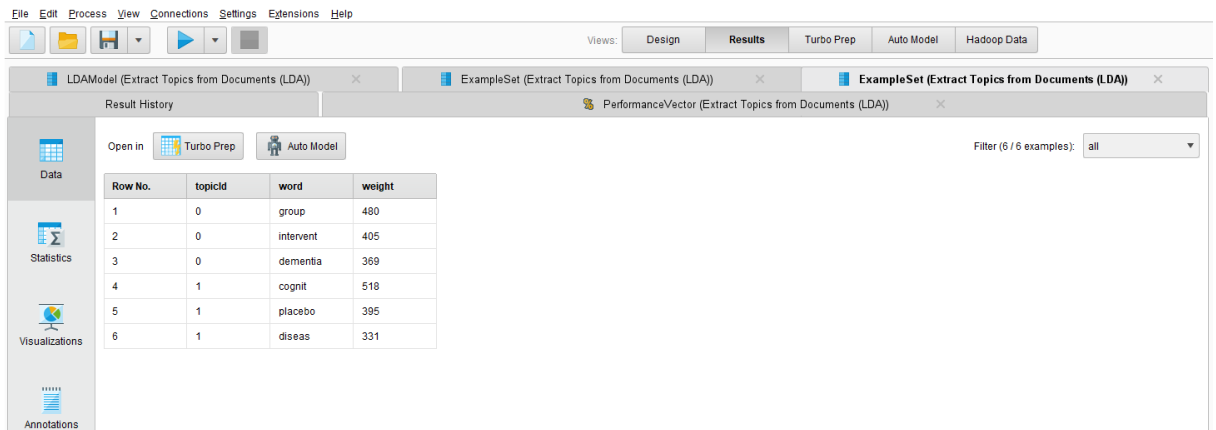
Εικόνα 4.56: Η τιμή Loglikelihood για  $k = 2$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.57: Η τιμή Perplexity για  $k = 2$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.58: Το μοντέλο LDA για  $k = 2$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.59: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για  $k = 2$  θέματα, Controlled Clinical Trial – περιλήψεις.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

LDAModel (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA))

Result History PerformanceVector (Extract Topics from Documents (LDA))

Open in Turbo Prep Auto Model Filter (307 / 307 examples): all

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_1	0.177	0.823	BACKGROU...
2	1	Topic_0	0.895	0.105	Considering L...
3	2	Topic_0	0.745	0.255	OBJECTIVE: ...
4	3	Topic_1	0.298	0.702	BACKGROU...
5	4	Topic_1	0.307	0.693	Early diagnos...
6	5	Topic_1	0.310	0.690	BACKGROU...
7	6	Topic_0	0.604	0.396	BACKGROU...
8	7	Topic_0	0.759	0.241	BACKGROU...
9	8	Topic_1	0.458	0.542	BACKGROU...
10	9	Topic_0	0.970	0.030	BACKGROU...
11	10	Topic_1	0.460	0.540	INTRODUCTL...
12	11	Topic_1	0.088	0.912	
13	12	Topic_1	0.054	0.946	INTRODUCTL...
14	13	Topic_1	0.063	0.937	Background ...
15	14	Topic_0	0.697	0.303	BACKGROU...
16	15	Topic_1	0.131	0.869	Whether HDL...
17	16	Topic_1	0.338	0.662	BACKGROU...
18	17	Topic_1	0.372	0.628	BACKGROU...
19	18	Topic_1	0.307	0.693	BACKGROU...
20	19	Topic_0	0.903	0.097	People with A...
21	20	Topic_0	0.988	0.012	BACKGROU...
22	21	Topic_1	0.104	0.896	BACKGROU...
23	22	Topic_0	0.542	0.458	BACKGROU...
24	23	Topic_1	0.257	0.743	IMPORTANC...
25	24	Topic_1	0.031	0.969	BACKGROU...
26	25	Topic_1	0.155	0.845	HYPOTHESI...
27	26	Topic_0	0.628	0.372	BACKGROU...
28	27	Topic_0	0.959	0.041	In reports of...
29	28	Topic_0	0.980	0.020	OBJECTIVE: ...
30	29	Topic_1	0.121	0.879	AIMS: Cholin...
31	30	Topic_0	0.993	0.007	OBJECTIVE: ...
32	31	Topic_1	0.266	0.734	Alzheimer's d...
33	32	Topic_1	0.410	0.590	BACKGROU...
34	33	Topic_0	0.763	0.237	BACKGROU...
35	34	Topic_0	0.705	0.295	OBJECTIVE: ...
36	35	Topic_1	0.000	1.000	BACKGROU...
37	36	Topic_0	0.904	0.096	The stigma a...
38	37	Topic_1	0.068	0.932	BACKGROU...
39	38	Topic_1	0.328	0.672	The LUCIND...
40	39	Topic_0	0.507	0.493	INTRODUCTL...
41	40	Topic_0	1.000	0.000	Adults with Al...
42	41	Topic_1	0.183	0.817	BACKGROU...
43	42	Topic_1	0.393	0.607	BACKGROU...
44	43	Topic_1	0.129	0.871	BACKGROU...
45	44	Topic_1	0.337	0.663	BACKGROU...
46	45	Topic_1	0.338	0.662	Alzheimer's d...
47	46	Topic_1	0.404	0.596	BACKGROU...
48	47	Topic_0	0.830	0.170	INTRODUCTL...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

49	48	Topic_1	0.412	0.588	Alzheimer's d...
50	49	Topic_0	0.908	0.092	BACKGROU...
51	50	Topic_1	0.118	0.882	Episodic me...
52	51	Topic_0	0.937	0.063	Technologies...
53	52	Topic_0	0.858	0.142	BACKGROU...
54	53	Topic_0	0.691	0.309	BACKGROU...
55	54	Topic_0	0.924	0.076	
56	55	Topic_0	0.905	0.095	BACKGROU...
57	56	Topic_0	0.884	0.116	BACKGROU...
58	57	Topic_1	0.139	0.861	BACKGROU...
59	58	Topic_0	0.918	0.082	BACKGROU...
60	59	Topic_0	0.621	0.379	The prevalen...
61	60	Topic_1	0.140	0.860	PURPOSE: T...
62	61	Topic_1	0.387	0.613	APOE encod...
63	62	Topic_0	0.755	0.245	PURPOSE: T...
64	63	Topic_1	0.250	0.750	BACKGROU...
65	64	Topic_1	0.160	0.840	Accurate mea...
66	65	Topic_1	0.076	0.924	Tau plays a k...
67	66	Topic_1	0.000	1.000	BACKGROU...
68	67	Topic_0	1.000	0.000	OBJECTIVE: ...
69	68	Topic_1	0.411	0.589	Mild cognitive...
70	69	Topic_1	0.414	0.586	BACKGROU...
71	70	Topic_1	0.295	0.705	Microglia play...
72	71	Topic_0	0.679	0.321	Neuroinflam...

73	72	Topic_0	0.566	0.434	BACKGROU...
74	73	Topic_1	0.101	0.899	BACKGROU...
75	74	Topic_0	0.827	0.173	INTRODUCTL...
76	75	Topic_0	0.590	0.410	BACKGROU...
77	76	Topic_0	1.000	0.000	BACKGROU...
78	77	Topic_1	0.322	0.678	INTRODUCTL...
79	78	Topic_0	0.689	0.311	AIM: Previous...
80	79	Topic_0	0.952	0.048	BACKGROU...
81	80	Topic_0	0.587	0.413	PURPOSE: M...
82	81	Topic_0	0.581	0.419	BACKGROU...
83	82	Topic_0	0.971	0.029	Efficacy of as...
84	83	Topic_0	0.722	0.278	OBJECTIVE: ...
85	84	Topic_1	0.064	0.916	BACKGROU...
86	85	Topic_1	0.130	0.870	BACKGROU...
87	86	Topic_1	0.478	0.522	The aim of th...
88	87	Topic_0	0.573	0.427	Spirulina max...
89	88	Topic_1	0.470	0.530	INTRODUCTL...
90	89	Topic_0	0.682	0.318	BACKGROU...
91	90	Topic_1	0.186	0.814	BACKGROU...
92	91	Topic_0	0.881	0.119	
93	92	Topic_1	0.185	0.815	BACKGROU...
94	93	Topic_1	0.118	0.882	BACKGROU...
95	94	Topic_0	0.794	0.206	BACKGROU...
96	95	Topic_1	0.279	0.721	Probiotic sup...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

97	96	Topic_1	0.469	0.531	The progress...
98	97	Topic_1	0.407	0.593	BACKGROU...
99	98	Topic_0	0.623	0.377	IMPORTANC...
100	99	Topic_0	0.690	0.310	BACKGROU...
101	100	Topic_1	0.308	0.692	OBJECTIVES...
102	101	Topic_1	0.169	0.831	BACKGROU...
103	102	Topic_1	0.313	0.687	A poor Fines...
104	103	Topic_1	0.459	0.541	IMPORTANC...
105	104	Topic_0	0.710	0.290	INTRODUCT...
106	105	Topic_1	0.264	0.736	BACKGROU...
107	106	Topic_0	0.742	0.258	BACKGROU...
108	107	Topic_1	0.342	0.658	Learning and...
109	108	Topic_0	0.998	0.002	BACKGROU...
110	109	Topic_1	0.275	0.725	
111	110	Topic_1	0.220	0.780	BACKGROU...
112	111	Topic_1	0.184	0.816	BACKGROU...
113	112	Topic_1	0.140	0.860	OBJECTIVE: ...
114	113	Topic_1	0.283	0.717	AIMS: HTL00...
115	114	Topic_0	0.524	0.476	INTRODUCT...
116	115	Topic_1	0.244	0.756	AIMS: HTL00...
117	116	Topic_1	0.078	0.922	Subjective m...
118	117	Topic_1	0.496	0.504	BACKGROU...
119	118	Topic_0	0.898	0.102	(1) Backgrou...
120	119	Topic_0	0.780	0.220	BACKGROU...

121	120	Topic_1	0.475	0.525	OBJECTIVE: ...
122	121	Topic_0	0.630	0.370	BACKGROU...
123	122	Topic_1	0.205	0.795	BACKGROU...
124	123	Topic_0	0.949	0.051	BACKGROU...
125	124	Topic_0	0.737	0.263	Nighttime agl...
126	125	Topic_1	0.049	0.951	Apolipoprotei...
127	126	Topic_1	0.322	0.678	Research ha...
128	127	Topic_0	0.705	0.295	BACKGROU...
129	128	Topic_0	0.618	0.382	This study util...
130	129	Topic_0	0.714	0.286	RATIONALE: ...
131	130	Topic_1	0.202	0.798	BACKGROU...
132	131	Topic_0	0.999	0.001	The "Great Vil...
133	132	Topic_1	0.000	1.000	BACKGROU...
134	133	Topic_0	0.867	0.133	OBJECTIVES...
135	134	Topic_1	0.223	0.777	
136	135	Topic_0	0.558	0.442	INTRODUCT...
137	136	Topic_1	0.303	0.697	BACKGROU...
138	137	Topic_0	0.624	0.376	BACKGROU...
139	138	Topic_0	0.967	0.033	AIMS: We con...
140	139	Topic_0	0.755	0.245	
141	140	Topic_0	1.000	0.000	BACKGROU...
142	141	Topic_0	0.698	0.302	BACKGROU...
143	142	Topic_0	1.000	0.000	OBJECTIVES...
144	143	Topic_1	0.059	0.941	



«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

145	144	Topic_0	0.989	0.011	BACKGROU...
146	145	Topic_1	0.163	0.837	The purpose ...
147	146	Topic_1	0.134	0.866	BACKGROU...
148	147	Topic_1	0.068	0.932	PURPOSE: S...
149	148	Topic_0	0.598	0.402	
150	149	Topic_1	0.320	0.680	BACKGROU...
151	150	Topic_1	0.282	0.718	
152	151	Topic_1	0.217	0.783	BACKGROU...
153	152	Topic_0	0.664	0.336	INTRODUCTI...
154	153	Topic_0	0.601	0.399	BACKGROU...
155	154	Topic_1	0.097	0.903	INTRODUCTI...
156	155	Topic_1	0.041	0.959	Hippocampal...
157	156	Topic_1	0.145	0.855	AIMS: TAK-07...
158	157	Topic_1	0.478	0.522	BACKGROU...
159	158	Topic_0	0.678	0.322	BACKGROU...
160	159	Topic_1	0.091	0.909	BACKGROU...
161	160	Topic_0	0.689	0.311	Serum amylo...
162	161	Topic_0	0.540	0.460	Long-term us...
163	162	Topic_0	0.932	0.068	OBJECTIVES...
164	163	Topic_1	0.433	0.567	IMPORTANC...
165	164	Topic_0	0.852	0.148	OBJECTIVES...
166	165	Topic_1	0.472	0.528	IMPORTANC...
167	166	Topic_0	0.930	0.070	BACKGROU...
168	167	Topic_1	0.243	0.757	BACKGROU...

169	168	Topic_0	0.511	0.489	IMPORTANC...
170	169	Topic_0	0.931	0.069	BACKGROU...
171	170	Topic_0	0.548	0.452	INTRODUCTI...
172	171	Topic_1	0.164	0.836	
173	172	Topic_1	0.361	0.639	BACKGROU...
174	173	Topic_1	0.448	0.552	Desalted Sali...
175	174	Topic_1	0.169	0.831	Background ...
176	175	Topic_0	0.632	0.368	BACKGROU...
177	176	Topic_0	0.520	0.480	OBJECTIVES...
178	177	Topic_0	0.825	0.175	
179	178	Topic_0	0.569	0.431	BACKGROU...
180	179	Topic_0	1.000	0.000	INTRODUCTI...
181	180	Topic_1	0.195	0.805	IMPORTANC...
182	181	Topic_0	0.898	0.102	BACKGROU...
183	182	Topic_0	1.000	0.000	BACKGROU...
184	183	Topic_0	0.575	0.425	BACKGROU...
185	184	Topic_1	0.204	0.796	OBJECTIVE: ...
186	185	Topic_0	1.000	0.000	OBJECTIVES...
187	186	Topic_1	0.248	0.752	The effects of...
188	187	Topic_1	0.164	0.836	IMPORTANC...
189	188	Topic_1	0.146	0.854	HIV-associat...
190	189	Topic_1	0.282	0.718	BACKGROU...
191	190	Topic_0	0.916	0.084	BACKGROU...
192	191	Topic_1	0.157	0.843	IMPORTANC...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

193	192	Topic_1	0.153	0.847	IMPORTANC...
194	193	Topic_0	0.870	0.130	OBJECTIVES...
195	194	Topic_1	0.238	0.762	The CAIDE (...)
196	195	Topic_1	0.414	0.586	OBJECTIVES...
197	196	Topic_0	0.798	0.202	BACKGROU...
198	197	Topic_0	0.758	0.242	OBJECTIVE: ...
199	198	Topic_1	0.068	0.932	The locus co...
200	199	Topic_0	0.996	0.004	OBJECTIVE: ...
201	200	Topic_0	0.506	0.494	INTRODUCTI...
202	201	Topic_0	0.844	0.156	BACKGROU...
203	202	Topic_0	0.604	0.396	We previousl...
204	203	Topic_0	0.930	0.070	BACKGROU...
205	204	Topic_0	0.875	0.125	PURPOSE: In...
206	205	Topic_0	0.553	0.447	AIM: To Invest...
207	206	Topic_0	0.537	0.463	
208	207	Topic_0	0.923	0.077	
209	208	Topic_1	0.073	0.927	IMPORTANC...
210	209	Topic_1	0.460	0.540	BACKGROU...
211	210	Topic_1	0.482	0.518	IMPORTANC...
212	211	Topic_0	0.685	0.315	IMPORTANC...
213	212	Topic_0	1.000	0.000	BACKGROU...
214	213	Topic_1	0.045	0.955	OBJECTIVE: ...
215	214	Topic_1	0.080	0.920	OBJECTIVE: ...
216	215	Topic_1	0.280	0.720	BACKGROU...
217	216	Topic_0	0.809	0.191	Transcranial ...
218	217	Topic_1	0.496	0.504	IMPORTANC...
219	218	Topic_1	0.339	0.661	
220	219	Topic_0	0.621	0.379	IMPORTANC...
221	220	Topic_0	0.917	0.083	BACKGROU...
222	221	Topic_1	0.327	0.673	BACKGROU...
223	222	Topic_0	0.659	0.341	BACKGROU...
224	223	Topic_0	0.999	0.001	IMPORTANC...
225	224	Topic_0	1.000	0.000	OBJECTIVES...
226	225	Topic_0	0.750	0.250	BACKGROU...
227	226	Topic_0	0.905	0.095	BACKGROU...
228	227	Topic_1	0.381	0.619	OBJECTIVE: ...
229	228	Topic_0	0.723	0.277	BACKGROU...
230	229	Topic_0	0.815	0.185	BACKGROU...
231	230	Topic_1	0.174	0.826	BACKGROU...
232	231	Topic_1	0.346	0.654	IMPORTANC...
233	232	Topic_1	0.012	0.988	OBJECTIVE: ...
234	233	Topic_0	0.724	0.276	IMPORTANC...
235	234	Topic_1	0.363	0.637	IMPORTANC...
236	235	Topic_0	0.636	0.364	IMPORTANC...
237	236	Topic_0	0.679	0.321	The use of ec...
238	237	Topic_1	0.262	0.738	BACKGROU...
239	238	Topic_1	0.398	0.602	White matter ...
240	239	Topic_1	0.180	0.820	BACKGROU...

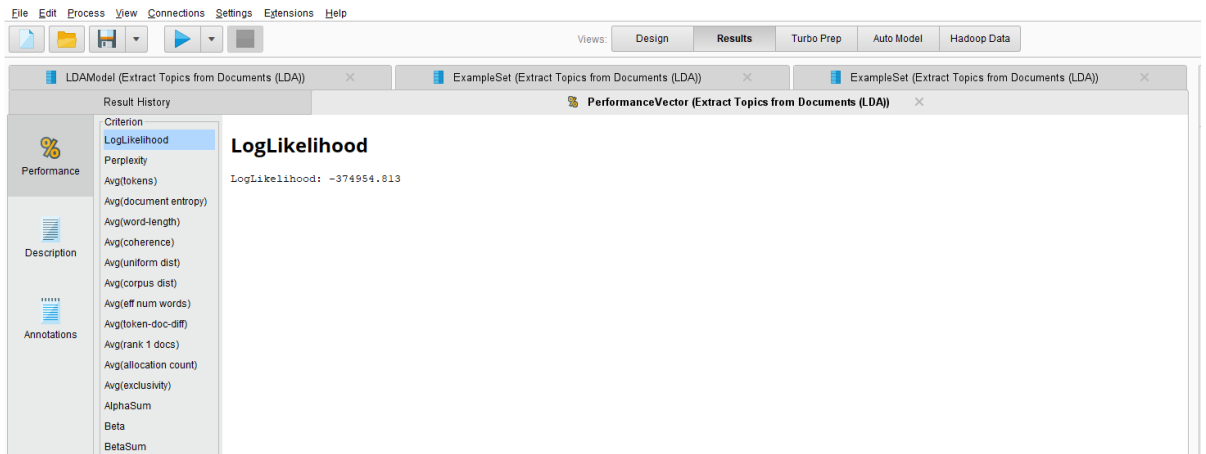
«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

241	240	Topic_0	0.552	0.448	INTRODUCTI...
242	241	Topic_1	0.296	0.704	BACKGROU...
243	242	Topic_1	0.226	0.774	BACKGROU...
244	243	Topic_1	0.360	0.640	Dominantly in...
245	244	Topic_0	0.523	0.477	Dietary suppl...
246	245	Topic_1	0.097	0.903	BACKGROU...
247	246	Topic_1	0.092	0.908	Dysregulatio...
248	247	Topic_0	0.838	0.162	BACKGROU...
249	248	Topic_0	0.581	0.419	BACKGROU...
250	249	Topic_1	0.349	0.651	The treatmen...
251	250	Topic_1	0.409	0.591	Dementia aff...
252	251	Topic_0	0.581	0.419	
253	252	Topic_0	0.726	0.274	Close to 6 mi...
254	253	Topic_1	0.030	0.970	OBJECTIVE: ...
255	254	Topic_1	0.149	0.851	Current thera...
256	255	Topic_1	0.434	0.566	BACKGROU...
257	256	Topic_0	0.692	0.308	We conducte...
258	257	Topic_1	0.455	0.545	BACKGROU...
259	258	Topic_1	0.266	0.734	OBJECTIVE: ...
260	259	Topic_1	0.002	0.998	Tau hyperpho...
261	260	Topic_0	0.684	0.316	OBJECTIVE: ...
262	261	Topic_1	0.010	0.990	In the phase ...
263	262	Topic_1	0.263	0.737	INTRODUCTI...
264	263	Topic_1	0.429	0.571	PURPOSE: V...
265	264	Topic_1	0.107	0.893	OBJECTIVE: ...
266	265	Topic_0	0.507	0.493	BACKGROU...
267	266	Topic_1	0.259	0.741	OBJECTIVE: ...
268	267	Topic_1	0.137	0.863	IMPORTANC...
269	268	Topic_0	0.988	0.012	Objectives: Pr...
270	269	Topic_1	0.137	0.863	Intranasal ins...
271	270	Topic_1	0.304	0.696	BACKGROU...
272	271	Topic_1	0.037	0.963	BACKGROU...
273	272	Topic_1	0.170	0.830	Non-invasive ...
274	273	Topic_1	0.192	0.808	IMPORTANC...
275	274	Topic_1	0.413	0.587	BACKGROU...
276	275	Topic_1	0.273	0.727	Repetitive tra...
277	276	Topic_1	0.068	0.932	OBJECTIVE: ...
278	277	Topic_0	0.813	0.187	OBJECTIVE: ...
279	278	Topic_0	0.782	0.218	BACKGROU...
280	279	Topic_1	0.047	0.953	RIPK1 is a m...
281	280	Topic_1	0.027	0.973	INTRODUCTI...
282	281	Topic_0	0.899	0.101	OBJECTIVES...
283	282	Topic_1	0.081	0.919	Mild cognitive...
284	283	Topic_0	0.692	0.308	No prior studi...
285	284	Topic_0	0.851	0.149	IMPORTANC...
286	285	Topic_0	0.553	0.447	BACKGROU...
287	286	Topic_1	0.339	0.661	Increasing ev...
288	287	Topic_0	0.903	0.097	We examined...
289	288	Topic_1	0.356	0.644	BACKGROU...
290	289	Topic_0	0.753	0.247	IMPORTANC...
291	290	Topic_1	0.295	0.705	Aerobic exerc...
292	291	Topic_1	0.438	0.562	IMPORTANC...
293	292	Topic_0	0.960	0.040	IMPORTANC...
294	293	Topic_1	0.071	0.929	
295	294	Topic_1	0.261	0.739	Probiotics co...
296	295	Topic_0	1.000	0.000	BACKGROU...
297	296	Topic_1	0.471	0.529	BACKGROU...
298	297	Topic_1	0.401	0.599	BACKGROU...
299	298	Topic_1	0.419	0.581	There is eme...
300	299	Topic_1	0.395	0.605	IMPORTANC...
301	300	Topic_0	0.740	0.260	BACKGROU...
302	301	Topic_1	0.246	0.754	Tau accumul...
303	302	Topic_0	0.697	0.303	OBJECTIVE: ...
304	303	Topic_0	0.794	0.206	BACKGROU...
305	304	Topic_0	0.689	0.311	BACKGROU...
306	305	Topic_0	0.682	0.318	Age-related d...
307	306	Topic_1	0.205	0.795	

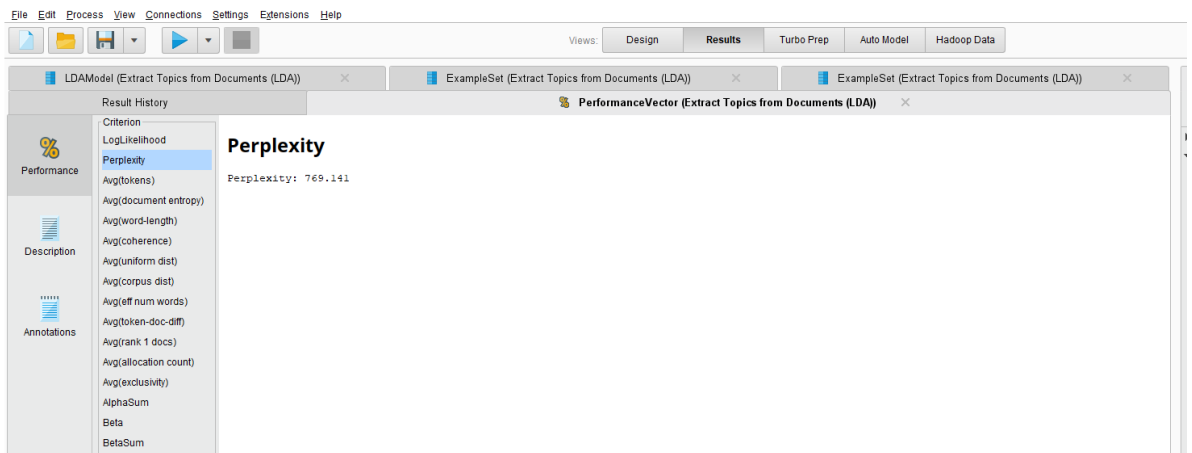
ExampleSet (307 examples, 4 special attributes, 1 regular attribute)

Εικόνα 4.60: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε  $k = 2$  θέματα

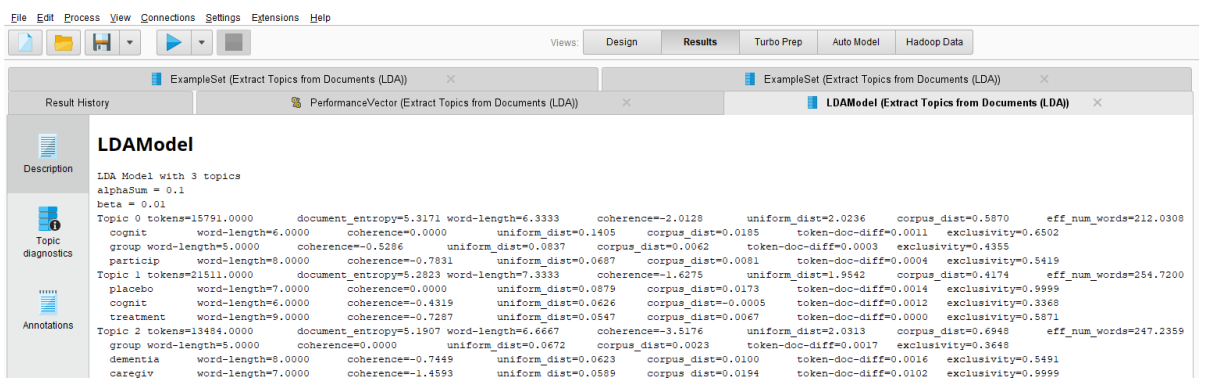
Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :



Εικόνα 4.61: Η τιμή Loglikelihood για  $k = 3$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.62: Η τιμή Perplexity για  $k = 3$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.63: Το μοντέλο LDA για  $k = 3$  θέματα, Controlled Clinical Trial – περιλήψεις.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Row No.	topicId	word	weight
1	0	cognit	468
2	0	group	306
3	0	particip	261
4	1	placebo	434
5	1	cognit	330
6	1	treatment	296
7	2	group	219
8	2	dementia	206
9	2	caregiv	197

Εικόνα 4.64: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για k = 3 θέματα, Controlled Clinical Trial – περιλήψεις.

Row No.	documentId	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_1	0.219	0.781	0.001	BACKGROU...
2	1	Topic_2	0.459	0.000	0.540	Considering L...
3	2	Topic_2	0.384	0.156	0.460	OBJECTIVE: ...
4	3	Topic_1	0.121	0.767	0.112	BACKGROU...
5	4	Topic_0	0.516	0.370	0.114	Early diagnos...
6	5	Topic_1	0.056	0.943	0.001	BACKGROU...
7	6	Topic_2	0.130	0.222	0.647	BACKGROU...
8	7	Topic_0	0.383	0.255	0.363	BACKGROU...
9	8	Topic_1	0.006	0.566	0.428	BACKGROU...
10	9	Topic_2	0.458	0.001	0.540	BACKGROU...
11	10	Topic_1	0.351	0.647	0.002	INTRODUCT...
12	11	Topic_1	0.019	0.980	0.000	
13	12	Topic_1	0.334	0.665	0.000	INTRODUCT...
14	13	Topic_1	0.000	1.000	0.000	Background: ...
15	14	Topic_0	0.558	0.382	0.060	BACKGROU...
16	15	Topic_1	0.420	0.578	0.001	Whether HDL...
17	16	Topic_1	0.274	0.726	0.000	BACKGROU...
18	17	Topic_2	0.000	0.167	0.833	BACKGROU...
19	18	Topic_1	0.041	0.782	0.177	BACKGROU...
20	19	Topic_0	0.533	0.000	0.467	People with A...
21	20	Topic_2	0.283	0.000	0.717	BACKGROU...
22	21	Topic_1	0.000	0.905	0.094	BACKGROU...
23	22	Topic_0	0.659	0.336	0.005	BACKGROU...
24	23	Topic_1	0.283	0.698	0.019	IMPORTANC...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

25	24	Topic_1	0.001	0.963	0.036	BACKGROU...
26	25	Topic_1	0.031	0.888	0.080	HYPOTHESL...
27	26	Topic_1	0.239	0.485	0.276	BACKGROU...
28	27	Topic_2	0.189	0.001	0.810	In reports of r...
29	28	Topic_0	0.596	0.002	0.403	OBJECTIVE: ...
30	29	Topic_1	0.000	0.564	0.435	AIMS: Cholin...
31	30	Topic_2	0.088	0.001	0.912	OBJECTIVE: ...
32	31	Topic_1	0.085	0.487	0.429	Alzheimer's d...
33	32	Topic_0	0.413	0.180	0.407	BACKGROU...
34	33	Topic_0	0.767	0.097	0.136	BACKGROU...
35	34	Topic_1	0.061	0.722	0.216	OBJECTIVE: ...
36	35	Topic_1	0.090	0.910	0.000	BACKGROU...
37	36	Topic_2	0.235	0.002	0.763	The stigma a...
38	37	Topic_1	0.045	0.954	0.000	BACKGROU...
39	38	Topic_1	0.000	0.750	0.250	The LUCIND...
40	39	Topic_1	0.062	0.624	0.315	INTRODUCTI...
41	40	Topic_2	0.323	0.000	0.676	Adults with Al...
42	41	Topic_1	0.084	0.542	0.374	BACKGROU...
43	42	Topic_0	0.757	0.041	0.201	BACKGROU...
44	43	Topic_1	0.250	0.747	0.003	BACKGROU...
45	44	Topic_2	0.349	0.301	0.350	BACKGROU...
46	45	Topic_0	0.526	0.400	0.073	Alzheimer's d...
47	46	Topic_1	0.346	0.478	0.175	BACKGROU...
48	47	Topic_2	0.204	0.157	0.638	INTRODUCTI...
49	48	Topic_0	0.677	0.095	0.228	Alzheimer's d...
50	49	Topic_0	0.947	0.053	0.001	BACKGROU...
51	50	Topic_2	0.040	0.213	0.747	Episodic me...
52	51	Topic_2	0.078	0.000	0.922	Technologies...
53	52	Topic_2	0.219	0.069	0.711	BACKGROU...
54	53	Topic_0	0.605	0.282	0.113	BACKGROU...
55	54	Topic_2	0.257	0.002	0.741	BACKGROU...
56	55	Topic_2	0.007	0.000	0.993	BACKGROU...
57	56	Topic_0	0.713	0.043	0.244	BACKGROU...
58	57	Topic_1	0.367	0.633	0.000	BACKGROU...
59	58	Topic_2	0.243	0.000	0.757	BACKGROU...
60	59	Topic_2	0.240	0.338	0.422	The prevalen...
61	60	Topic_1	0.000	0.914	0.086	PURPOSE: T...
62	61	Topic_1	0.179	0.722	0.099	APOE encod...
63	62	Topic_0	0.520	0.156	0.325	PURPOSE: T...
64	63	Topic_1	0.145	0.855	0.000	BACKGROU...
65	64	Topic_1	0.311	0.544	0.145	Accurate mea...
66	65	Topic_1	0.000	0.975	0.024	Tau plays a k...
67	66	Topic_1	0.171	0.829	0.000	BACKGROU...
68	67	Topic_0	0.535	0.000	0.465	OBJECTIVE: ...
69	68	Topic_0	0.795	0.000	0.205	Mild cognitiv...
70	69	Topic_0	0.737	0.156	0.107	BACKGROU...
71	70	Topic_1	0.404	0.505	0.091	Microglia play...
72	71	Topic_0	0.503	0.497	0.000	Neuroinflam...
73	72	Topic_0	0.489	0.483	0.028	BACKGROU...
74	73	Topic_1	0.054	0.871	0.075	BACKGROU...
75	74	Topic_0	0.688	0.309	0.002	INTRODUCTI...
76	75	Topic_0	0.600	0.000	0.399	BACKGROU...
77	76	Topic_2	0.366	0.000	0.634	BACKGROU...
78	77	Topic_1	0.178	0.822	0.000	INTRODUCTI...
79	78	Topic_1	0.000	0.586	0.414	AIM: Previous...
80	79	Topic_2	0.487	0.000	0.512	BACKGROU...
81	80	Topic_2	0.407	0.000	0.593	PURPOSE: M...
82	81	Topic_0	0.578	0.422	0.000	BACKGROU...
83	82	Topic_2	0.233	0.000	0.767	Efficacy of as...
84	83	Topic_0	0.609	0.386	0.005	OBJECTIVE: ...
85	84	Topic_1	0.427	0.533	0.040	BACKGROU...
86	85	Topic_1	0.213	0.783	0.004	BACKGROU...
87	86	Topic_0	0.838	0.017	0.145	The aim of th...
88	87	Topic_1	0.328	0.419	0.253	Spirulina max...
89	88	Topic_0	0.528	0.467	0.005	INTRODUCTI...
90	89	Topic_1	0.244	0.660	0.096	BACKGROU...
91	90	Topic_1	0.211	0.652	0.138	BACKGROU...
92	91	Topic_2	0.129	0.028	0.843	BACKGROU...
93	92	Topic_1	0.237	0.674	0.089	BACKGROU...
94	93	Topic_1	0.076	0.922	0.002	BACKGROU...
95	94	Topic_2	0.304	0.066	0.629	BACKGROU...
96	95	Topic_1	0.314	0.534	0.152	Probiotic sup...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

97	96	Topic_0	0.455	0.157	0.388	The progress...
98	97	Topic_0	0.522	0.473	0.005	BACKGROU...
99	98	Topic_1	0.199	0.430	0.371	IMPORTANC...
100	99	Topic_0	0.723	0.245	0.032	BACKGROU...
101	100	Topic_1	0.001	0.778	0.221	OBJECTIVES...
102	101	Topic_1	0.133	0.748	0.119	BACKGROU...
103	102	Topic_0	0.827	0.132	0.041	A poor Fibres...
104	103	Topic_1	0.214	0.645	0.141	IMPORTANC...
105	104	Topic_0	0.995	0.004	0.000	INTRODUCTI...
106	105	Topic_1	0.032	0.768	0.200	BACKGROU...
107	106	Topic_0	0.616	0.186	0.198	BACKGROU...
108	107	Topic_2	0.361	0.254	0.385	Learning and...
109	108	Topic_2	0.000	0.000	0.999	BACKGROU...
110	109	Topic_1	0.137	0.715	0.148	
111	110	Topic_1	0.202	0.798	0.000	BACKGROU...
112	111	Topic_1	0.403	0.463	0.134	BACKGROU...
113	112	Topic_1	0.000	0.868	0.131	OBJECTIVE...
114	113	Topic_1	0.084	0.916	0.000	AIMS: HTL00...
115	114	Topic_2	0.150	0.381	0.469	INTRODUCTI...
116	115	Topic_1	0.005	0.929	0.067	AIMS: HTL00...
117	116	Topic_2	0.001	0.339	0.660	Subjective m...
118	117	Topic_2	0.225	0.068	0.708	BACKGROU...
119	118	Topic_2	0.248	0.000	0.752	(1) Backgrou...
120	119	Topic_0	0.704	0.012	0.285	BACKGROU...
121	120	Topic_1	0.000	0.646	0.354	OBJECTIVE: ...
122	121	Topic_1	0.122	0.628	0.249	BACKGROU...
123	122	Topic_1	0.070	0.930	0.000	BACKGROU...
124	123	Topic_2	0.369	0.005	0.626	BACKGROU...
125	124	Topic_2	0.110	0.089	0.802	Nighttime agi...
126	125	Topic_1	0.000	0.974	0.026	Apollipoprotel...
127	126	Topic_1	0.000	0.851	0.148	Research ha...
128	127	Topic_2	0.465	0.000	0.534	BACKGROU...
129	128	Topic_0	0.660	0.339	0.000	This study utili...
130	129	Topic_1	0.004	0.835	0.161	RATIONALE: ...
131	130	Topic_1	0.488	0.492	0.021	BACKGROU...
132	131	Topic_2	0.003	0.000	0.997	The "Great Vill...
133	132	Topic_1	0.004	0.946	0.050	BACKGROU...
134	133	Topic_2	0.301	0.000	0.699	OBJECTIVES...
135	134	Topic_0	0.931	0.046	0.023	
136	135	Topic_0	0.876	0.000	0.124	INTRODUCTI...
137	136	Topic_0	0.656	0.343	0.000	BACKGROU...
138	137	Topic_0	0.565	0.068	0.366	BACKGROU...
139	138	Topic_2	0.396	0.003	0.601	AIMS: We con...
140	139	Topic_1	0.275	0.723	0.003	
141	140	Topic_2	0.030	0.000	0.970	BACKGROU...
142	141	Topic_0	0.450	0.201	0.349	BACKGROU...
143	142	Topic_2	0.145	0.000	0.854	OBJECTIVES...
144	143	Topic_1	0.140	0.823	0.038	
145	144	Topic_2	0.274	0.000	0.725	BACKGROU...
146	145	Topic_0	1.000	0.000	0.000	The purpose ...
147	146	Topic_1	0.055	0.945	0.000	BACKGROU...
148	147	Topic_1	0.000	1.000	0.000	PURPOSE: S...
149	148	Topic_1	0.480	0.493	0.027	
150	149	Topic_1	0.188	0.677	0.135	BACKGROU...
151	150	Topic_0	0.898	0.102	0.000	
152	151	Topic_1	0.048	0.881	0.071	BACKGROU...
153	152	Topic_2	0.415	0.072	0.513	INTRODUCTI...
154	153	Topic_0	0.591	0.027	0.382	BACKGROU...
155	154	Topic_1	0.003	0.874	0.123	INTRODUCTI...
156	155	Topic_1	0.072	0.673	0.255	Hippocampal...
157	156	Topic_1	0.000	0.873	0.127	AIMS: TAK-07...
158	157	Topic_0	0.690	0.231	0.079	BACKGROU...
159	158	Topic_0	0.607	0.154	0.238	BACKGROU...
160	159	Topic_1	0.002	0.885	0.114	BACKGROU...
161	160	Topic_2	0.003	0.355	0.642	Serum amylo...
162	161	Topic_0	0.838	0.161	0.001	Long-term us...
163	162	Topic_2	0.207	0.001	0.791	OBJECTIVES...
164	163	Topic_0	0.634	0.339	0.026	IMPORTANC...
165	164	Topic_2	0.158	0.038	0.804	OBJECTIVES...
166	165	Topic_1	0.459	0.541	0.000	IMPORTANC...
167	166	Topic_2	0.181	0.004	0.815	BACKGROU...
168	167	Topic_1	0.419	0.487	0.093	BACKGROU...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

169	168	Topic_1	0.168	0.832	0.000	IMPORTANC...
170	169	Topic_2	0.233	0.000	0.767	BACKGROU...
171	170	Topic_2	0.185	0.365	0.451	INTRODUCTI...
172	171	Topic_1	0.415	0.582	0.003	
173	172	Topic_0	0.624	0.141	0.235	BACKGROU...
174	173	Topic_1	0.311	0.414	0.274	Desalted Sal...
175	174	Topic_1	0.038	0.880	0.082	Background: ...
176	175	Topic_0	0.725	0.026	0.249	BACKGROU...
177	176	Topic_1	0.063	0.838	0.099	OBJECTIVES...
178	177	Topic_0	0.783	0.064	0.153	
179	178	Topic_1	0.092	0.671	0.237	BACKGROU...
180	179	Topic_2	0.185	0.002	0.813	INTRODUCTI...
181	180	Topic_1	0.384	0.516	0.100	IMPORTANC...
182	181	Topic_2	0.187	0.000	0.813	BACKGROU...
183	182	Topic_2	0.290	0.000	0.710	BACKGROU...
184	183	Topic_2	0.064	0.400	0.536	BACKGROU...
185	184	Topic_1	0.000	0.595	0.405	OBJECTIVE: ...
186	185	Topic_2	0.280	0.000	0.719	OBJECTIVES...
187	186	Topic_1	0.353	0.559	0.088	The effects of...
188	187	Topic_1	0.099	0.901	0.000	IMPORTANC...
189	188	Topic_1	0.103	0.653	0.244	HIV-assodat...
190	189	Topic_1	0.001	0.536	0.463	BACKGROU...
191	190	Topic_2	0.273	0.000	0.727	BACKGROU...
192	191	Topic_1	0.325	0.647	0.028	IMPORTANC...
193	192	Topic_1	0.110	0.826	0.064	IMPORTANC...
194	193	Topic_2	0.329	0.038	0.634	OBJECTIVES...
195	194	Topic_0	0.673	0.327	0.000	The CAIDE (...)
196	195	Topic_2	0.368	0.175	0.457	OBJECTIVES...
197	196	Topic_2	0.142	0.127	0.730	BACKGROU...
198	197	Topic_0	0.925	0.000	0.075	OBJECTIVE: ...
199	198	Topic_1	0.044	0.891	0.065	The locus co...
200	199	Topic_2	0.205	0.000	0.795	OBJECTIVE: ...
201	200	Topic_1	0.000	0.945	0.055	INTRODUCTI...
202	201	Topic_2	0.249	0.243	0.508	BACKGROU...
203	202	Topic_1	0.151	0.665	0.184	We previous...
204	203	Topic_2	0.432	0.000	0.568	BACKGROU...
205	204	Topic_0	0.748	0.000	0.252	PURPOSE: In...
206	205	Topic_2	0.467	0.000	0.513	AIM: To invest...
207	206	Topic_1	0.137	0.576	0.287	
208	207	Topic_2	0.248	0.000	0.752	
209	208	Topic_1	0.074	0.923	0.003	IMPORTANC...
210	209	Topic_1	0.400	0.455	0.144	BACKGROU...
211	210	Topic_1	0.409	0.450	0.141	IMPORTANC...
212	211	Topic_1	0.038	0.552	0.410	IMPORTANC...
213	212	Topic_2	0.283	0.000	0.717	BACKGROU...
214	213	Topic_1	0.070	0.930	0.000	OBJECTIVE: ...
215	214	Topic_1	0.000	0.935	0.064	OBJECTIVE: ...
216	215	Topic_1	0.113	0.591	0.296	BACKGROU...



«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

217	216	Topic_2	0.269	0.001	0.730	Transcranial ...
218	217	Topic_1	0.062	0.759	0.178	IMPORTANC...
219	218	Topic_1	0.357	0.542	0.100	
220	219	Topic_1	0.252	0.511	0.237	IMPORTANC...
221	220	Topic_0	0.899	0.004	0.097	BACKGROU...
222	221	Topic_1	0.392	0.598	0.010	BACKGROU...
223	222	Topic_0	0.437	0.183	0.381	BACKGROU...
224	223	Topic_2	0.484	0.000	0.516	IMPORTANC...
225	224	Topic_2	0.120	0.001	0.879	OBJECTIVES...
226	225	Topic_2	0.396	0.003	0.601	BACKGROU...
227	226	Topic_0	0.541	0.000	0.459	BACKGROU...
228	227	Topic_1	0.125	0.640	0.235	OBJECTIVE: ...
229	228	Topic_0	0.655	0.045	0.300	BACKGROU...
230	229	Topic_0	0.550	0.000	0.450	BACKGROU...
231	230	Topic_0	0.544	0.455	0.001	BACKGROU...
232	231	Topic_0	0.481	0.342	0.177	IMPORTANC...
233	232	Topic_2	0.028	0.419	0.553	OBJECTIVE: ...
234	233	Topic_0	0.589	0.411	0.000	IMPORTANC...
235	234	Topic_1	0.094	0.683	0.223	IMPORTANC...
236	235	Topic_1	0.301	0.651	0.048	IMPORTANC...
237	236	Topic_2	0.381	0.099	0.520	The use of ec...
238	237	Topic_1	0.159	0.841	0.000	BACKGROU...
239	238	Topic_0	0.660	0.340	0.000	White matter ...
240	239	Topic_1	0.246	0.696	0.058	BACKGROU...

241	240	Topic_1	0.285	0.712	0.003	INTRODUCT...
242	241	Topic_1	0.278	0.527	0.195	BACKGROU...
243	242	Topic_1	0.049	0.951	0.000	BACKGROU...
244	243	Topic_1	0.374	0.624	0.002	Dominantly in...
245	244	Topic_1	0.007	0.563	0.431	Dietary suppl...
246	245	Topic_1	0.238	0.527	0.235	BACKGROU...
247	246	Topic_1	0.085	0.882	0.032	Dysregulatio...
248	247	Topic_2	0.086	0.252	0.662	BACKGROU...
249	248	Topic_0	0.756	0.136	0.108	BACKGROU...
250	249	Topic_2	0.263	0.200	0.537	The treatmen...
251	250	Topic_0	0.564	0.192	0.244	Dementia aff...
252	251	Topic_1	0.228	0.711	0.061	
253	252	Topic_0	0.732	0.000	0.268	Close to 6 mi...
254	253	Topic_2	0.000	0.365	0.634	OBJECTIVE: ...
255	254	Topic_1	0.000	0.788	0.212	Current thera...
256	255	Topic_1	0.207	0.611	0.182	BACKGROU...
257	256	Topic_1	0.004	0.732	0.264	We conducte...
258	257	Topic_2	0.000	0.443	0.557	BACKGROU...
259	258	Topic_2	0.215	0.368	0.417	OBJECTIVE: ...
260	259	Topic_1	0.002	0.997	0.000	Tau hyperpho...
261	260	Topic_0	0.916	0.082	0.002	OBJECTIVE: ...
262	261	Topic_1	0.000	0.955	0.045	In the phase ...
263	262	Topic_1	0.493	0.506	0.000	INTRODUCT...
264	263	Topic_0	0.680	0.135	0.186	PURPOSE: V...

265	264	Topic_1	0.108	0.867	0.025	OBJECTIVE: ...
266	265	Topic_0	0.550	0.179	0.271	BACKGROU...
267	266	Topic_0	0.774	0.226	0.000	OBJECTIVE: ...
268	267	Topic_1	0.080	0.920	0.000	IMPORTANC...
269	268	Topic_2	0.097	0.000	0.903	Objectives: Pr...
270	269	Topic_1	0.074	0.887	0.040	Intranasal ins...
271	270	Topic_1	0.328	0.671	0.000	BACKGROU...
272	271	Topic_1	0.000	0.964	0.036	BACKGROU...
273	272	Topic_1	0.000	0.706	0.293	Non-invasive ...
274	273	Topic_1	0.088	0.911	0.000	IMPORTANC...
275	274	Topic_0	0.506	0.468	0.026	BACKGROU...
276	275	Topic_1	0.053	0.530	0.418	Repetitive tra...
277	276	Topic_1	0.125	0.875	0.000	OBJECTIVE: ...
278	277	Topic_2	0.414	0.000	0.586	OBJECTIVE: ...
279	278	Topic_1	0.150	0.442	0.408	BACKGROU...
280	279	Topic_1	0.061	0.939	0.000	RIPK1 is a m...
281	280	Topic_2	0.064	0.402	0.534	INTRODUCT...
282	281	Topic_0	0.613	0.002	0.385	OBJECTIVES...
283	282	Topic_1	0.057	0.825	0.118	Mild cognitiv...
284	283	Topic_0	0.599	0.349	0.052	No prior stud...
285	284	Topic_0	0.704	0.148	0.147	IMPORTANC...
286	285	Topic_0	0.660	0.194	0.146	BACKGROU...
287	286	Topic_1	0.288	0.681	0.032	Increasing ev...
288	287	Topic_2	0.115	0.001	0.884	We examined...

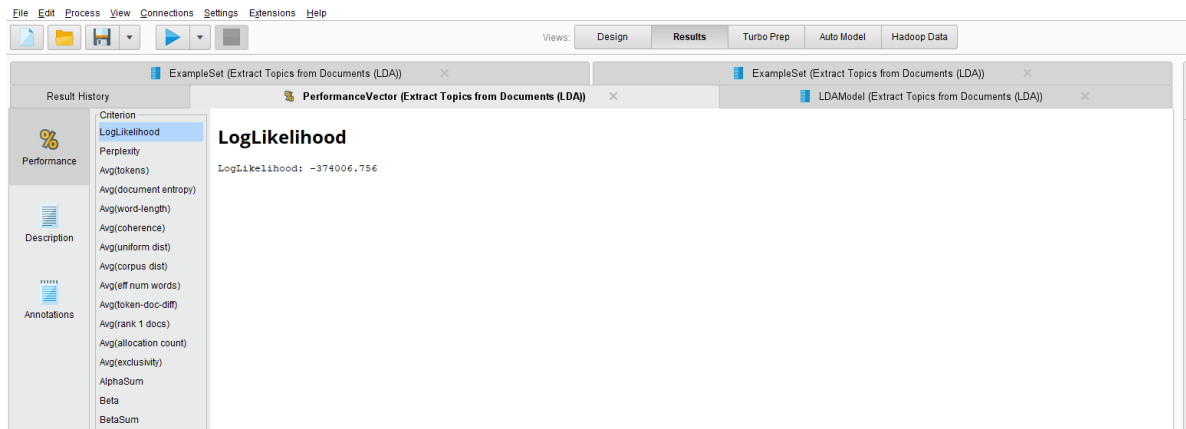
«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

289	288	Topic_1	0.319	0.668	0.013	BACKGROU...
290	289	Topic_2	0.000	0.430	0.570	IMPORTANC...
291	290	Topic_0	0.576	0.417	0.008	Aerobic exerc...
292	291	Topic_1	0.167	0.798	0.035	IMPORTANC...
293	292	Topic_2	0.244	0.000	0.756	IMPORTANC...
294	293	Topic_1	0.101	0.733	0.166	
295	294	Topic_0	0.536	0.464	0.000	Probiotics co...
296	295	Topic_2	0.406	0.010	0.584	BACKGROU...
297	296	Topic_0	0.877	0.018	0.104	BACKGROU...
298	297	Topic_1	0.284	0.460	0.256	BACKGROU...
299	298	Topic_0	0.712	0.247	0.042	There is eme...
300	299	Topic_1	0.470	0.530	0.000	IMPORTANC...
301	300	Topic_0	0.699	0.217	0.083	BACKGROU...
302	301	Topic_1	0.135	0.865	0.000	Tau accumul...
303	302	Topic_2	0.001	0.363	0.636	OBJECTIVE:...
304	303	Topic_0	0.809	0.041	0.150	BACKGROU...
305	304	Topic_0	0.745	0.168	0.086	BACKGROU...
306	305	Topic_0	0.728	0.058	0.214	Age-related d...
307	306	Topic_0	0.658	0.330	0.012	

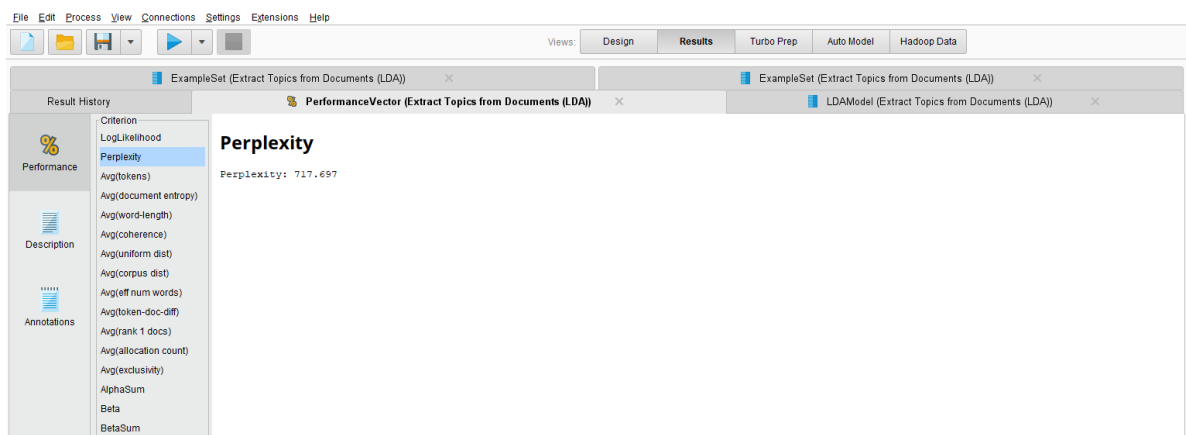
ExampleSet (307 examples, 5 special attributes, 1 regular attribute)

Εικόνα 4.65: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε  $k = 3$  θέματα.

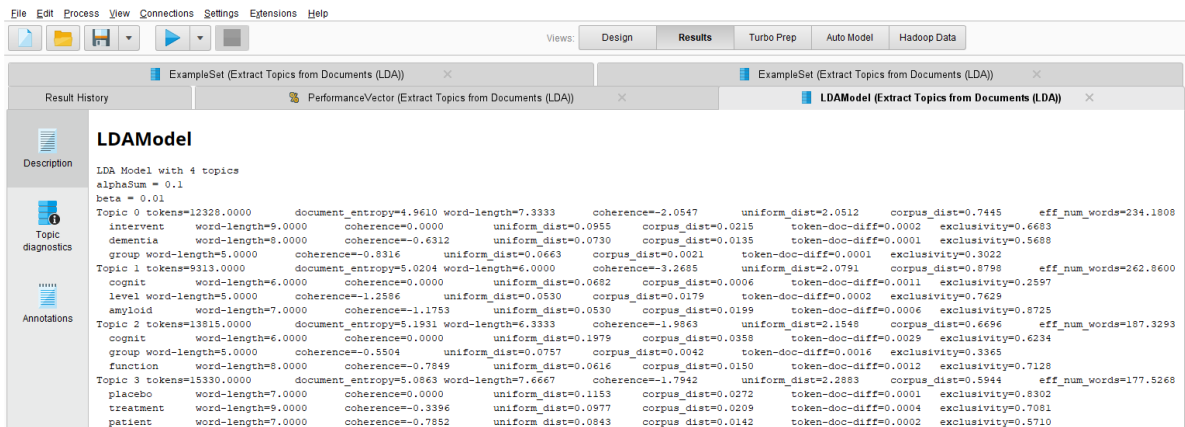
Αποτελέσματα για αριθμό θεμάτων  $k = 4$ :



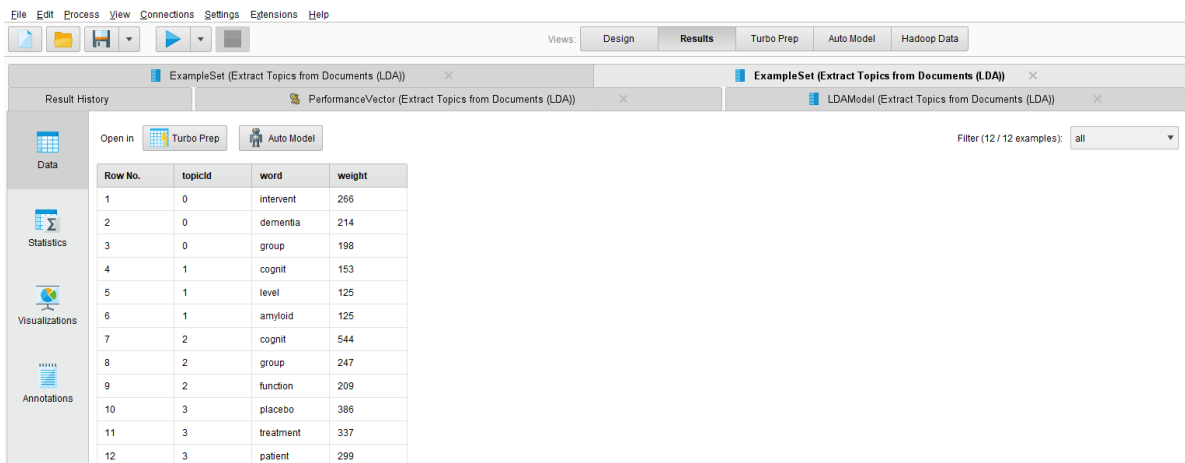
Εικόνα 4.66: Η τιμή Loglikelihood για  $k = 4$  θέματα Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.67: Η τιμή Perplexity για  $k = 4$  θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.68: Το μοντέλο LDA για k = 4 θέματα, Controlled Clinical Trial – περιλήψεις.



Εικόνα 4.69: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για k = 4 θέματα, Controlled Clinical Trial – περιλήψεις.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

Result History ExampleSet (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA)) LDAModel (Extract Topics from Documents (LDA))

Open in Turbo Prep Auto Model Filter (307 / 307 examples): all

Row No.	documentid	prediction(T...	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_3	0.000	0.079	0.014	0.907	BACKGROU...
2	1	Topic_0	0.803	0.019	0.104	0.073	Considering L...
3	2	Topic_2	0.343	0.148	0.418	0.091	OBJECTIVE: ...
4	3	Topic_3	0.042	0.036	0.097	0.825	BACKGROU...
5	4	Topic_0	0.359	0.321	0.317	0.003	Early diagnos...
6	5	Topic_1	0.001	0.361	0.345	0.294	BACKGROU...
7	6	Topic_2	0.179	0.120	0.361	0.340	BACKGROU...
8	7	Topic_0	0.397	0.306	0.112	0.185	BACKGROU...
9	8	Topic_3	0.000	0.000	0.333	0.666	BACKGROU...
10	9	Topic_0	0.792	0.188	0.018	0.002	BACKGROU...
11	10	Topic_3	0.061	0.105	0.000	0.834	INTRODUCTI...
12	11	Topic_3	0.002	0.310	0.083	0.605	
13	12	Topic_2	0.000	0.203	0.796	0.000	INTRODUCTI...
14	13	Topic_1	0.000	0.559	0.164	0.277	Background: ...
15	14	Topic_3	0.139	0.000	0.074	0.787	BACKGROU...
16	15	Topic_1	0.000	0.684	0.310	0.005	Whether HDL...
17	16	Topic_3	0.000	0.079	0.000	0.921	BACKGROU...
18	17	Topic_2	0.177	0.001	0.562	0.260	BACKGROU...
19	18	Topic_3	0.089	0.149	0.000	0.762	BACKGROU...
20	19	Topic_0	0.864	0.062	0.000	0.074	People with A...
21	20	Topic_0	0.772	0.070	0.122	0.035	BACKGROU...
22	21	Topic_1	0.000	0.683	0.194	0.122	BACKGROU...
23	22	Topic_0	0.444	0.002	0.246	0.308	BACKGROU...
24	23	Topic_3	0.004	0.085	0.000	0.911	IMPORTANC...
25	24	Topic_3	0.000	0.329	0.039	0.632	BACKGROU...
26	25	Topic_3	0.030	0.089	0.002	0.879	HYPOTHESE...
27	26	Topic_1	0.269	0.425	0.000	0.305	BACKGROU...
28	27	Topic_0	0.913	0.004	0.074	0.009	In reports of r...
29	28	Topic_0	0.860	0.000	0.139	0.000	OBJECTIVE: ...
30	29	Topic_3	0.015	0.315	0.000	0.670	AIMS: Cholin...
31	30	Topic_0	0.777	0.200	0.021	0.002	OBJECTIVE: ...
32	31	Topic_2	0.085	0.086	0.574	0.255	Alzheimer's d...
33	32	Topic_2	0.303	0.002	0.479	0.217	BACKGROU...
34	33	Topic_2	0.143	0.334	0.522	0.000	BACKGROU...
35	34	Topic_3	0.172	0.000	0.000	0.828	OBJECTIVE: ...
36	35	Topic_2	0.000	0.000	0.610	0.389	BACKGROU...
37	36	Topic_2	0.284	0.000	0.715	0.001	The stigma a...
38	37	Topic_3	0.000	0.211	0.000	0.789	BACKGROU...
39	38	Topic_3	0.019	0.319	0.000	0.662	The LUCIND...
40	39	Topic_3	0.154	0.000	0.214	0.632	INTRODUCTI...
41	40	Topic_0	0.785	0.001	0.214	0.000	Adults with Al...
42	41	Topic_2	0.054	0.312	0.606	0.029	BACKGROU...
43	42	Topic_2	0.003	0.015	0.806	0.176	BACKGROU...
44	43	Topic_3	0.003	0.383	0.001	0.612	BACKGROU...
45	44	Topic_0	0.521	0.014	0.090	0.375	BACKGROU...
46	45	Topic_2	0.173	0.089	0.607	0.131	Alzheimer's d...
47	46	Topic_2	0.193	0.215	0.397	0.196	BACKGROU...
48	47	Topic_0	0.572	0.110	0.103	0.215	INTRODUCTI...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Statistics	49	48	Topic_0	0.667	0.001	0.238	0.094	Alzheimer's d...	
	50	49	Topic_2	0.000	0.116	0.864	0.019	BACKGROU...	
	51	50	Topic_2	0.144	0.139	0.716	0.000	Episodic me...	
	52	51	Topic_0	0.999	0.000	0.000	0.000	Technologies...	
	Visualizations	53	52	Topic_1	0.157	0.541	0.000	0.301	BACKGROU...
		54	53	Topic_1	0.468	0.525	0.002	0.006	BACKGROU...
	Annotations	55	54	Topic_0	0.748	0.000	0.000	0.251	
		56	55	Topic_0	0.934	0.000	0.065	0.000	BACKGROU...
		57	56	Topic_0	0.789	0.001	0.206	0.005	BACKGROU...
		58	57	Topic_2	0.000	0.475	0.525	0.000	BACKGROU...
		59	58	Topic_0	0.924	0.008	0.068	0.000	BACKGROU...
		60	59	Topic_3	0.159	0.000	0.330	0.511	The prevalen...
		61	60	Topic_3	0.000	0.000	0.431	0.569	PURPOSE: T...
		62	61	Topic_1	0.118	0.578	0.154	0.149	APOE encod...
63		62	Topic_0	0.565	0.111	0.225	0.099	PURPOSE: T...	
64		63	Topic_3	0.000	0.475	0.031	0.493	BACKGROU...	
65		64	Topic_2	0.056	0.144	0.775	0.024	Accurate mea...	
66		65	Topic_3	0.000	0.280	0.000	0.719	Tau plays a K...	
67		66	Topic_3	0.000	0.215	0.000	0.785	BACKGROU...	
68		67	Topic_0	0.967	0.033	0.000	0.000	OBJECTIVE: ...	
69	68	Topic_2	0.318	0.293	0.388	0.000	Mild cognitive...		
70	69	Topic_3	0.129	0.001	0.430	0.440	BACKGROU...		
71	70	Topic_1	0.280	0.368	0.285	0.068	Microglia play...		
72	71	Topic_1	0.000	0.547	0.208	0.245	Neuroinflam...		
Statistics	73	72	Topic_3	0.001	0.093	0.257	0.649	BACKGROU...	
	74	73	Topic_3	0.000	0.099	0.197	0.704	BACKGROU...	
	75	74	Topic_3	0.130	0.000	0.188	0.682	INTRODUCTI...	
	76	75	Topic_0	0.905	0.001	0.079	0.015	BACKGROU...	
	Visualizations	77	76	Topic_0	0.981	0.000	0.002	0.018	BACKGROU...
		78	77	Topic_1	0.103	0.577	0.319	0.001	INTRODUCTI...
	Annotations	79	78	Topic_3	0.140	0.289	0.001	0.570	AIM: Previous...
		80	79	Topic_0	0.768	0.000	0.229	0.003	BACKGROU...
		81	80	Topic_2	0.486	0.000	0.510	0.003	PURPOSE: M...
		82	81	Topic_2	0.000	0.082	0.579	0.339	BACKGROU...
		83	82	Topic_0	0.956	0.000	0.043	0.000	Efficacy of as...
		84	83	Topic_3	0.338	0.008	0.000	0.654	OBJECTIVE: ...
		85	84	Topic_2	0.026	0.000	0.845	0.129	BACKGROU...
		86	85	Topic_1	0.000	0.523	0.081	0.397	BACKGROU...
87		86	Topic_2	0.134	0.054	0.808	0.004	The aim of thi...	
88		87	Topic_2	0.189	0.000	0.434	0.377	Spirulina max...	
89		88	Topic_3	0.042	0.113	0.354	0.491	INTRODUCTI...	
90		89	Topic_3	0.004	0.000	0.093	0.903	BACKGROU...	
91		90	Topic_2	0.064	0.003	0.661	0.272	BACKGROU...	
92		91	Topic_0	0.850	0.000	0.000	0.149		
93	92	Topic_2	0.124	0.093	0.505	0.278	BACKGROU...		
94	93	Topic_1	0.089	0.832	0.048	0.031	BACKGROU...		
95	94	Topic_0	0.428	0.365	0.093	0.114	BACKGROU...		
96	95	Topic_2	0.307	0.015	0.624	0.053	Probiotic sup...		
Statistics	97	96	Topic_0	0.481	0.193	0.257	0.068	The progress...	
	98	97	Topic_1	0.054	0.841	0.098	0.007	BACKGROU...	
	99	98	Topic_3	0.110	0.307	0.010	0.573	IMPORTANC...	
	100	99	Topic_1	0.061	0.854	0.084	0.000	BACKGROU...	
	Visualizations	101	100	Topic_1	0.000	0.760	0.078	0.162	OBJECTIVES...
		102	101	Topic_2	0.000	0.257	0.565	0.178	BACKGROU...
	Annotations	103	102	Topic_2	0.028	0.195	0.777	0.000	A poor Fines...
		104	103	Topic_3	0.078	0.073	0.032	0.816	IMPORTANC...
		105	104	Topic_2	0.119	0.000	0.873	0.007	INTRODUCTI...
		106	105	Topic_2	0.000	0.000	0.677	0.322	BACKGROU...
		107	106	Topic_0	0.572	0.000	0.289	0.139	BACKGROU...
		108	107	Topic_2	0.232	0.041	0.727	0.000	Learning and...
		109	108	Topic_0	0.915	0.002	0.082	0.001	BACKGROU...
		110	109	Topic_2	0.232	0.144	0.516	0.107	
111		110	Topic_3	0.001	0.217	0.162	0.620	BACKGROU...	
112		111	Topic_3	0.061	0.374	0.001	0.564	BACKGROU...	
113		112	Topic_1	0.061	0.540	0.269	0.130	OBJECTIVE: ...	
114		113	Topic_3	0.007	0.086	0.000	0.906	AIMS: HTL00...	
115		114	Topic_0	0.491	0.271	0.238	0.000	INTRODUCTI...	
116		115	Topic_3	0.064	0.000	0.000	0.936	AIMS: HTL00...	
117	116	Topic_2	0.000	0.434	0.475	0.091	Subjective m...		
118	117	Topic_0	0.564	0.014	0.005	0.418	BACKGROU...		
119	118	Topic_0	0.829	0.000	0.170	0.000	(1) Backgrou...		
120	119	Topic_0	0.518	0.001	0.480	0.001	BACKGROU...		

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Statistics	121	120	Topic_1	0.000	0.635	0.126	0.239	OBJECTIVE: ...
	122	121	Topic_3	0.090	0.000	0.000	0.909	BACKGROU...
	123	122	Topic_1	0.075	0.465	0.000	0.460	BACKGROU...
	124	123	Topic_0	0.609	0.000	0.390	0.000	BACKGROU...
	125	124	Topic_0	0.584	0.000	0.116	0.300	Nighttime agl...
	126	125	Topic_1	0.000	0.835	0.165	0.000	Apolliprote...
	127	126	Topic_3	0.064	0.320	0.000	0.616	Research ha...
	128	127	Topic_3	0.145	0.000	0.366	0.489	BACKGROU...
	129	128	Topic_2	0.229	0.004	0.767	0.000	This study utl...
	130	129	Topic_3	0.325	0.000	0.000	0.675	RATIONALE: ...
	131	130	Topic_2	0.106	0.227	0.668	0.000	BACKGROU...
	132	131	Topic_0	0.998	0.000	0.002	0.000	The "Great VII...
	133	132	Topic_1	0.000	0.766	0.109	0.125	BACKGROU...
	134	133	Topic_0	0.879	0.001	0.019	0.101	OBJECTIVES...
	135	134	Topic_2	0.000	0.150	0.848	0.002	
	136	135	Topic_2	0.350	0.000	0.648	0.001	INTRODUCTI...
	137	136	Topic_2	0.141	0.169	0.690	0.000	BACKGROU...
	138	137	Topic_2	0.267	0.002	0.543	0.188	BACKGROU...
	139	138	Topic_0	0.912	0.000	0.002	0.086	AIMS: We con...
	140	139	Topic_3	0.218	0.003	0.237	0.541	
141	140	Topic_0	0.945	0.000	0.055	0.000	BACKGROU...	
142	141	Topic_1	0.403	0.475	0.059	0.064	BACKGROU...	
143	142	Topic_0	0.999	0.000	0.000	0.000	OBJECTIVES...	
144	143	Topic_1	0.040	0.492	0.044	0.424		

Statistics	145	144	Topic_0	0.870	0.000	0.000	0.129	BACKGROU...
	146	145	Topic_2	0.000	0.000	1.000	0.000	The purpose ...
	147	146	Topic_3	0.000	0.314	0.000	0.686	BACKGROU...
	148	147	Topic_3	0.006	0.044	0.036	0.914	PURPOSE: S...
	149	148	Topic_3	0.125	0.089	0.000	0.786	
	150	149	Topic_3	0.000	0.000	0.305	0.695	BACKGROU...
	151	150	Topic_2	0.000	0.000	1.000	0.000	
	152	151	Topic_3	0.000	0.046	0.159	0.795	BACKGROU...
	153	152	Topic_1	0.002	0.722	0.272	0.004	INTRODUCTI...
	154	153	Topic_0	0.694	0.000	0.297	0.009	BACKGROU...
	155	154	Topic_1	0.000	0.573	0.042	0.385	INTRODUCTI...
	156	155	Topic_2	0.075	0.000	0.859	0.056	Hippocampal...
	157	156	Topic_3	0.083	0.317	0.002	0.598	AIMS: TAK-07...
	158	157	Topic_1	0.244	0.402	0.348	0.007	BACKGROU...
	159	158	Topic_2	0.352	0.080	0.568	0.000	BACKGROU...
	160	159	Topic_2	0.254	0.076	0.517	0.152	BACKGROU...
	161	160	Topic_1	0.024	0.802	0.012	0.162	Serum amylo...
	162	161	Topic_2	0.033	0.348	0.608	0.011	Long-term us...
	163	162	Topic_0	0.876	0.000	0.124	0.000	OBJECTIVES...
	164	163	Topic_1	0.001	0.392	0.272	0.334	IMPORTANC...
165	164	Topic_0	0.704	0.000	0.286	0.010	OBJECTIVES...	
166	165	Topic_3	0.152	0.000	0.221	0.627	IMPORTANC...	
167	166	Topic_0	0.896	0.001	0.000	0.103	BACKGROU...	
168	167	Topic_1	0.000	0.546	0.345	0.110	BACKGROU...	

Statistics	169	168	Topic_3	0.000	0.343	0.002	0.655	IMPORTANC...
	170	169	Topic_0	0.849	0.001	0.000	0.150	BACKGROU...
	171	170	Topic_3	0.161	0.000	0.207	0.632	INTRODUCTI...
	172	171	Topic_2	0.035	0.227	0.592	0.146	
	173	172	Topic_2	0.242	0.001	0.690	0.067	BACKGROU...
	174	173	Topic_3	0.091	0.005	0.271	0.633	Desalted Sali...
	175	174	Topic_3	0.000	0.233	0.180	0.586	Background: ...
	176	175	Topic_0	0.511	0.151	0.338	0.000	BACKGROU...
	177	176	Topic_3	0.109	0.018	0.059	0.814	OBJECTIVES...
	178	177	Topic_0	0.509	0.382	0.108	0.000	
	179	178	Topic_1	0.110	0.615	0.000	0.276	BACKGROU...
	180	179	Topic_0	0.995	0.003	0.000	0.002	INTRODUCTI...
	181	180	Topic_2	0.087	0.218	0.584	0.111	IMPORTANC...
	182	181	Topic_0	0.922	0.000	0.078	0.000	BACKGROU...
	183	182	Topic_0	0.941	0.058	0.000	0.000	BACKGROU...
	184	183	Topic_1	0.003	0.581	0.329	0.087	BACKGROU...
	185	184	Topic_3	0.053	0.279	0.082	0.587	OBJECTIVE: ...
	186	185	Topic_0	0.858	0.140	0.001	0.001	OBJECTIVES...
	187	186	Topic_2	0.104	0.179	0.622	0.094	The effects of...
	188	187	Topic_3	0.000	0.441	0.079	0.480	IMPORTANC...
189	188	Topic_3	0.000	0.317	0.249	0.434	HIV-associat...	
190	189	Topic_2	0.000	0.290	0.545	0.164	BACKGROU...	
191	190	Topic_0	0.901	0.000	0.099	0.000	BACKGROU...	
192	191	Topic_2	0.057	0.008	0.673	0.263	IMPORTANC...	

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

193	192	Topic_1	0.077	0.710	0.077	0.137	IMPORTANC...
194	193	Topic_0	0.877	0.001	0.015	0.106	OBJECTIVES...
195	194	Topic_2	0.000	0.105	0.893	0.001	The CAIDE (...)
196	195	Topic_2	0.132	0.000	0.435	0.433	OBJECTIVES...
197	196	Topic_0	0.657	0.179	0.000	0.164	BACKGROU...
198	197	Topic_2	0.330	0.000	0.669	0.001	OBJECTIVE: ...
199	198	Topic_1	0.011	0.442	0.227	0.319	The locus co...
200	199	Topic_0	0.980	0.011	0.000	0.009	OBJECTIVE: ...
201	200	Topic_3	0.174	0.000	0.000	0.826	INTRODUCTI...
202	201	Topic_1	0.400	0.588	0.002	0.000	BACKGROU...
203	202	Topic_2	0.103	0.000	0.475	0.422	We previousl...
204	203	Topic_0	0.583	0.156	0.003	0.259	BACKGROU...
205	204	Topic_2	0.405	0.003	0.592	0.000	PURPOSE: In...
206	205	Topic_2	0.388	0.000	0.611	0.000	AIM: To invest...
207	206	Topic_3	0.000	0.054	0.185	0.760	
208	207	Topic_0	0.932	0.000	0.066	0.002	
209	208	Topic_3	0.000	0.284	0.046	0.671	IMPORTANC...
210	209	Topic_3	0.239	0.000	0.202	0.559	BACKGROU...
211	210	Topic_3	0.125	0.103	0.160	0.612	IMPORTANC...
212	211	Topic_1	0.068	0.440	0.063	0.429	IMPORTANC...
213	212	Topic_0	0.877	0.122	0.000	0.000	BACKGROU...
214	213	Topic_1	0.001	0.520	0.425	0.055	OBJECTIVE: ...
215	214	Topic_2	0.000	0.418	0.582	0.000	OBJECTIVE: ...
216	215	Topic_3	0.116	0.024	0.166	0.695	BACKGROU...

217	216	Topic_1	0.311	0.436	0.175	0.078	Transcranial ...
218	217	Topic_3	0.001	0.156	0.030	0.813	IMPORTANC...
219	218	Topic_3	0.078	0.093	0.259	0.569	
220	219	Topic_1	0.264	0.734	0.002	0.000	IMPORTANC...
221	220	Topic_2	0.000	0.001	0.724	0.274	BACKGROU...
222	221	Topic_3	0.000	0.083	0.000	0.917	BACKGROU...
223	222	Topic_3	0.201	0.011	0.303	0.485	BACKGROU...
224	223	Topic_0	0.691	0.282	0.027	0.001	IMPORTANC...
225	224	Topic_0	0.961	0.001	0.000	0.038	OBJECTIVES...
226	225	Topic_3	0.000	0.088	0.159	0.752	BACKGROU...
227	226	Topic_0	0.711	0.002	0.287	0.000	BACKGROU...
228	227	Topic_1	0.000	0.622	0.156	0.222	OBJECTIVE: ...
229	228	Topic_0	0.469	0.226	0.222	0.082	BACKGROU...
230	229	Topic_0	0.894	0.097	0.000	0.008	BACKGROU...
231	230	Topic_2	0.002	0.269	0.727	0.002	BACKGROU...
232	231	Topic_3	0.216	0.000	0.000	0.784	IMPORTANC...
233	232	Topic_1	0.000	0.523	0.410	0.067	OBJECTIVE: ...
234	233	Topic_3	0.108	0.219	0.006	0.667	IMPORTANC...
235	234	Topic_3	0.000	0.370	0.000	0.630	IMPORTANC...
236	235	Topic_3	0.054	0.355	0.000	0.591	IMPORTANC...
237	236	Topic_2	0.440	0.094	0.465	0.000	The use of ec...
238	237	Topic_1	0.003	0.493	0.348	0.156	BACKGROU...
239	238	Topic_2	0.096	0.080	0.824	0.000	White matter ...
240	239	Topic_1	0.009	0.497	0.230	0.264	BACKGROU...

241	240	Topic_3	0.000	0.039	0.057	0.904	INTRODUCTI...
242	241	Topic_2	0.255	0.123	0.353	0.270	BACKGROU...
243	242	Topic_3	0.000	0.044	0.317	0.639	BACKGROU...
244	243	Topic_3	0.001	0.432	0.000	0.567	Dominantly in...
245	244	Topic_1	0.000	0.607	0.393	0.000	Dietary suppl...
246	245	Topic_2	0.000	0.310	0.690	0.000	BACKGROU...
247	246	Topic_3	0.000	0.083	0.331	0.586	Dysregulatio...
248	247	Topic_0	0.525	0.000	0.000	0.474	BACKGROU...
249	248	Topic_2	0.198	0.000	0.726	0.076	BACKGROU...
250	249	Topic_2	0.173	0.000	0.692	0.135	The treatmen...
251	250	Topic_2	0.099	0.000	0.901	0.001	Dementia aff...
252	251	Topic_3	0.000	0.169	0.000	0.831	
253	252	Topic_2	0.236	0.000	0.569	0.195	Close to 6 mi...
254	253	Topic_2	0.000	0.461	0.506	0.033	OBJECTIVE: ...
255	254	Topic_3	0.001	0.284	0.269	0.445	Current thera...
256	255	Topic_3	0.001	0.003	0.081	0.915	BACKGROU...
257	256	Topic_3	0.000	0.021	0.034	0.944	We conducte...
258	257	Topic_3	0.129	0.139	0.000	0.731	BACKGROU...
259	258	Topic_2	0.071	0.010	0.643	0.276	OBJECTIVE: ...
260	259	Topic_1	0.002	0.592	0.406	0.000	Tau hyperpho...
261	260	Topic_2	0.172	0.000	0.828	0.000	OBJECTIVE: ...
262	261	Topic_2	0.004	0.130	0.610	0.256	In the phase ...
263	262	Topic_3	0.000	0.000	0.184	0.816	INTRODUCTI...
264	263	Topic_2	0.095	0.000	0.804	0.100	PURPOSE: V...

265	264	Topic_1	0.257	0.577	0.162	0.005	OBJECTIVE: ...
266	265	Topic_0	0.408	0.000	0.279	0.312	BACKGROU...
267	266	Topic_2	0.058	0.055	0.500	0.387	OBJECTIVE: ...
268	267	Topic_1	0.000	0.610	0.000	0.390	IMPORTANC...
269	268	Topic_0	0.865	0.135	0.000	0.000	Objectives: Pr...
270	269	Topic_1	0.000	0.529	0.312	0.158	Intranasal ins...
271	270	Topic_2	0.152	0.000	0.634	0.214	BACKGROU...
272	271	Topic_1	0.000	0.862	0.089	0.048	BACKGROU...
273	272	Topic_3	0.105	0.032	0.340	0.523	Non-invasve ...
274	273	Topic_3	0.000	0.160	0.001	0.839	IMPORTANC...
275	274	Topic_2	0.036	0.157	0.807	0.000	BACKGROU...
276	275	Topic_2	0.000	0.001	0.515	0.483	Repetitive tra...
277	276	Topic_1	0.000	0.732	0.260	0.007	OBJECTIVE: ...
278	277	Topic_2	0.304	0.026	0.464	0.207	OBJECTIVE: ...
279	278	Topic_2	0.096	0.000	0.903	0.000	BACKGROU...
280	279	Topic_1	0.000	0.554	0.000	0.446	RIPK1 is a m...
281	280	Topic_2	0.000	0.000	0.827	0.173	INTRODUCTI...
282	281	Topic_0	0.601	0.000	0.317	0.082	OBJECTIVES...
283	282	Topic_2	0.023	0.209	0.763	0.005	Mild cognitive...
284	283	Topic_2	0.073	0.008	0.535	0.383	No prior studi...
285	284	Topic_3	0.243	0.007	0.308	0.442	IMPORTANC...
286	285	Topic_2	0.226	0.248	0.522	0.004	BACKGROU...
287	286	Topic_2	0.018	0.253	0.728	0.000	Increasing ev...
288	287	Topic_0	0.386	0.000	0.371	0.243	We examined...
289	288	Topic_1	0.000	0.547	0.372	0.082	BACKGROU...
290	289	Topic_3	0.156	0.072	0.000	0.772	IMPORTANC...
291	290	Topic_2	0.004	0.000	0.880	0.115	Aerobic exerc...
292	291	Topic_3	0.000	0.115	0.001	0.884	IMPORTANC...
293	292	Topic_0	0.729	0.112	0.000	0.158	IMPORTANC...
294	293	Topic_3	0.038	0.187	0.113	0.661	
295	294	Topic_2	0.115	0.000	0.688	0.197	Probiotics co...
296	295	Topic_0	0.739	0.051	0.208	0.002	BACKGROU...
297	296	Topic_2	0.352	0.128	0.520	0.000	BACKGROU...
298	297	Topic_3	0.099	0.150	0.324	0.428	BACKGROU...
299	298	Topic_2	0.091	0.276	0.633	0.000	There is eme...
300	299	Topic_2	0.000	0.044	0.594	0.361	IMPORTANC...
301	300	Topic_2	0.340	0.000	0.374	0.286	BACKGROU...
302	301	Topic_3	0.000	0.097	0.005	0.897	Tau accumul...
303	302	Topic_1	0.000	0.623	0.000	0.377	OBJECTIVE: ...
304	303	Topic_2	0.117	0.057	0.825	0.001	BACKGROU...
305	304	Topic_3	0.035	0.080	0.398	0.488	BACKGROU...
306	305	Topic_2	0.222	0.000	0.655	0.122	Age-related d...
307	306	Topic_1	0.096	0.429	0.391	0.084	

ExampleSet (307 examples, 6 special attributes, 1 regular attribute)

Εικόνα 4.70: Η κατηγοριοποίηση των περιλήψεων Controlled Clinical Trial σε  $k = 4$  θέματα.

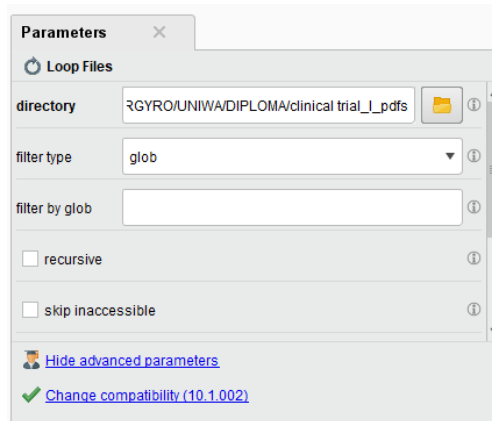
## 4.4 Εφαρμογή σε βιοϊατρικά πλήρη κείμενα και αποτελέσματα

Στην ενότητα 4.4 εισάγονται στην δομημένη διεργασία **τα πλήρη κείμενα (full texts)** των κατηγοριών Clinical Trial I, Clinical Trial II και Controlled Clinical Trial, με τους αντίστοιχους αριθμούς κειμένων που αναφέρθηκαν στην ενότητα 4.3. Ακολούθως θα παρατεθούν τα αποτελέσματα του RapidMiner για αριθμό θεμάτων  $k = 2$ ,  $k = 3$  και  $k = 4$ , τα οποία θα σχολιαστούν στην ενότητα 4.5.

### 4.4.1 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας Clinical Trial I

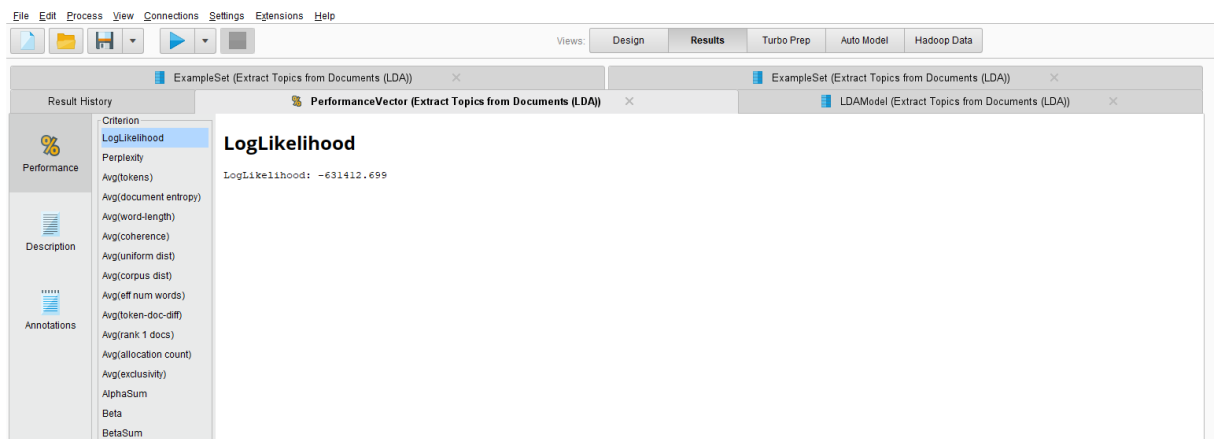
Στην ακόλουθη εικόνα το αρχείο που τοποθετείται στο directory του χειριστή “Loop Files” είναι το “clinical\_trial\_I\_pdfs” για να εισαχθούν τα πλήρη κείμενα της ομάδας Clinical Trial I.



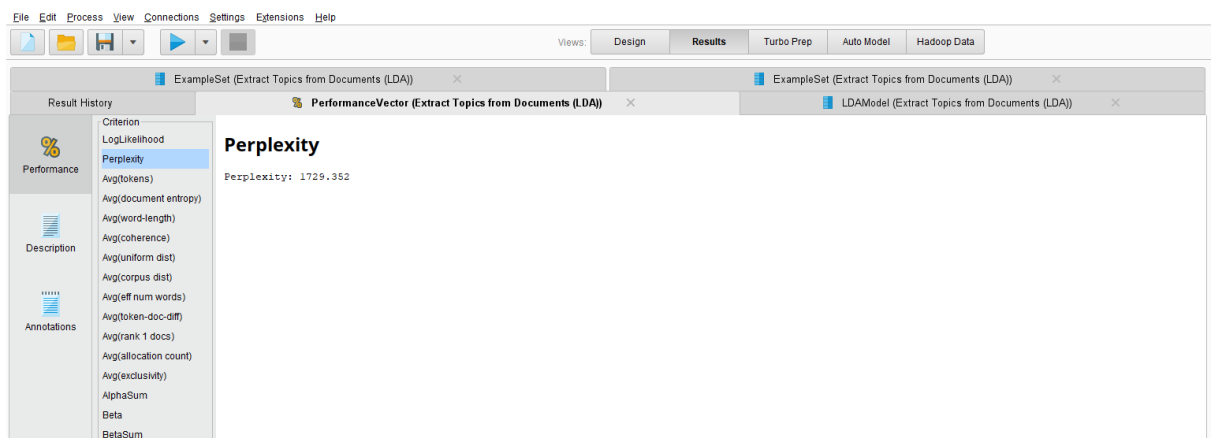


Εικόνα 4.71: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων του Clinical Trial I.

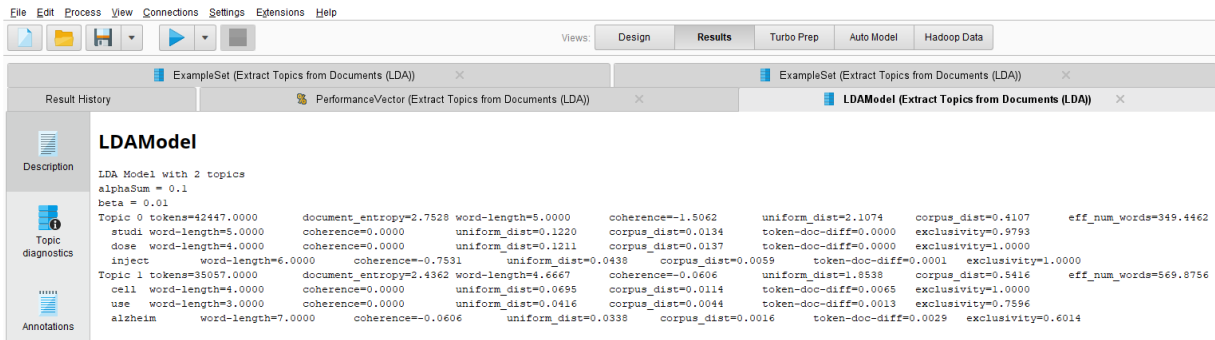
Αποτελέσματα για αριθμό θεμάτων  $k = 2$ :



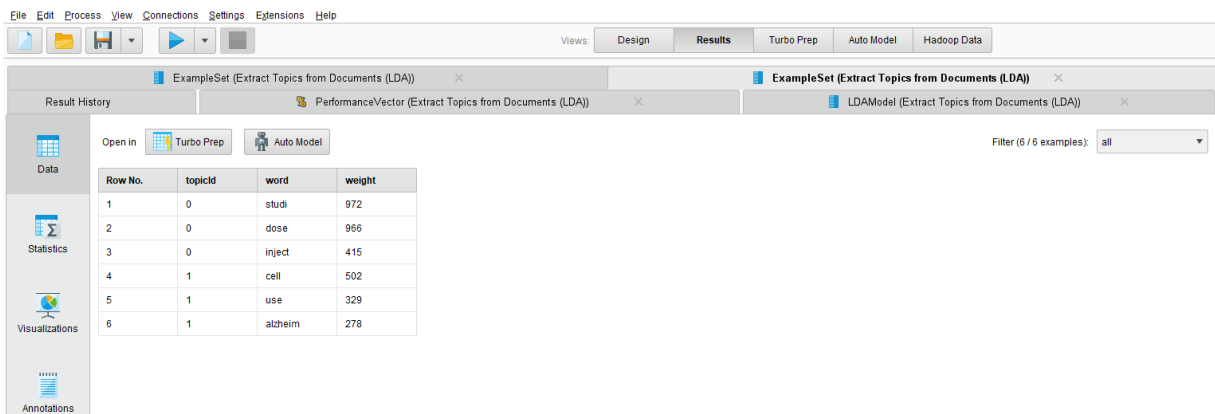
Εικόνα 4.72: Η τιμή Loglikelihood για  $k = 2$  θέματα, Clinical Trial I - πλήρη κείμενα.



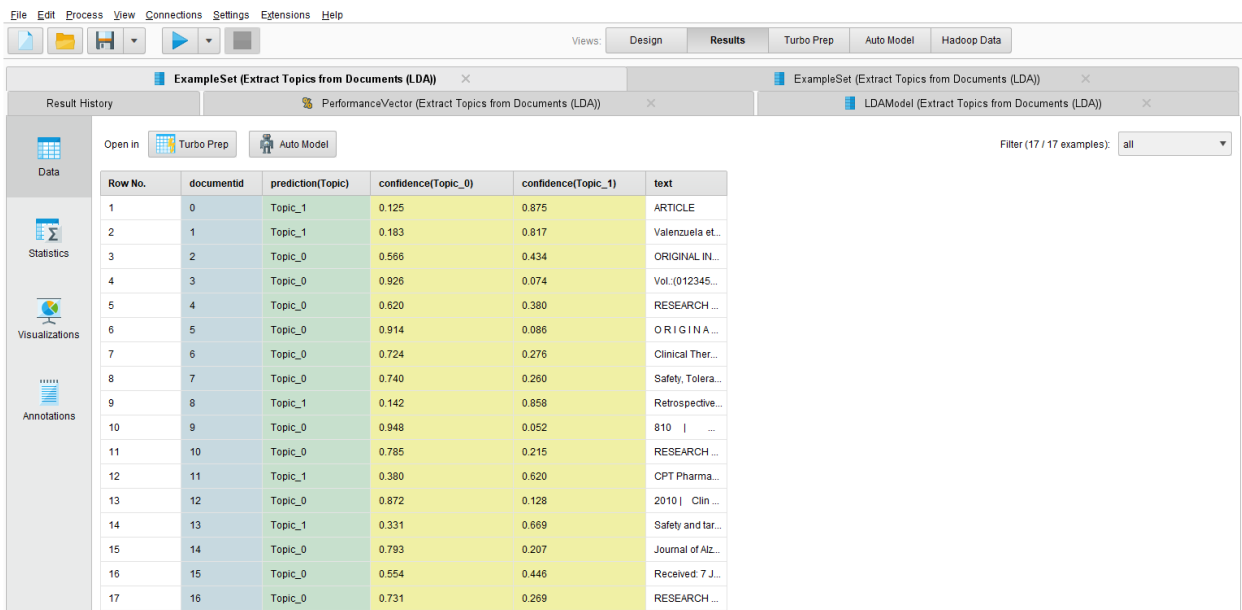
Εικόνα 4.73: Η τιμή Perplexity για  $k = 2$  θέματα, Clinical Trial I – πλήρη κείμενα.



Εικόνα 4.74: Το μοντέλο LDA για k = 2 θέματα, Clinical Trial I – πλήρη κείμενα.

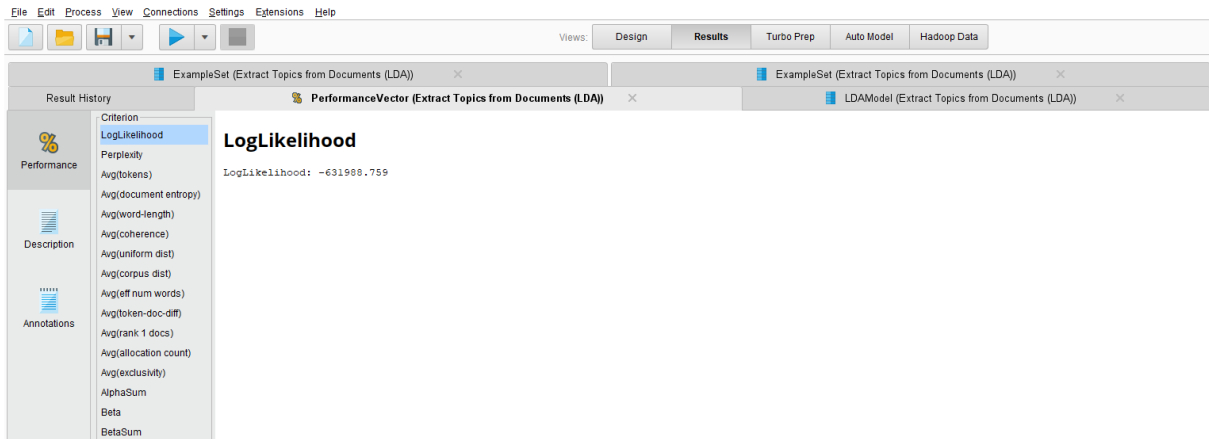


Εικόνα 4.75: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για k = 2 θέματα, Clinical Trial I – πλήρη κείμενα.

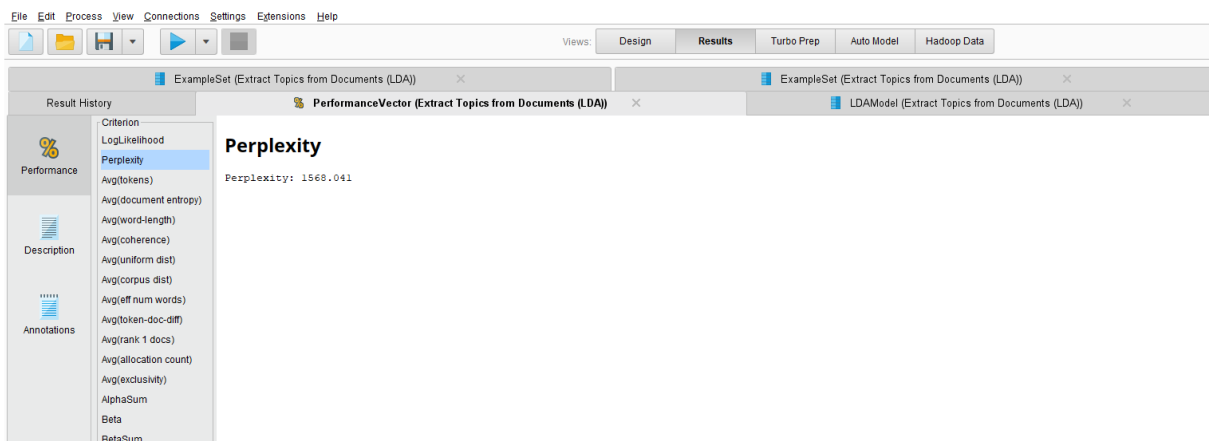


Εικόνα 4.76: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε k = 2 θέματα.

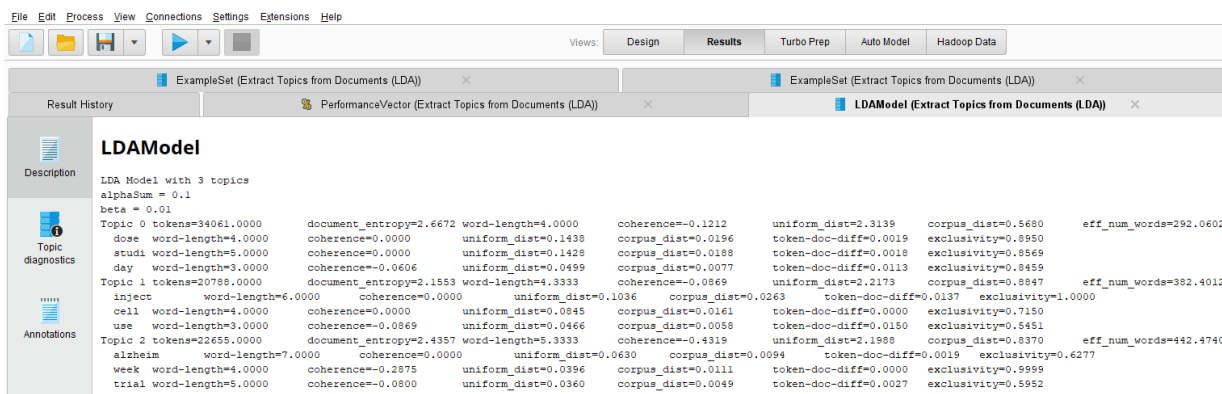
Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :



Εικόνα 4.77: Η τιμή Loglikelihood για  $k = 3$  θέματα, Clinical Trial I – πλήρη κείμενα.



Εικόνα 4.78: Η τιμή Perplexity για  $k = 3$  θέματα Clinical Trial I - πλήρη κείμενα.



Εικόνα 4.79: Το μοντέλο LDA για  $k = 3$  θέματα, Clinical Trial – πλήρη κείμενα.

Row No.	topicId	word	weight
1	0	dose	896
2	0	studi	891
3	0	day	371
4	1	inject	415
5	1	cell	350
6	1	use	214
7	2	alzheim	299
8	2	week	204
9	2	trial	189

Εικόνα 4.80: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 3$  θέματα, Clinical Trial I – πλήρη κείμενα.

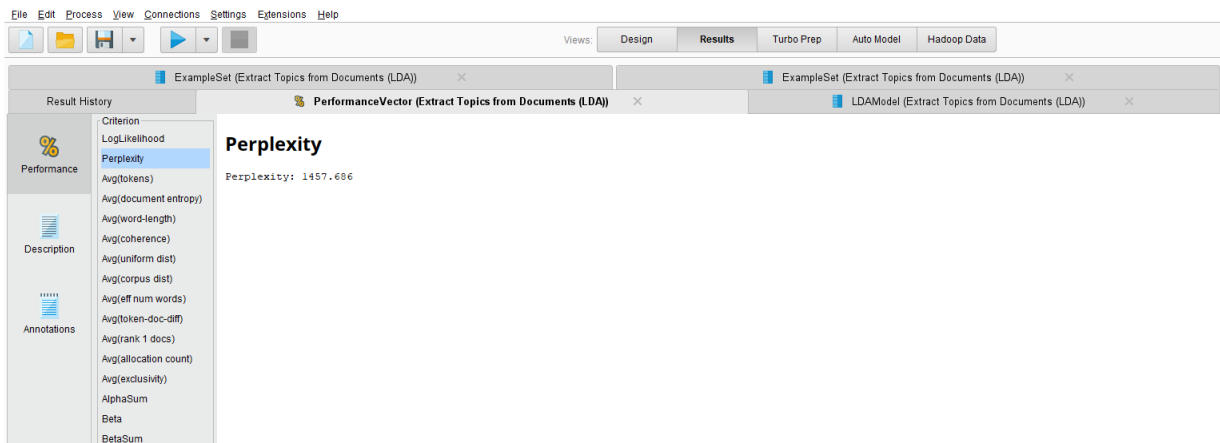
Row No.	documentId	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_1	0.112	0.840	0.048	ARTICLE
2	1	Topic_2	0.072	0.363	0.565	Valenzuela et...
3	2	Topic_0	0.759	0.095	0.146	ORIGINAL IN...
4	3	Topic_0	0.899	0.015	0.086	Vol.:012345...
5	4	Topic_0	0.505	0.136	0.359	RESEARCH...
6	5	Topic_0	0.868	0.039	0.093	O R I G I N A...
7	6	Topic_1	0.410	0.512	0.077	Clinical Ther...
8	7	Topic_0	0.702	0.050	0.248	Safety, Toler...
9	8	Topic_2	0.122	0.029	0.849	Retrospective...
10	9	Topic_0	0.887	0.012	0.101	810   I ...
11	10	Topic_1	0.269	0.383	0.328	RESEARCH...
12	11	Topic_1	0.402	0.488	0.110	CPT Pharma...
13	12	Topic_0	0.800	0.076	0.124	2010   Clin ...
14	13	Topic_2	0.228	0.040	0.733	Safety and tar...
15	14	Topic_0	0.693	0.109	0.197	Journal of Alz...
16	15	Topic_2	0.406	0.079	0.515	Received: 7 J...
17	16	Topic_1	0.110	0.759	0.131	RESEARCH...

Εικόνα 4.81: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε  $k = 3$  θέματα.

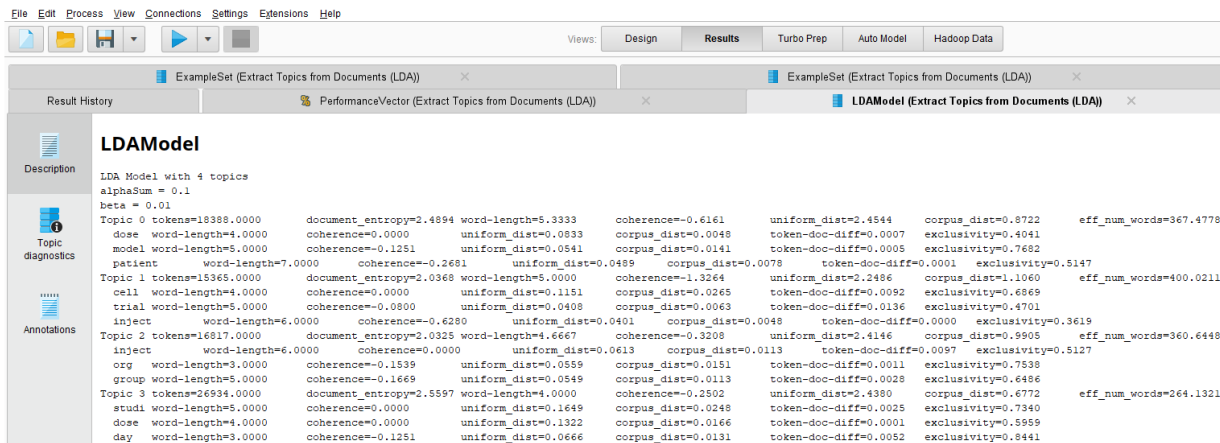
Αποτελέσματα για αριθμό θεμάτων  $k = 4$ :

**LogLikelihood**  
LogLikelihood: -634819.019

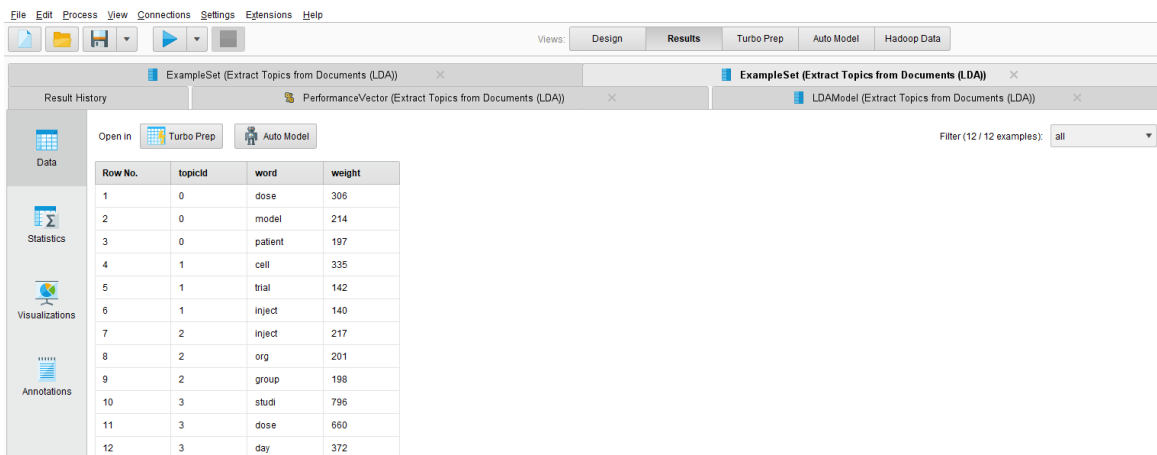
Εικόνα 4.82: Η τιμή Loglikelihood για  $k = 4$  θέματα, Clinical Trial I – πλήρη κείμενα.



Εικόνα 4.83: Η τιμή Perplexity για k = 4 θέματα, Clinical Trial I – πλήρη κείμενα.



Εικόνα 4.84: Το μοντέλο LDA για k = 4 θέματα, Clinical Trial I – πλήρη κείμενα.



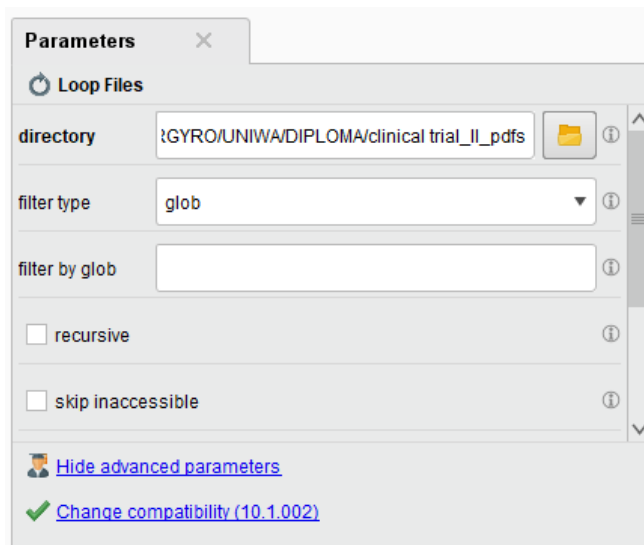
Εικόνα 4.85: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για k = 4 θέματα, Clinical Trial I – πλήρη κείμενα.

Row No.	documentId	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_2	0.060	0.103	0.762	0.075	ARTICLE
2	1	Topic_1	0.082	0.802	0.053	0.063	Valenzuela et...
3	2	Topic_3	0.083	0.095	0.200	0.622	ORIGINAL IN...
4	3	Topic_0	0.496	0.031	0.032	0.441	Vol.(012345...
5	4	Topic_0	0.717	0.019	0.036	0.227	RESEARCH ...
6	5	Topic_3	0.059	0.075	0.015	0.851	O R I G I N A ...
7	6	Topic_2	0.176	0.035	0.496	0.293	Clinical Ther...
8	7	Topic_3	0.073	0.092	0.122	0.713	Safety, Toler...
9	8	Topic_2	0.073	0.046	0.745	0.136	Retrospective...
10	9	Topic_3	0.053	0.044	0.007	0.897	810   I ...
11	10	Topic_1	0.128	0.631	0.057	0.183	RESEARCH ...
12	11	Topic_0	0.827	0.023	0.030	0.121	CPT Pharma...
13	12	Topic_3	0.111	0.068	0.039	0.781	2010   Clin ...
14	13	Topic_0	0.633	0.132	0.114	0.120	Safety and tar...
15	14	Topic_3	0.198	0.108	0.063	0.632	Journal of Alz...
16	15	Topic_2	0.140	0.233	0.350	0.277	Received. 7 J...
17	16	Topic_0	0.656	0.253	0.051	0.041	RESEARCH ...

Εικόνα 4.86: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial I σε k = 4 θέματα.

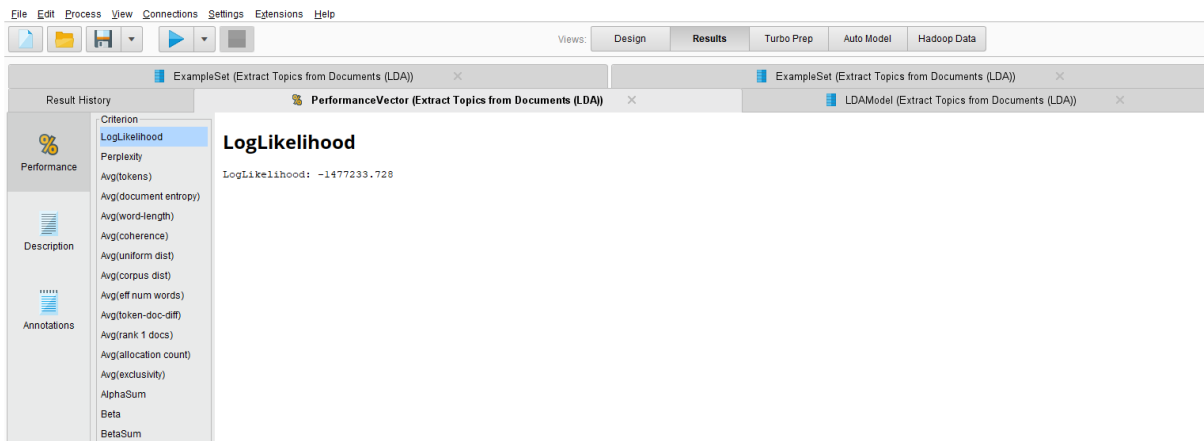
#### 4.4.2 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας Clinical Trial II

Η εισαγωγή των κειμένων Clinical Trial II πραγματοποιήθηκε με αλλαγή του directory του χειριστή “Loop Files” στο αρχείο “clinical\_trial\_II\_pdfs”.

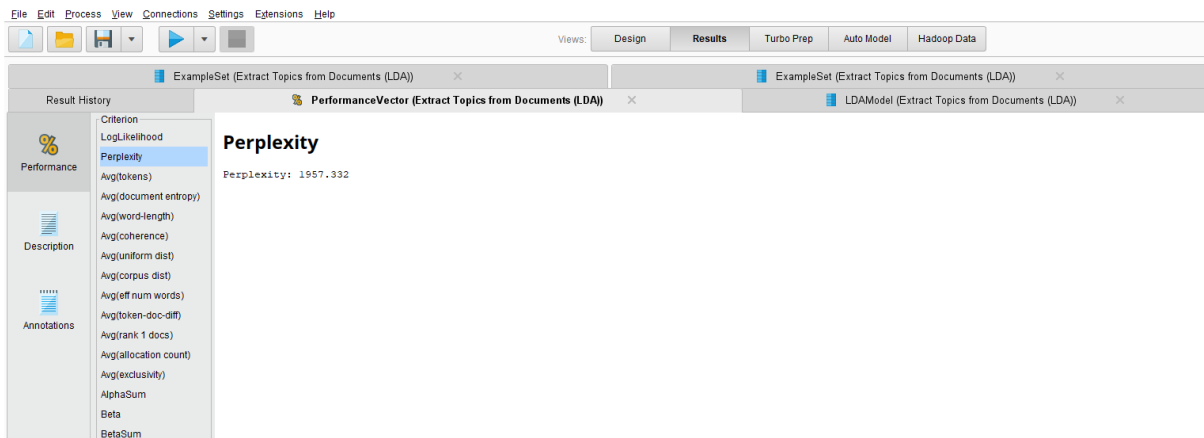


Εικόνα 4.87: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων της ομάδας Clinical Trial II.

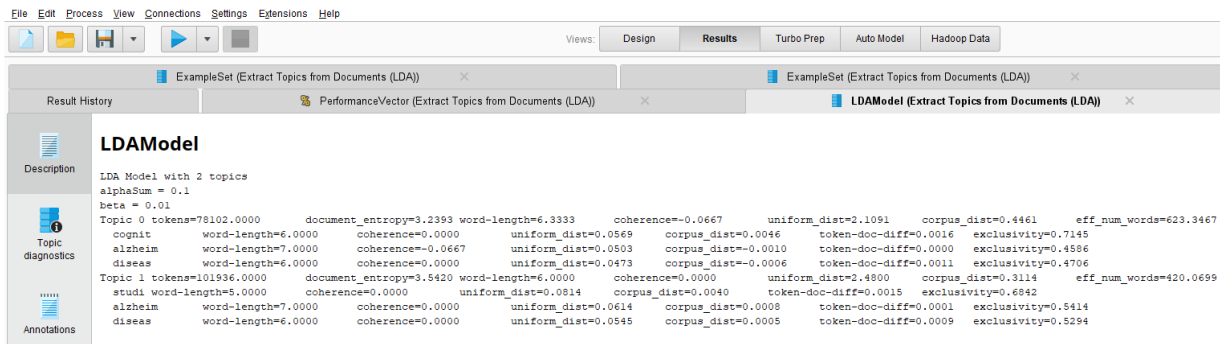
Αποτελέσματα για αριθμό θεμάτων  $k = 2$ :



Εικόνα 4.88: Η τιμή Loglikelihood για  $k = 2$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4.89: Η τιμή Perplexity για  $k = 2$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4.90: Το μοντέλο LDA για  $k = 2$  θέματα, Clinical Trial II – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Row No.	topicId	word	weight
1	0	cognit	867
2	0	alzheimer	783
3	0	diseas	744
4	1	studi	1529
5	1	alzheimer	1206
6	1	diseas	1092

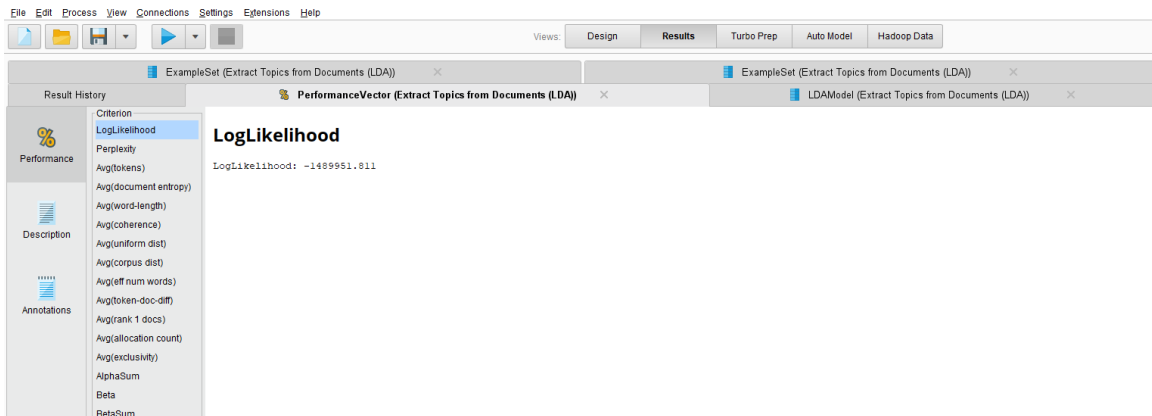
Εικόνα 4.91: : Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 2$  θέματα, Clinical Trial II – πλήρη κείμενα.

Row No.	documentId	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_1	0.031	0.969	RESEARCH ...
2	1	Topic_1	0.307	0.693	www.thelan...
3	2	Topic_1	0.044	0.956	Dhadda et al...
4	3	Topic_1	0.096	0.904	Received: 23...
5	4	Topic_0	0.755	0.245	Yulug et al. Tr...
6	5	Topic_0	0.649	0.351	Riluzole, a gl...
7	6	Topic_1	0.372	0.628	RESEARCH ...
8	7	Topic_1	0.083	0.917	CPT Pharma...
9	8	Topic_1	0.091	0.909	Safety and Eff...
10	9	Topic_1	0.498	0.502	2010   Clin ...
11	10	Topic_1	0.151	0.849	https://doi.org...
12	11	Topic_0	0.561	0.439	https://doi.org...
13	12	Topic_0	0.668	0.332	Intranasal Ins...
14	13	Topic_1	0.187	0.813	R E S E A R ...
15	14	Topic_1	0.103	0.897	Received: 12...
16	15	Topic_1	0.203	0.797	RESEARCH ...
17	16	Topic_1	0.459	0.541	Journal of Alz...
18	17	Topic_1	0.211	0.789	Parkinsonis...
19	18	Topic_0	0.791	0.209	The MARBLE ...
20	19	Topic_1	0.222	0.778	RESEARCH ...
21	20	Topic_1	0.239	0.761	1
22	21	Topic_0	0.847	0.153	Rationale an...
23	22	Topic_1	0.044	0.956	McDade et al...
24	23	Topic_0	0.809	0.191	Valenzuela et...
25	24	Topic_0	0.761	0.239	STUDY PRO...
26	25	Topic_1	0.474	0.526	
27	26	Topic_1	0.187	0.813	ARTICLE OP...
28	27	Topic_1	0.225	0.775	RESEARCH ...
29	28	Topic_0	0.530	0.470	Vol.(012345...
30	29	Topic_0	0.825	0.175	Retrospective...
31	30	Topic_1	0.118	0.882	Findings of Ef...
32	31	Topic_1	0.237	0.763	ORIGINAL R...
33	32	Topic_0	0.718	0.282	Benfotamine...
34	33	Topic_1	0.401	0.599	Noninvasive ...
35	34	Topic_0	0.707	0.293	22
36	35	Topic_0	0.540	0.460	Effect of Level...
37	36	Topic_1	0.035	0.965	Association o...
38	37	Topic_0	0.566	0.434	A phase II stu...

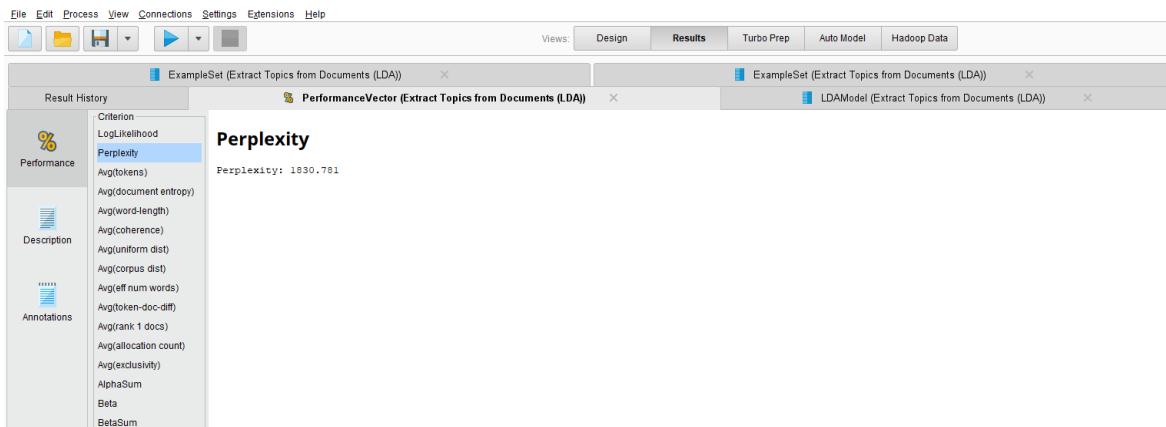
Εικόνα 4.92: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε  $k = 2$  θέματα.



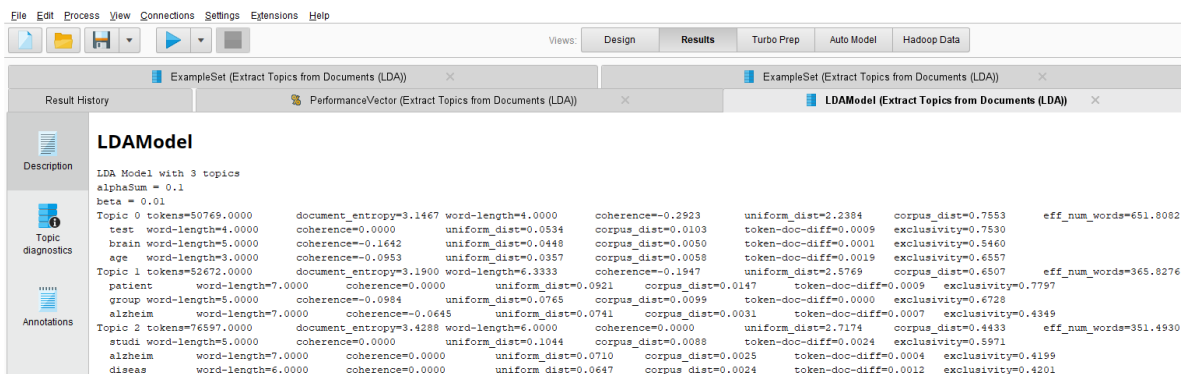
Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :



Εικόνα 4.93: Η τιμή Loglikelihood για  $k = 3$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4.94: Η τιμή Perplexity για  $k = 3$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4. 95: Το μοντέλο LDA για  $k = 3$  θέματα, Clinical Trial II – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

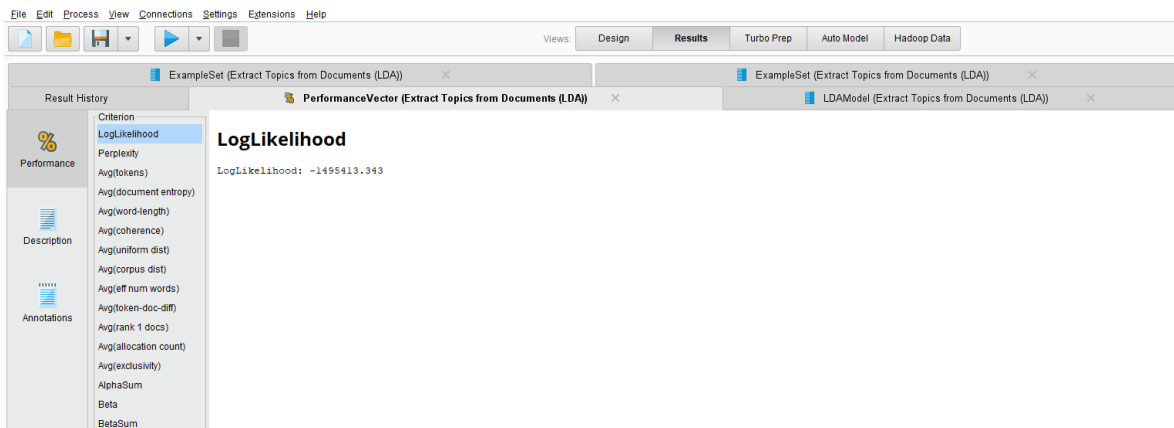
Row No.	topicId	word	weight
1	0	test	535
2	0	brain	462
3	0	age	383
4	1	patient	877
5	1	group	750
6	1	alzhaim	730
7	2	studl	1419
8	2	alzhaim	1024
9	2	diseas	947

Εικόνα 4.96: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για  $k = 3$  θέματα, Clinical Trial II – πλήρη κείμενα.

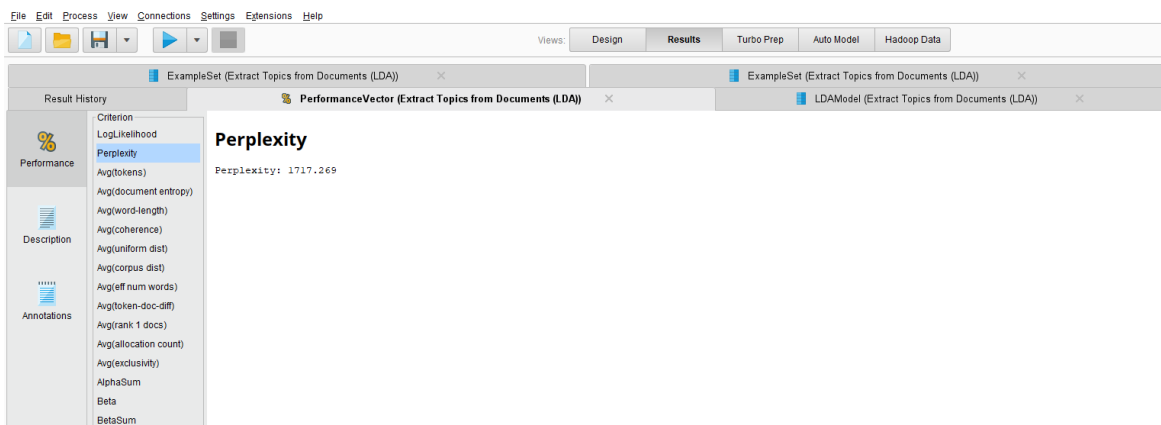
Row No.	document...	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_2	0.051	0.015	0.935	RESEARCH ...
2	1	Topic_2	0.253	0.171	0.576	www.thelanc...
3	2	Topic_2	0.045	0.005	0.950	Dhadda et al...
4	3	Topic_1	0.022	0.626	0.352	Received: 23...
5	4	Topic_0	0.535	0.306	0.159	Yulug et al. Tr...
6	5	Topic_1	0.217	0.515	0.268	Rikuzole, a gl...
7	6	Topic_2	0.146	0.284	0.571	RESEARCH ...
8	7	Topic_2	0.076	0.101	0.823	CPT Pharma...
9	8	Topic_2	0.077	0.034	0.889	Safety and Eff...
10	9	Topic_2	0.346	0.069	0.585	2010   Clin...
11	10	Topic_2	0.086	0.117	0.797	https://doi.org...
12	11	Topic_1	0.135	0.588	0.277	https://doi.org...
13	12	Topic_1	0.373	0.392	0.234	Intranasal Ins...
14	13	Topic_1	0.048	0.576	0.376	R E S E A R ...
15	14	Topic_1	0.023	0.751	0.226	Received: 12...
16	15	Topic_2	0.173	0.120	0.707	RESEARCH ...
17	16	Topic_1	0.161	0.488	0.351	Journal of Alz...
18	17	Topic_1	0.090	0.555	0.355	Parkinsons...
19	18	Topic_1	0.175	0.703	0.123	The MARBLE ...
20	19	Topic_2	0.201	0.176	0.623	RESEARCH ...
21	20	Topic_0	0.653	0.045	0.303	1
22	21	Topic_1	0.373	0.569	0.059	Rationale an...
23	22	Topic_2	0.049	0.021	0.930	McDade et al...
24	23	Topic_0	0.705	0.205	0.089	Valenzuela et...
25	24	Topic_0	0.661	0.212	0.127	STUDY PRO...
26	25	Topic_0	0.893	0.065	0.042	
27	26	Topic_1	0.050	0.760	0.190	ARTICLE OP...
28	27	Topic_2	0.042	0.279	0.679	RESEARCH ...
29	28	Topic_0	0.485	0.136	0.378	Vol.:012345...
30	29	Topic_1	0.136	0.448	0.416	Retrospective...
31	30	Topic_2	0.080	0.028	0.892	Findings of Ef...
32	31	Topic_0	0.814	0.037	0.150	ORIGINAL R...
33	32	Topic_1	0.128	0.697	0.174	Benfotamine...
34	33	Topic_0	0.782	0.071	0.147	Noninvasive ...
35	34	Topic_0	0.625	0.081	0.294	22
36	35	Topic_2	0.165	0.064	0.772	Effect of Lev...
37	36	Topic_2	0.067	0.012	0.920	Association o...
38	37	Topic_2	0.326	0.207	0.467	A phase II stu...

Εικόνα 4.97: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε  $k = 3$  θέματα.

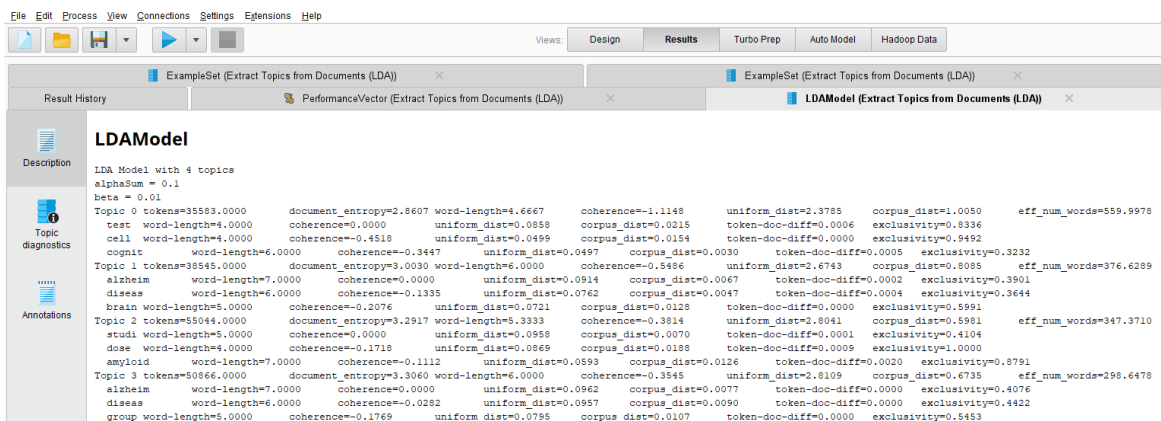
Αποτελέσματα για αριθμό κειμένων  $k = 4$ :



Εικόνα 4. 98: Η τιμή Loglikelihood για  $k = 4$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4.99: Η τιμή Perplexity για  $k = 4$  θέματα, Clinical Trial II – πλήρη κείμενα.



Εικόνα 4.100: Το μοντέλο LDA για  $k = 4$  θέματα, Clinical Trial II – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Row No.	topicid	word	weight
1	0	test	558
2	0	cell	354
3	0	cognit	353
4	1	alzheim	638
5	1	diseas	547
6	1	brain	522
7	2	studi	948
8	2	dose	873
9	2	amyloid	633
10	3	alzheim	879
11	3	diseas	875
12	3	group	748

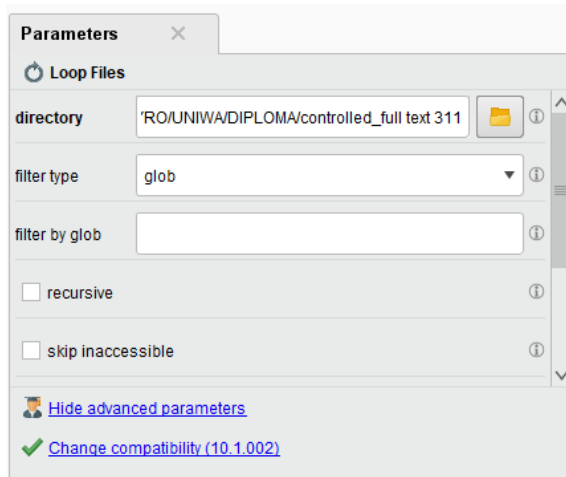
Εικόνα 4.101: Οι 3 πιο συχνές λέξεις κάθε θέματος και τα βάρη τους για k = 4 θέματα, Clinical Trial II – πλήρη κείμενα.

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_2	0.014	0.022	0.853	0.111	RESEARCH ...
2	1	Topic_1	0.059	0.531	0.094	0.315	www.thelanc...
3	2	Topic_2	0.016	0.000	0.918	0.066	Dhadda et al...
4	3	Topic_3	0.015	0.023	0.372	0.590	Received: 23...
5	4	Topic_0	0.548	0.146	0.176	0.130	Yulug et al. Tr...
6	5	Topic_1	0.269	0.361	0.175	0.195	Riluzole, a gl...
7	6	Topic_3	0.113	0.097	0.256	0.534	RESEARCH ...
8	7	Topic_2	0.043	0.039	0.899	0.019	CPT Pharma...
9	8	Topic_3	0.015	0.043	0.355	0.587	Safety and Eff...
10	9	Topic_2	0.194	0.018	0.589	0.099	2010   Clin ...
11	10	Topic_2	0.026	0.056	0.510	0.409	https://doi.org...
12	11	Topic_3	0.051	0.134	0.148	0.666	https://doi.org...
13	12	Topic_1	0.036	0.739	0.101	0.124	Intranasal Ins...
14	13	Topic_3	0.033	0.036	0.225	0.706	R E S E A R ...
15	14	Topic_3	0.020	0.011	0.305	0.665	Received: 12...
16	15	Topic_3	0.113	0.020	0.431	0.435	RESEARCH ...
17	16	Topic_2	0.178	0.171	0.622	0.028	Journal of Alz...
18	17	Topic_3	0.132	0.034	0.246	0.587	Parkinsons...
19	18	Topic_1	0.322	0.403	0.050	0.225	The MARBLE ...
20	19	Topic_3	0.093	0.052	0.212	0.643	RESEARCH ...
21	20	Topic_1	0.037	0.765	0.157	0.041	1
22	21	Topic_1	0.198	0.629	0.017	0.156	Rationale an...
23	22	Topic_2	0.012	0.014	0.906	0.068	McDade et al...
24	23	Topic_0	0.740	0.074	0.104	0.082	Valenzuela et...
25	24	Topic_0	0.450	0.289	0.045	0.216	STUDY PRO...
26	25	Topic_0	0.868	0.007	0.022	0.103	
27	26	Topic_2	0.048	0.035	0.604	0.313	ARTICLE OP...
28	27	Topic_2	0.034	0.007	0.915	0.044	RESEARCH ...
29	28	Topic_2	0.281	0.280	0.282	0.157	Vol. :012345...
30	29	Topic_0	0.478	0.327	0.061	0.134	Retrospective...
31	30	Topic_3	0.031	0.013	0.253	0.703	Findings of EF...
32	31	Topic_1	0.091	0.775	0.102	0.032	ORIGINAL R...
33	32	Topic_1	0.110	0.603	0.103	0.184	Benfotamine...
34	33	Topic_1	0.133	0.718	0.106	0.043	Noninvasiv...
35	34	Topic_1	0.247	0.455	0.120	0.179	22
36	35	Topic_3	0.039	0.028	0.074	0.860	Effect of Lev...
37	36	Topic_2	0.004	0.053	0.744	0.199	Association o...
38	37	Topic_3	0.228	0.184	0.209	0.380	A phase II stu...

Εικόνα 4.102: Η κατηγοριοποίηση των πλήρων κειμένων Clinical Trial II σε k = 4 θέματα.

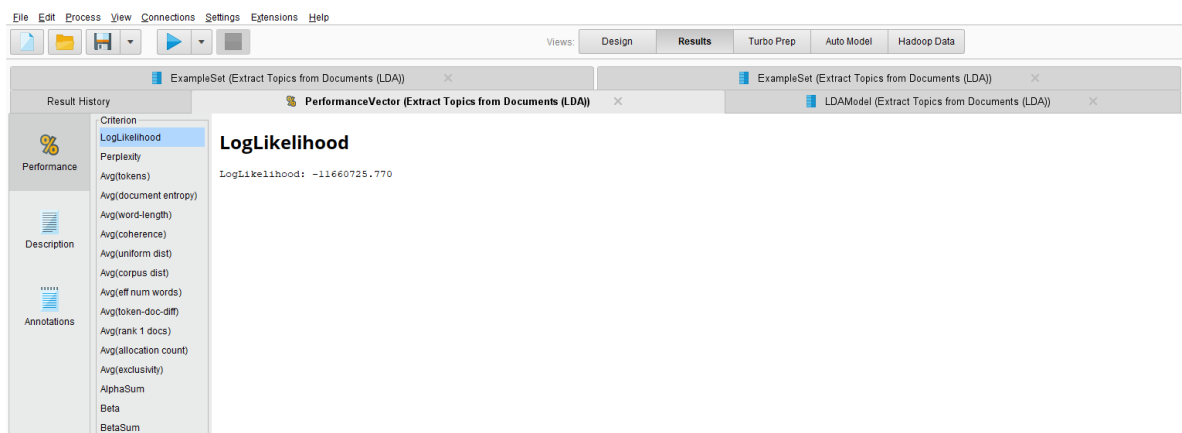
### 4.4.3 Ανάλυση και αποτελέσματα πλήρους κειμένου της ομάδας **Controlled Clinical Trial**

Το directory του χειριστή “Loop Files” αλλάζει σε “controlled\_full text 311” ώστε να εισαχθούν τα κείμενα Controlled Clinical Trial στο πρόγραμμα για να αναλυθούν.

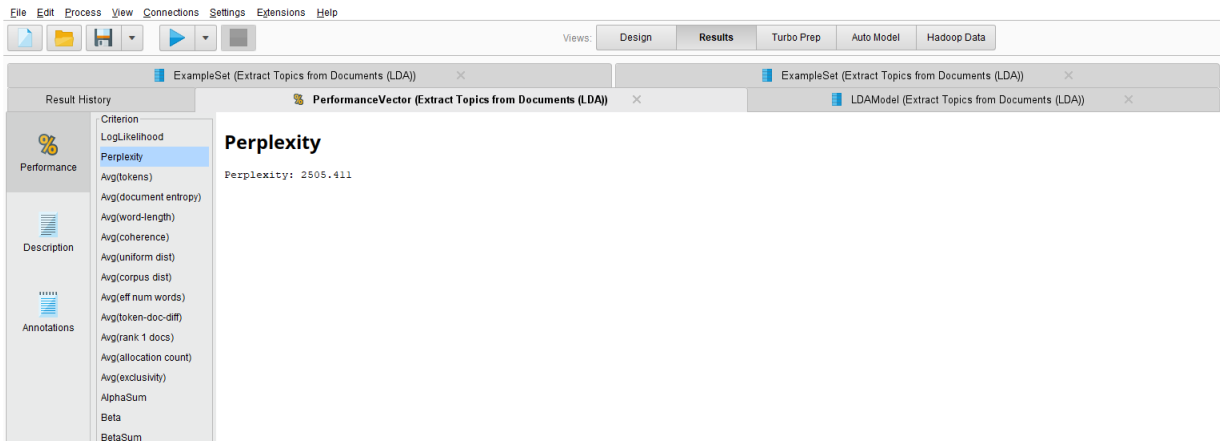


Εικόνα 4.103: Αλλαγή directory για εύρεση των πλήρων κειμένων των άρθρων της ομάδας Controlled Clinical Trial.

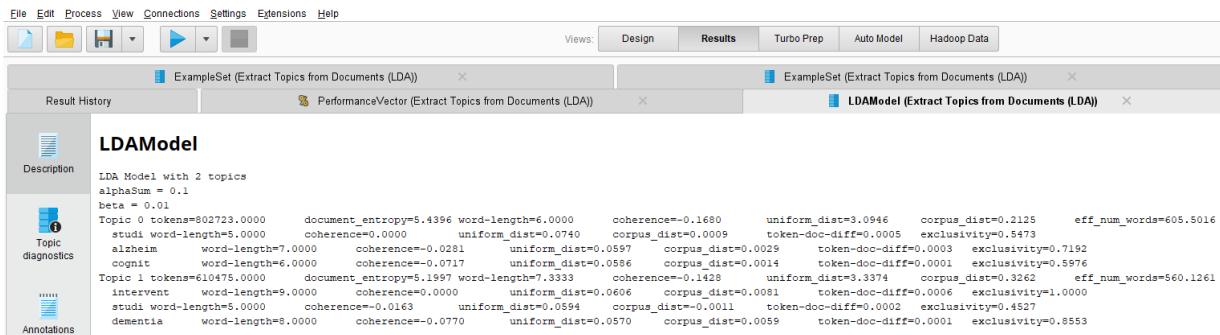
Αποτελέσματα για αριθμό θεμάτων  $k = 2$ :



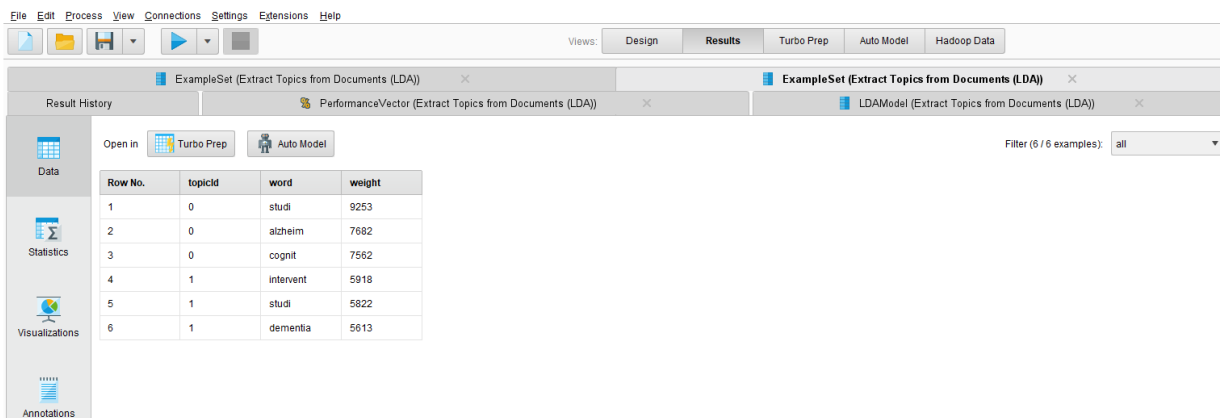
Εικόνα 4.104: Η τιμή Loglikelihood για  $k = 2$  θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.105: Η τιμή Perplexity για k = 2 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.106: Το μοντέλο LDA για k = 2 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.107: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για k = 2 θέματα, Controlled Clinical Trial – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

Result History ExampleSet (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA))

Open In Turbo Prep Auto Model Filter (307 / 307 examples): all

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	text
1	0	Topic_0	0.947	0.053	Alzheimer's
2	1	Topic_1	0.068	0.932	Research Arti...
3	2	Topic_1	0.193	0.807	The relative c...
4	3	Topic_0	0.828	0.172	780
5	4	Topic_0	0.771	0.229	Citation: Yu, X...
6	5	Topic_0	0.830	0.170	Intranasal Ins...
7	6	Topic_1	0.481	0.519	Journal of Alz...
8	7	Topic_1	0.361	0.639	233
9	8	Topic_0	0.793	0.207	Journal of Alz...
10	9	Topic_1	0.032	0.968	Original Paper
11	10	Topic_0	0.679	0.321	1Kondo T, et ...
12	11	Topic_0	0.944	0.056	Lerner et al. A...
13	12	Topic_0	0.930	0.070	Received: 7 ...
14	13	Topic_0	0.934	0.066	Nutrición
15	14	Topic_0	0.762	0.238	RESEAR ...
16	15	Topic_0	0.909	0.091	This article is...
17	16	Topic_0	0.797	0.203	Wunderlich et...
18	17	Topic_0	0.704	0.296	lable at Scien...
19	18	Topic_0	0.942	0.058	Journal of Alz...
20	19	Topic_1	0.068	0.932	Citation: Mufl...
21	20	Topic_1	0.064	0.936	RESEARCH ...
22	21	Topic_0	0.971	0.029	Journal of Alz...
23	22	Topic_0	0.660	0.340	Quantifying th...
24	23	Topic_0	0.928	0.072	Amyloid-Rela...
25	24	Topic_0	0.971	0.029	RESEARCH ...
26	25	Topic_0	0.863	0.137	Received: 7 J...
27	26	Topic_0	0.778	0.222	Received: 27 ...
28	27	Topic_1	0.011	0.989	Attention Con...
29	28	Topic_1	0.066	0.934	TYPE Origina...
30	29	Topic_0	0.883	0.117	OR I G I N A L...
31	30	Topic_1	0.150	0.850	Current Topic...
32	31	Topic_0	0.813	0.187	Vol.: (012345...
33	32	Topic_0	0.738	0.262	lable at Scien...
34	33	Topic_1	0.213	0.787	Baseline prev...
35	34	Topic_0	0.518	0.482	Am J of Geria...
36	35	Topic_0	0.929	0.071	Journal of Alz...
37	36	Topic_1	0.017	0.983	RESEARCH ...
38	37	Topic_0	0.978	0.022	635
39	38	Topic_0	0.574	0.426	Rationale, st...
40	39	Topic_0	0.892	0.108	Received: 12...
41	40	Topic_1	0.023	0.977	A Dyadic Appr...
42	41	Topic_0	0.896	0.104	J. Integr. Neu...
43	42	Topic_0	0.693	0.307	2169
44	43	Topic_0	0.923	0.077	Journal of Alz...
45	44	Topic_1	0.165	0.835	Public opinio...
46	45	Topic_1	0.409	0.591	Mediterranea...
47	46	Topic_0	0.577	0.423	RESEARCH ...
48	47	Topic_1	0.242	0.758	Research Arti...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

49	48	Topic_0	0.720	0.280	1
50	49	Topic_1	0.217	0.783	Available onli...
51	50	Topic_0	0.762	0.238	1
52	51	Topic_1	0.000	1.000	It's Like a Cy...
53	52	Topic_1	0.133	0.867	Measuring ap...
54	53	Topic_1	0.320	0.680	646
55	54	Topic_1	0.002	0.998	Implications ...
56	55	Topic_1	0.022	0.978	Gaugler et al...
57	56	Topic_1	0.088	0.912	International ...
58	57	Topic_0	0.945	0.055	Journal of Alz...
59	58	Topic_1	0.000	1.000	The design a...
60	59	Topic_0	0.883	0.117	Whole-Brain ...
61	60	Topic_0	0.949	0.051	Vol. (012345...
62	61	Topic_0	0.880	0.120	RESEARCH ...
63	62	Topic_1	0.274	0.726	CLINICAL ST...
64	63	Topic_0	0.934	0.066	RESEARCH ...
65	64	Topic_0	0.772	0.228	Optimizing qu...
66	65	Topic_0	0.943	0.057	Nature Medici...
67	66	Topic_0	0.970	0.030	McDade et al...
68	67	Topic_1	0.000	1.000	One Call Mak...
69	68	Topic_0	0.796	0.204	Citation: LI. B...
70	69	Topic_0	0.539	0.461	Potashman e...
71	70	Topic_0	0.778	0.222	ARTICLE OP...
72	71	Topic_0	0.923	0.077	
73	72	Topic_0	0.710	0.290	OR I G I N A L...
74	73	Topic_0	0.975	0.025	Journal of Alz...
75	74	Topic_0	0.583	0.417	Parkinsons...
76	75	Topic_1	0.041	0.959	Bentley et al...
77	76	Topic_1	0.000	1.000	Implementati...
78	77	Topic_0	0.799	0.201	Received: 21...
79	78	Topic_0	0.887	0.113	REGULAR A...
80	79	Topic_1	0.026	0.974	Cheung et al...
81	80	Topic_0	0.518	0.482	717www.eym...
82	81	Topic_0	0.953	0.047	Journal of Alz...
83	82	Topic_1	0.012	0.988	S H O R T R ...
84	83	Topic_0	0.929	0.071	190
85	84	Topic_0	0.707	0.293	Journal of Alz...
86	85	Topic_0	0.681	0.319	The MARBLE ...
87	86	Topic_1	0.325	0.675	Journal of Alz...
88	87	Topic_0	0.858	0.142	Citation: Chol...
89	88	Topic_0	0.772	0.228	Received: 9 J...
90	89	Topic_0	0.815	0.185	RESEARCH ...
91	90	Topic_0	0.915	0.085	Diaz-Galvan ...
92	91	Topic_1	0.033	0.967	Lessons fro...
93	92	Topic_0	0.851	0.149	RESEARCH ...
94	93	Topic_0	0.941	0.059	Biomedicine ...
95	94	Topic_1	0.220	0.780	Journal of Alz...
96	95	Topic_0	0.784	0.216	TaggedEndG...
97	96	Topic_1	0.265	0.735	RESEARCH ...
98	97	Topic_0	0.669	0.331	Prospective a...
99	98	Topic_1	0.443	0.557	Effect of Meth...
100	99	Topic_0	0.799	0.201	RESEARCH ...
101	100	Topic_0	0.549	0.451	[18]FFDNP ...
102	101	Topic_0	0.897	0.103	Received: De...
103	102	Topic_1	0.114	0.886	©Journal of S...
104	103	Topic_0	0.925	0.075	Efficacy and ...
105	104	Topic_1	0.378	0.622	Received: 25 ...
106	105	Topic_0	0.912	0.088	Kile et al. BM...
107	106	Topic_1	0.279	0.721	Rationale an...
108	107	Topic_1	0.377	0.623	Mnemonic str...
109	108	Topic_1	0.061	0.939	Vol. (012345...
110	109	Topic_0	0.703	0.297	32
111	110	Topic_0	0.707	0.293	Benfotamine...
112	111	Topic_0	0.905	0.095	
113	112	Topic_0	0.624	0.376	Estradiol trea...
114	113	Topic_0	0.958	0.042	OR I G I N A L...
115	114	Topic_1	0.399	0.601	A Randomize...
116	115	Topic_0	0.942	0.058	OR I G I N A L...
117	116	Topic_0	0.821	0.179	Vol. (012345...
118	117	Topic_1	0.135	0.865	Journal of Alz...
119	118	Topic_1	0.067	0.933	International ...
120	119	Topic_1	0.064	0.936	Using Smart...



«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

121	120	Topic_1	0.489	0.511	Preliminary e...
122	121	Topic_0	0.839	0.161	RESEARCH ...
123	122	Topic_0	0.942	0.058	RESEARCH ...
124	123	Topic_1	0.082	0.918	REGULAR A...
125	124	Topic_1	0.152	0.848	Nighttime Agi...
126	125	Topic_0	0.983	0.017	RESEARCH ...
127	126	Topic_0	0.824	0.176	ORIGINAL IN...
128	127	Topic_1	0.465	0.535	RESEARCH ...
129	128	Topic_1	0.414	0.586	Few Structura...
130	129	Topic_0	0.957	0.043	Vol. (012345...
131	130	Topic_1	0.476	0.524	STUDY PRO...
132	131	Topic_1	0.001	0.999	The Effect of L...
133	132	Topic_0	0.901	0.099	EBioMedicine...
134	133	Topic_1	0.087	0.913	1017
135	134	Topic_0	0.835	0.165	
136	135	Topic_1	0.185	0.815	Received: 10 ...
137	136	Topic_0	0.871	0.129	RESEARCH ...
138	137	Topic_0	0.703	0.297	Received: 19 ...
139	138	Topic_1	0.105	0.895	1412   ...
140	139	Topic_0	0.846	0.154	ARTICLE OP...
141	140	Topic_1	0.041	0.959	https://doi.org...
142	141	Topic_1	0.231	0.769	Courtois-Am...
143	142	Topic_1	0.005	0.995	Citation: Xu, L...
144	143	Topic_0	0.867	0.133	22
145	144	Topic_1	0.017	0.983	Teles et al. B...
146	145	Topic_0	0.795	0.205	RESEARCH ...
147	146	Topic_0	0.969	0.031	RESEARCH ...
148	147	Topic_0	0.950	0.050	O R I G I N A ...
149	148	Topic_0	0.911	0.089	Clinical Ther...
150	149	Topic_0	0.548	0.452	Shin et al. BM...
151	150	Topic_0	0.824	0.176	Journal of Alz...
152	151	Topic_0	0.949	0.051	RESEARCH ...
153	152	Topic_0	0.652	0.348	Neuropsych...
154	153	Topic_1	0.032	0.968	Partnering a...
155	154	Topic_0	0.965	0.035	Received: 20 ...
156	155	Topic_0	0.896	0.104	
157	156	Topic_0	0.984	0.016	O R I G I N A L...
158	157	Topic_0	0.574	0.426	Received: 28 ...
159	158	Topic_1	0.264	0.736	Whitfield et al.
160	159	Topic_0	0.861	0.139	Journal of Alz...
161	160	Topic_0	0.835	0.165	Journal of Ps...
162	161	Topic_0	0.849	0.151	1
163	162	Topic_1	0.012	0.988	1Morgan- Tri...
164	163	Topic_0	0.825	0.175	Original Inves...
165	164	Topic_1	0.116	0.884	Current Topic...
166	165	Topic_0	0.808	0.192	Effect of Lev...
167	166	Topic_1	0.012	0.988	616
168	167	Topic_0	0.753	0.247	Ca Mg Ratio, ...
169	168	Topic_0	0.855	0.145	10/7/23, 1:36 ...
170	169	Topic_1	0.044	0.956	STUDY PRO...
171	170	Topic_1	0.451	0.549	Study Protoco...
172	171	Topic_0	0.863	0.137	nutrients
173	172	Topic_0	0.500	0.500	Journal of Alz...
174	173	Topic_0	0.800	0.200	1
175	174	Topic_0	0.937	0.063	ORIGINAL R...
176	175	Topic_1	0.145	0.855	RESEARCH ...
177	176	Topic_0	0.876	0.124	JAMDA 23 (2...
178	177	Topic_1	0.143	0.857	Smith et al. B...
179	178	Topic_0	0.879	0.121	Vol. (012345...
180	179	Topic_1	0.010	0.990	Protocol for a...
181	180	Topic_0	0.756	0.244	Association o...
182	181	Topic_1	0.004	0.996	Clare et al. B...
183	182	Topic_1	0.000	1.000	Satisfaction, ...
184	183	Topic_0	0.543	0.457	Combined tre...
185	184	Topic_0	0.745	0.255	Retrospective...
186	185	Topic_1	0.062	0.938	Australas J A...
187	186	Topic_0	0.596	0.404	Citation: Oga...
188	187	Topic_0	0.987	0.013	Association o...
189	188	Topic_0	0.903	0.097	Vol. (123456...
190	189	Topic_0	0.915	0.085	Received: 20 ...
191	190	Topic_1	0.017	0.983	STUDY PRO...
192	191	Topic_0	0.899	0.101	Original Inves...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

193	192	Topic_0	0.951	0.049	Alzheimer Di...
194	193	Topic_1	0.005	0.995	Aging & Ment...
195	194	Topic_0	0.726	0.274	1407
196	195	Topic_1	0.490	0.510	BRIEF COMM...
197	196	Topic_1	0.067	0.913	RESEARCH ...
198	197	Topic_1	0.034	0.966	RESIDENT & ...
199	198	Topic_0	0.943	0.057	A phase II stu...
200	199	Topic_1	0.002	0.998	Effectiveness...
201	200	Topic_0	0.978	0.022	Vol. (012345...
202	201	Topic_0	0.782	0.218	The Journal o...
203	202	Topic_0	0.836	0.164	nutrients
204	203	Topic_1	0.044	0.956	Risks and be...
205	204	Topic_1	0.206	0.794	O R I G I N A ...
206	205	Topic_1	0.153	0.847	RESEARCH ...
207	206	Topic_0	0.921	0.079	Original Inves...
208	207	Topic_1	0.001	0.999	Journals Libr...
209	208	Topic_0	0.953	0.047	Original Inves...
210	209	Topic_0	0.533	0.467	www.thelanc...
211	210	Topic_0	0.748	0.252	Clinical Effect...
212	211	Topic_0	0.816	0.184	Original Inves...
213	212	Topic_1	0.009	0.991	Advance Car...
214	213	Topic_0	0.836	0.164	ARTICLE OP...
215	214	Topic_0	0.575	0.425	Cognitive co...
216	215	Topic_0	0.838	0.162	RESEARCH ...
217	216	Topic_1	0.450	0.550	Citation: Che...
218	217	Topic_0	0.960	0.040	Findings of Ef...
219	218	Topic_0	0.759	0.241	1
220	219	Topic_0	0.582	0.418	Short-term P...
221	220	Topic_1	0.278	0.722	Journal of Alz...
222	221	Topic_0	0.908	0.092	Dhadda et al...
223	222	Topic_1	0.344	0.656	Study Protoco...
224	223	Topic_1	0.098	0.902	The Trial to R...
225	224	Topic_1	0.002	0.998	Effects of a vi...
226	225	Topic_1	0.367	0.633	Study Protoco...
227	226	Topic_1	0.057	0.943	Original Paper
228	227	Topic_0	0.918	0.082	ARTICLE CL...
229	228	Topic_1	0.074	0.926	Original Paper
230	229	Topic_1	0.019	0.981	RESEARCH ...
231	230	Topic_0	0.820	0.180	Levakov, Kapl...
232	231	Topic_0	0.746	0.254	Original Inves...
233	232	Topic_0	0.826	0.174	RESEARCH ...
234	233	Topic_0	0.605	0.395	Minocycline a...
235	234	Topic_0	0.931	0.069	Reactions to ...
236	235	Topic_0	0.874	0.126	Safety, Efficac...
237	236	Topic_1	0.139	0.861	International ...
238	237	Topic_0	0.739	0.261	lable at Scien...
239	238	Topic_0	0.687	0.313	NeuroImage ...
240	239	Topic_0	0.843	0.157	Viña et al. Alz...
241	240	Topic_0	0.914	0.086	Received: 23...
242	241	Topic_1	0.386	0.614	STUDY PRO...
243	242	Topic_0	0.955	0.045	Yulug et al. Tr...
244	243	Topic_0	0.735	0.265	A trial of gant...
245	244	Topic_0	0.898	0.102	nutrients
246	245	Topic_1	0.500	0.500	Journal of Alz...
247	246	Topic_0	0.957	0.043	Riluzole, a gl...
248	247	Topic_1	0.177	0.823	Articles
249	248	Topic_1	0.265	0.735	Journal of Alz...
250	249	Topic_0	0.749	0.251	lable at Scien...
251	250	Topic_0	0.581	0.419	Contents list...
252	251	Topic_0	0.902	0.098	Dubois et al. ...
253	252	Topic_1	0.212	0.788	Clinical Trial ...
254	253	Topic_0	0.825	0.175	lable at Scien...
255	254	Topic_0	0.923	0.077	ArticleFrom st...
256	255	Topic_0	0.892	0.108	RESEARCH ...
257	256	Topic_0	0.910	0.090	1
258	257	Topic_0	0.915	0.085	lable at Scien...
259	258	Topic_1	0.335	0.665	Low dose libh...
260	259	Topic_0	0.768	0.232	Sleep depriva...
261	260	Topic_1	0.199	0.801	Properties of ...
262	261	Topic_0	0.984	0.016	BACE inhibiti...
263	262	Topic_1	0.493	0.507	Received: 24 ...
264	263	Topic_0	0.602	0.398	Vol. (012345...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

265	264	Topic_0	0.926	0.074	Research Art...
266	265	Topic_0	0.670	0.330	RESEARCH ...
267	266	Topic_0	0.503	0.497	NULL HYPOT...
268	267	Topic_0	0.986	0.014	Association o...
269	268	Topic_1	0.000	1.000	Transitioning ...
270	269	Topic_0	0.983	0.037	1
271	270	Topic_0	0.927	0.073	Journal of Alz...
272	271	Topic_0	0.866	0.134	Effect of APO...
273	272	Topic_0	0.751	0.249	RESEARCH ...
274	273	Topic_0	0.934	0.066	Safety and Eff...
275	274	Topic_0	0.544	0.456	RESEARCH ...
276	275	Topic_0	0.888	0.112	https://doi.org...
277	276	Topic_0	0.929	0.071	RESEARCH ...
278	277	Topic_1	0.449	0.551	Received: 10...
279	278	Topic_1	0.383	0.617	ORIGINAL AR...
280	279	Topic_0	0.967	0.033	2010   Clin ...
281	280	Topic_0	0.844	0.156	lable at Scien...
282	281	Topic_1	0.110	0.890	Mindfulness i...
283	282	Topic_0	0.921	0.079	Citation: Kad...
284	283	Topic_1	0.428	0.572	ARTICLE
285	284	Topic_1	0.252	0.748	Original Inves...
286	285	Topic_1	0.139	0.861	The Preventin...
287	286	Topic_0	0.774	0.226	Frontiers in E...
288	287	Topic_1	0.376	0.624	10   Ne...
289	288	Topic_0	0.951	0.049	Journal of Alz...
290	289	Topic_1	0.486	0.514	Original Inves...
291	290	Topic_0	0.532	0.468	Brain Perfus...
292	291	Topic_0	0.917	0.083	Evaluating th...
293	292	Topic_1	0.141	0.859	Comparison ...
294	293	Topic_0	0.944	0.056	Frontiers in L...
295	294	Topic_0	0.768	0.232	Citation: Shi...
296	295	Topic_1	0.139	0.861	RESEARCH ...
297	296	Topic_1	0.379	0.621	30
298	297	Topic_0	0.774	0.226	Journal of Alz...
299	298	Topic_0	0.823	0.177	RESEARCH ...
300	299	Topic_0	0.908	0.092	Association o...
301	300	Topic_1	0.100	0.900	547
302	301	Topic_0	0.933	0.067	https://doi.org...
303	302	Topic_1	0.254	0.746	A randomized...
304	303	Topic_1	0.297	0.703	See correspo...
305	304	Topic_1	0.444	0.556	Intensive vers...
306	305	Topic_1	0.215	0.785	International ...
307	306	Topic_0	0.857	0.143	RESEARCH ...

Εικόνα 4.108: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε  $k = 2$  θέματα.

Αποτελέσματα για αριθμό θεμάτων  $k = 3$ :

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

ExampleSet (Extract Topics from Documents (LDA))

Result History PerformanceVector (Extract Topics from Documents (LDA)) LDAModel (Extract Topics from Documents (LDA))

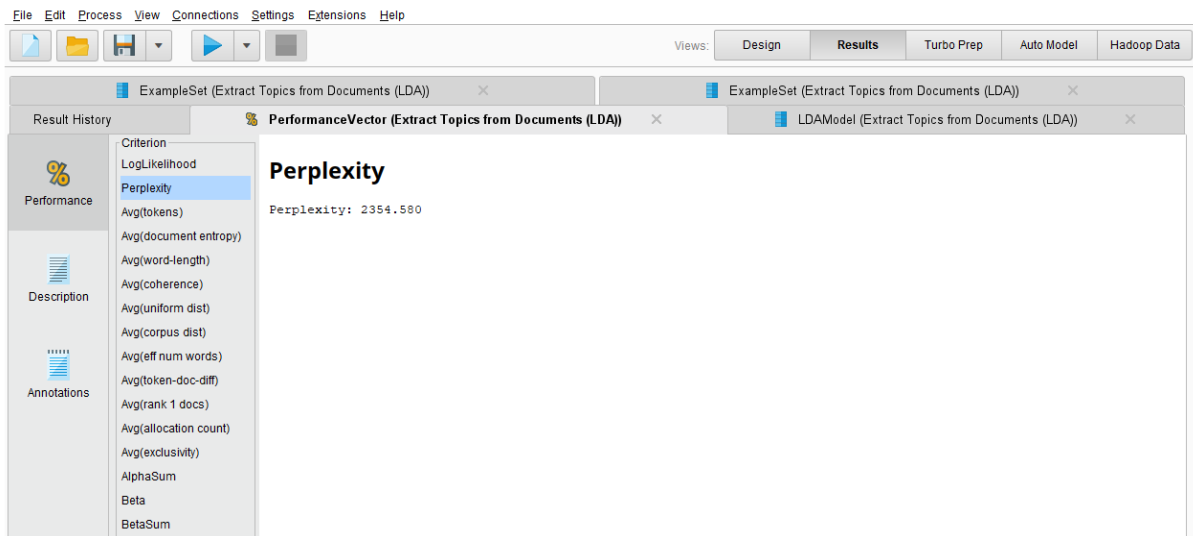
Criterion

- LogLikelihood
- Perplexity
- Avg(tokens)
- Avg(document entropy)
- Avg(word-length)
- Avg(coherence)
- Avg(uniform dist)
- Avg(corpus dist)
- Avg(eff num words)
- Avg(token-doc-diff)
- Avg(rank 1 docs)
- Avg(allocation count)
- Avg(exclusivity)
- AlphaSum
- Beta
- BetaSum

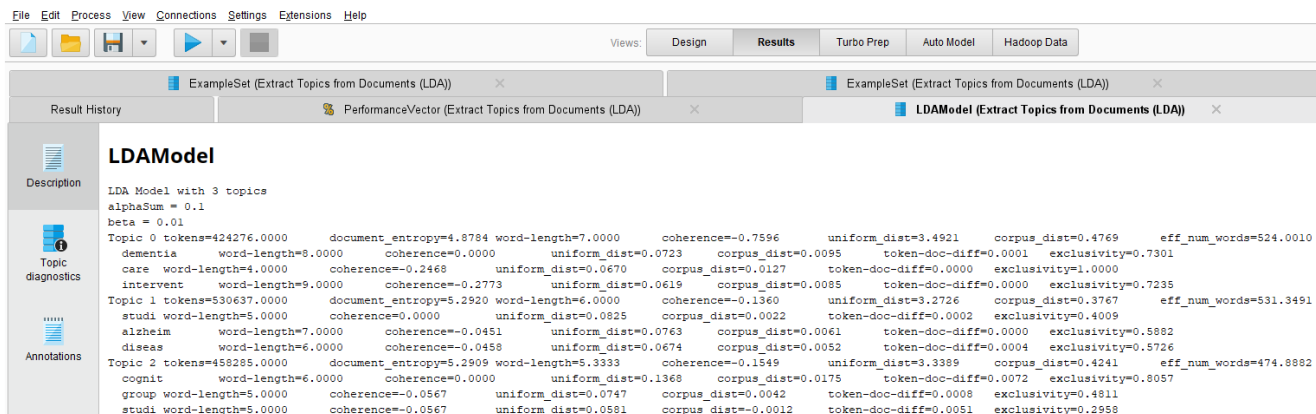
**LogLikelihood**

LogLikelihood: -11796788.272

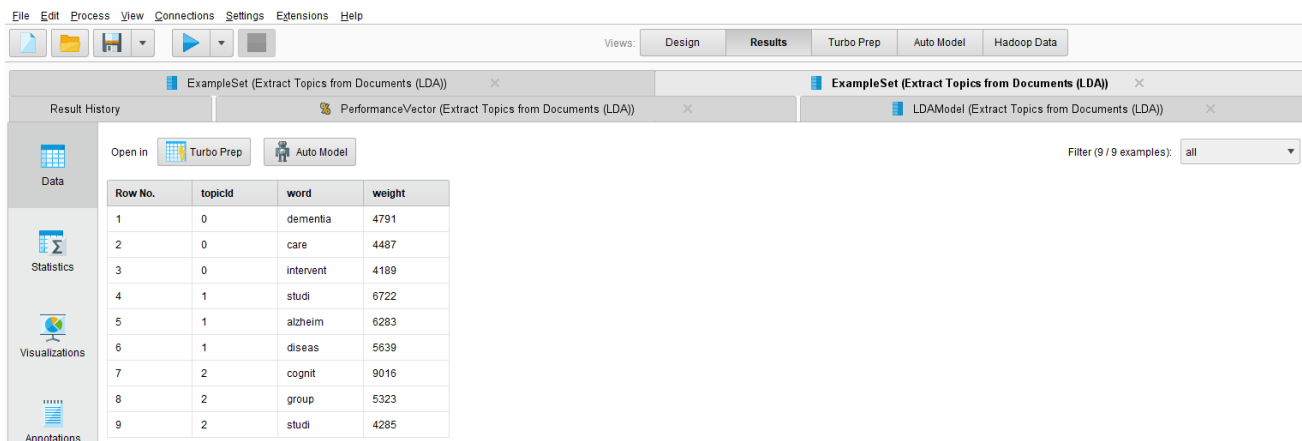
Εικόνα 4.109: Η τιμή Loglikelihood για  $k = 3$  θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.110: Η τιμή Perplexity για k = 3 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.111: Το μοντέλο LDA για k = 3 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.112: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για k = 3 θέματα, Controlled Clinical Trial – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

ExampleSet (Extract Topics from Documents (LDA))

Result History PerformanceVector (Extract Topics from Documents (LDA)) LDAModel (Extract Topics from Documents (LDA))

Open in Turbo Prep Auto Model Filter (307 / 307 examples): all

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	text
1	0	Topic_1	0.049	0.946	0.005	Alzheimer's
2	1	Topic_0	0.610	0.012	0.379	Research Arti...
3	2	Topic_0	0.629	0.143	0.227	The relative c...
4	3	Topic_1	0.150	0.826	0.024	780
5	4	Topic_1	0.149	0.526	0.325	Citation: Yu, X...
6	5	Topic_1	0.008	0.502	0.490	Intranasal Ins...
7	6	Topic_2	0.302	0.250	0.449	Journal of Alz...
8	7	Topic_0	0.437	0.243	0.321	233
9	8	Topic_1	0.121	0.645	0.234	Journal of Alz...
10	9	Topic_0	0.846	0.014	0.140	Original Paper
11	10	Topic_1	0.251	0.713	0.036	Ikondo T, et...
12	11	Topic_1	0.038	0.865	0.096	Lerner et al. A...
13	12	Topic_2	0.022	0.447	0.531	Received: 7 ...
14	13	Topic_1	0.061	0.764	0.175	Nutrición
15	14	Topic_1	0.145	0.732	0.123	R E S E A R ...
16	15	Topic_1	0.044	0.763	0.193	This article is...
17	16	Topic_1	0.175	0.791	0.033	Wunderlich et...
18	17	Topic_2	0.060	0.175	0.766	lable at Scien...
19	18	Topic_1	0.030	0.729	0.241	Journal of Alz...
20	19	Topic_0	0.731	0.033	0.236	Citation: Muñ...
21	20	Topic_0	0.853	0.072	0.075	RESEARCH ...
22	21	Topic_1	0.012	0.835	0.153	Journal of Alz...
23	22	Topic_1	0.255	0.618	0.127	Quantifying th...
24	23	Topic_1	0.060	0.935	0.005	Amyloid-Rela...
25	24	Topic_1	0.030	0.946	0.024	RESEARCH ...
26	25	Topic_1	0.089	0.759	0.152	Received: 7 J...
27	26	Topic_1	0.108	0.618	0.274	Received: 27 ...
28	27	Topic_0	0.903	0.004	0.093	Attention Con...
29	28	Topic_0	0.772	0.017	0.211	TYPE Origina...
30	29	Topic_1	0.063	0.596	0.341	OR I G I N A L
31	30	Topic_0	0.696	0.071	0.233	Current Topic...
32	31	Topic_2	0.052	0.221	0.727	Vol.: (012345...
33	32	Topic_2	0.030	0.116	0.854	lable at Scien...
34	33	Topic_2	0.375	0.082	0.543	Baseline prev...
35	34	Topic_1	0.281	0.708	0.011	Am J of Geria...
36	35	Topic_1	0.028	0.551	0.421	Journal of Alz...
37	36	Topic_0	0.747	0.000	0.253	RESEARCH ...
38	37	Topic_1	0.022	0.949	0.029	635
39	38	Topic_1	0.279	0.444	0.277	Rationale, st...
40	39	Topic_1	0.076	0.785	0.139	Received: 12...
41	40	Topic_0	0.692	0.002	0.305	A Dyadic Appr...
42	41	Topic_2	0.028	0.395	0.577	J. Integr. Neu...
43	42	Topic_2	0.148	0.155	0.698	2169
44	43	Topic_1	0.091	0.900	0.009	Journal of Alz...
45	44	Topic_0	0.768	0.222	0.010	Public opinio...
46	45	Topic_2	0.231	0.065	0.704	Mediterranea...
47	46	Topic_2	0.122	0.125	0.753	RESEARCH ...
48	47	Topic_0	0.627	0.278	0.095	Research Arti...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

49	48	Topic_0	0.468	0.260	0.272	1
50	49	Topic_2	0.264	0.075	0.661	Available onli...
51	50	Topic_2	0.051	0.049	0.900	1
52	51	Topic_0	0.966	0.001	0.032	It's Like a Cy...
53	52	Topic_0	0.696	0.198	0.105	Measuring ap...
54	53	Topic_0	0.468	0.243	0.289	646
55	54	Topic_0	0.983	0.000	0.017	Implications ...
56	55	Topic_0	0.942	0.017	0.041	Gaugler et al...
57	56	Topic_0	0.671	0.006	0.323	International ...
58	57	Topic_1	0.005	0.592	0.403	Journal of Alz...
59	58	Topic_0	0.984	0.000	0.016	The design a...
60	59	Topic_2	0.063	0.451	0.486	Whole-Brain ...
61	60	Topic_1	0.022	0.742	0.236	Vol.:012345...
62	61	Topic_1	0.039	0.636	0.325	RESEARCH ...
63	62	Topic_0	0.464	0.121	0.414	CLINICAL ST...
64	63	Topic_1	0.051	0.841	0.108	RESEARCH ...
65	64	Topic_2	0.034	0.325	0.641	Optimizing qu...
66	65	Topic_1	0.058	0.926	0.016	Nature Medici...
67	66	Topic_1	0.033	0.967	0.000	McDade et al...
68	67	Topic_0	0.984	0.000	0.016	One Call Mak...
69	68	Topic_2	0.086	0.196	0.718	Citation: LI, B.
70	69	Topic_1	0.336	0.467	0.198	Potashman e...
71	70	Topic_2	0.052	0.212	0.737	ARTICLE OP...
72	71	Topic_1	0.028	0.584	0.388	
73	72	Topic_2	0.120	0.299	0.581	OR I G I N A L...
74	73	Topic_1	0.012	0.827	0.161	Journal of Alz...
75	74	Topic_1	0.220	0.544	0.236	Parkinsonis...
76	75	Topic_0	0.900	0.034	0.065	Bentley et al. ...
77	76	Topic_0	0.966	0.000	0.034	Implementab...
78	77	Topic_1	0.053	0.492	0.455	Received: 21...
79	78	Topic_1	0.070	0.769	0.161	REGULAR A...
80	79	Topic_0	0.903	0.010	0.087	Cheung et al...
81	80	Topic_2	0.082	0.187	0.731	717www.eym...
82	81	Topic_1	0.012	0.525	0.463	Journal of Alz...
83	82	Topic_0	0.843	0.000	0.157	S H O R T R ...
84	83	Topic_1	0.026	0.709	0.265	190
85	84	Topic_2	0.040	0.109	0.851	Journal of Alz...
86	85	Topic_1	0.319	0.412	0.270	The MARBLE ...
87	86	Topic_2	0.179	0.116	0.705	Journal of Alz...
88	87	Topic_2	0.020	0.300	0.680	Citation: Cho...
89	88	Topic_1	0.189	0.683	0.128	Received: 9 J...
90	89	Topic_1	0.096	0.733	0.171	RESEARCH ...
91	90	Topic_1	0.050	0.716	0.234	Diaz-Galvan ...
92	91	Topic_0	0.946	0.032	0.022	Lessons fro...
93	92	Topic_1	0.062	0.651	0.287	RESEARCH ...
94	93	Topic_1	0.057	0.857	0.086	Biomedicine ...
95	94	Topic_0	0.560	0.112	0.328	Journal of Alz...
96	95	Topic_2	0.097	0.180	0.722	TaggedEndG...
97	96	Topic_2	0.420	0.025	0.555	RESEARCH ...
98	97	Topic_1	0.271	0.551	0.178	Prospective a...
99	98	Topic_1	0.351	0.602	0.047	Effect of Meth...
100	99	Topic_0	0.516	0.316	0.168	RESEARCH ...
101	100	Topic_2	0.055	0.319	0.626	[18]FDDNP ...
102	101	Topic_1	0.022	0.570	0.408	Received: De...
103	102	Topic_2	0.166	0.011	0.823	@Journal of S...
104	103	Topic_1	0.057	0.930	0.013	Efficacy and ...
105	104	Topic_2	0.400	0.146	0.455	Received: 25 ...
106	105	Topic_1	0.073	0.809	0.119	Kile et al. BM...
107	106	Topic_2	0.297	0.056	0.647	Rationale an...
108	107	Topic_2	0.066	0.001	0.933	Mnemonic str...
109	108	Topic_0	0.853	0.032	0.115	Vol.:012345...
110	109	Topic_2	0.119	0.201	0.680	32
111	110	Topic_2	0.083	0.449	0.468	Benfotamine...
112	111	Topic_1	0.080	0.888	0.032	
113	112	Topic_2	0.031	0.169	0.800	Estradiol trea...
114	113	Topic_1	0.010	0.760	0.230	OR I G I N A L...
115	114	Topic_0	0.436	0.355	0.210	A Randomize...
116	115	Topic_1	0.011	0.745	0.244	OR I G I N A L...
117	116	Topic_2	0.014	0.063	0.923	Vol.:012345...
118	117	Topic_0	0.671	0.272	0.056	Journal of Alz...
119	118	Topic_0	0.719	0.005	0.276	International ...
120	119	Topic_0	0.588	0.004	0.409	Using Smart...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

121	120	Topic_2	0.211	0.281	0.508	Preliminary e...
122	121	Topic_1	0.106	0.733	0.161	RESEARCH ...
123	122	Topic_1	0.058	0.889	0.053	RESEARCH ...
124	123	Topic_0	0.671	0.032	0.297	REGULAR A...
125	124	Topic_0	0.772	0.178	0.050	Nighttime Agl...
126	125	Topic_1	0.013	0.892	0.094	RESEARCH ...
127	126	Topic_1	0.096	0.589	0.315	ORIGINAL IN...
128	127	Topic_2	0.214	0.136	0.650	RESEARCH ...
129	128	Topic_2	0.091	0.057	0.852	Few Structura...
130	129	Topic_1	0.144	0.833	0.023	Vol. (012345...
131	130	Topic_2	0.225	0.157	0.619	STUDY PRO...
132	131	Topic_0	0.869	0.001	0.130	The Effect of...
133	132	Topic_1	0.067	0.741	0.193	EBioMedicine...
134	133	Topic_0	0.849	0.092	0.059	1017
135	134	Topic_2	0.076	0.113	0.810	
136	135	Topic_0	0.588	0.119	0.294	Received: 10 ...
137	136	Topic_2	0.025	0.476	0.499	RESEARCH ...
138	137	Topic_2	0.090	0.121	0.789	Received: 19 ...
139	138	Topic_0	0.733	0.047	0.220	1412   ...
140	139	Topic_1	0.181	0.523	0.296	ARTICLE OP...
141	140	Topic_0	0.911	0.028	0.061	https://doi.org...
142	141	Topic_0	0.521	0.174	0.205	Courtois-Am...
143	142	Topic_0	0.946	0.000	0.054	Citation: Xu, L...
144	143	Topic_1	0.071	0.742	0.187	22
145	144	Topic_0	0.957	0.006	0.036	Teles et al. B...
146	145	Topic_2	0.010	0.208	0.782	RESEARCH ...
147	146	Topic_1	0.035	0.965	0.000	RESEARCH ...
148	147	Topic_1	0.026	0.812	0.162	O R I G I N A ...
149	148	Topic_1	0.114	0.800	0.086	Clinical Ther...
150	149	Topic_2	0.088	0.310	0.602	Shin et al. BM...
151	150	Topic_2	0.028	0.186	0.786	Journal of Alz...
152	151	Topic_1	0.016	0.752	0.232	RESEARCH ...
153	152	Topic_1	0.214	0.683	0.104	Neuropsych...
154	153	Topic_0	0.642	0.000	0.358	Partnering a ...
155	154	Topic_1	0.025	0.895	0.080	Received: 20...
156	155	Topic_2	0.007	0.146	0.847	
157	156	Topic_1	0.008	0.836	0.156	O R I G I N A L...
158	157	Topic_2	0.215	0.376	0.410	Received: 28 ...
159	158	Topic_0	0.488	0.195	0.316	Whitfield et al.
160	159	Topic_2	0.041	0.192	0.767	Journal of Alz...
161	160	Topic_1	0.098	0.625	0.277	Journal of Ps...
162	161	Topic_1	0.082	0.675	0.243	1
163	162	Topic_0	0.966	0.006	0.028	1Morgan- Tri...
164	163	Topic_1	0.104	0.701	0.196	Original Inves...
165	164	Topic_0	0.762	0.039	0.199	Current Topic...
166	165	Topic_1	0.286	0.518	0.196	Effect of Level...
167	166	Topic_0	0.916	0.000	0.083	616
168	167	Topic_2	0.042	0.293	0.665	Ca/Mg Ratio, ...
169	168	Topic_1	0.129	0.800	0.071	10/7/23, 1.36 ...
170	169	Topic_0	0.890	0.035	0.075	STUDY PRO...
171	170	Topic_2	0.235	0.139	0.627	Study Proto...
172	171	Topic_2	0.019	0.352	0.629	nutrients
173	172	Topic_2	0.071	0.116	0.814	Journal of Alz...
174	173	Topic_2	0.054	0.437	0.509	1
175	174	Topic_1	0.025	0.672	0.303	ORIGINAL R...
176	175	Topic_0	0.748	0.063	0.189	RESEARCH ...
177	176	Topic_1	0.051	0.807	0.141	JAMDA 23 (2...
178	177	Topic_0	0.790	0.128	0.082	Smith et al. B...
179	178	Topic_1	0.064	0.664	0.272	Vol. (012345...
180	179	Topic_0	0.942	0.011	0.047	Protocol for a...
181	180	Topic_1	0.083	0.699	0.218	Association o...
182	181	Topic_0	0.978	0.001	0.021	Clare et al. B...
183	182	Topic_0	0.993	0.000	0.007	Satisfaction, ...
184	183	Topic_2	0.036	0.178	0.786	Combined tre...
185	184	Topic_2	0.042	0.364	0.594	Retrospective...
186	185	Topic_0	0.892	0.067	0.041	Australas J A...
187	186	Topic_2	0.131	0.111	0.758	Citation: Oga...
188	187	Topic_1	0.011	0.982	0.006	Association o...
189	188	Topic_1	0.028	0.602	0.370	Vol. (123456...
190	189	Topic_2	0.040	0.412	0.548	Received: 20...
191	190	Topic_0	0.928	0.006	0.066	STUDY PRO...
192	191	Topic_2	0.036	0.352	0.612	Original Inves...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

193	192	Topic_1	0.049	0.941	0.010	Alzheimer Di...
194	193	Topic_0	0.988	0.010	0.002	Aging & Ment...
195	194	Topic_2	0.091	0.228	0.682	1407
196	195	Topic_2	0.239	0.268	0.493	BRIEF COMM...
197	196	Topic_0	0.810	0.067	0.123	RESEARCH ...
198	197	Topic_0	0.905	0.017	0.078	RESIDENT & ...
199	198	Topic_1	0.019	0.765	0.216	A phase II stu...
200	199	Topic_0	0.972	0.002	0.025	Effectiveness...
201	200	Topic_1	0.010	0.921	0.068	Vol.:(012345...
202	201	Topic_1	0.094	0.497	0.409	The Journal o...
203	202	Topic_2	0.015	0.354	0.631	nutrients
204	203	Topic_0	0.869	0.018	0.113	Risks and be...
205	204	Topic_2	0.184	0.007	0.809	O R I G I N A ...
206	205	Topic_2	0.238	0.001	0.761	RESEARCH ...
207	206	Topic_1	0.016	0.616	0.368	Original Inves...
208	207	Topic_0	0.998	0.001	0.001	Journals Libr...
209	208	Topic_1	0.040	0.921	0.039	Original Inves...
210	209	Topic_1	0.413	0.428	0.159	www.thelanc...
211	210	Topic_1	0.221	0.751	0.028	Clinical Effect...
212	211	Topic_1	0.099	0.710	0.191	Original Inves...
213	212	Topic_0	0.957	0.005	0.038	Advance Car...
214	213	Topic_1	0.092	0.788	0.120	ARTICLE OP...
215	214	Topic_2	0.080	0.046	0.874	Cognitve co...
216	215	Topic_1	0.159	0.703	0.138	RESEARCH ...
217	216	Topic_2	0.401	0.073	0.527	Citation: Che...
218	217	Topic_1	0.034	0.925	0.041	Findings of Ef...
219	218	Topic_1	0.122	0.556	0.322	1
220	219	Topic_1	0.302	0.619	0.079	Short-term P...
221	220	Topic_2	0.309	0.124	0.567	Journal of Alz...
222	221	Topic_1	0.077	0.922	0.000	Dhadda et al...
223	222	Topic_2	0.208	0.187	0.604	Study Protoco...
224	223	Topic_0	0.870	0.100	0.030	The Trial to R...
225	224	Topic_0	0.976	0.000	0.024	Effects of a Vi...
226	225	Topic_2	0.273	0.141	0.586	Study Protoco...
227	226	Topic_0	0.765	0.007	0.228	Original Paper
228	227	Topic_1	0.070	0.891	0.039	ARTICLE CL...
229	228	Topic_0	0.632	0.000	0.368	Original Paper
230	229	Topic_0	0.949	0.018	0.034	RESEARCH ...
231	230	Topic_2	0.042	0.118	0.839	Levakov, Kapl...
232	231	Topic_1	0.186	0.814	0.000	Original Inves...
233	232	Topic_2	0.030	0.168	0.801	RESEARCH ...
234	233	Topic_1	0.294	0.664	0.042	Minocycline a...
235	234	Topic_1	0.054	0.931	0.015	Reactions to ...
236	235	Topic_1	0.084	0.828	0.088	Safety, Efficac...
237	236	Topic_2	0.465	0.032	0.503	International ...
238	237	Topic_2	0.102	0.414	0.483	lable at Scien...
239	238	Topic_2	0.037	0.069	0.895	NeuroImage ...
240	239	Topic_1	0.090	0.569	0.341	Vilfa et al. Alz...
241	240	Topic_1	0.080	0.888	0.032	Received: 23...
242	241	Topic_0	0.602	0.318	0.080	STUDY PRO...
243	242	Topic_1	0.012	0.737	0.252	Yulug et al. Tr...
244	243	Topic_1	0.181	0.724	0.094	Atrial of gant...
245	244	Topic_1	0.025	0.650	0.325	nutrients
246	245	Topic_2	0.161	0.094	0.745	Journal of Alz...
247	246	Topic_1	0.013	0.727	0.259	Riluzole, a gl...
248	247	Topic_0	0.732	0.257	0.010	Articles
249	248	Topic_2	0.168	0.039	0.793	Journal of Alz...
250	249	Topic_2	0.037	0.120	0.844	lable at Scien...
251	250	Topic_2	0.089	0.090	0.821	Contents list...
252	251	Topic_1	0.085	0.906	0.010	Dubois et al. ...
253	252	Topic_2	0.289	0.025	0.687	Clinical Trial ...
254	253	Topic_2	0.011	0.133	0.856	lable at Scien...
255	254	Topic_1	0.033	0.690	0.277	ArticleFrom st...
256	255	Topic_1	0.074	0.782	0.144	RESEARCH ...
257	256	Topic_1	0.054	0.820	0.125	1
258	257	Topic_2	0.007	0.231	0.762	lable at Scien...
259	258	Topic_1	0.375	0.390	0.235	Low dose lith...
260	259	Topic_1	0.045	0.682	0.273	Sleep depriva...
261	260	Topic_2	0.183	0.007	0.809	Properties of ...
262	261	Topic_1	0.009	0.872	0.119	BACE inhibiti...
263	262	Topic_1	0.161	0.800	0.039	Received: 24 ...
264	263	Topic_2	0.058	0.152	0.790	Vol.:(012345...



«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

265	264	Topic_1	0.068	0.857	0.074	Research Arti...
266	265	Topic_1	0.245	0.452	0.303	RESEARCH...
267	266	Topic_0	0.387	0.346	0.267	NULL HYPOT...
268	267	Topic_1	0.019	0.981	0.000	Association o...
269	268	Topic_0	0.987	0.000	0.013	Transitioning ...
270	269	Topic_1	0.018	0.730	0.253	1
271	270	Topic_2	0.025	0.245	0.730	Journal of Alz...
272	271	Topic_1	0.011	0.745	0.244	Effect of APO...
273	272	Topic_2	0.046	0.177	0.777	RESEARCH...
274	273	Topic_1	0.049	0.945	0.006	Safety and Eff...
275	274	Topic_2	0.087	0.291	0.622	RESEARCH...
276	275	Topic_2	0.028	0.328	0.644	https://doi.org...
277	276	Topic_1	0.117	0.673	0.210	RESEARCH...
278	277	Topic_1	0.209	0.553	0.238	Received: 10...
279	278	Topic_1	0.169	0.591	0.240	ORIGINAL AR...
280	279	Topic_1	0.040	0.940	0.021	2010   Clin ...
281	280	Topic_2	0.026	0.118	0.856	lable at Scien...
282	281	Topic_0	0.513	0.005	0.481	Mindfulness i...
283	282	Topic_2	0.088	0.261	0.651	Citation: Kad...
284	283	Topic_1	0.336	0.473	0.191	ARTICLE
285	284	Topic_2	0.291	0.162	0.548	Original Inves...
286	285	Topic_0	0.547	0.047	0.406	The Preventin...
287	286	Topic_2	0.009	0.336	0.655	Frontiers in E...
288	287	Topic_0	0.406	0.330	0.264	10   Ne...
289	288	Topic_1	0.015	0.552	0.433	Journal of Alz...
290	289	Topic_1	0.308	0.659	0.032	Original Inves...
291	290	Topic_2	0.034	0.075	0.891	Brain Perfus...
292	291	Topic_1	0.070	0.930	0.000	Evaluating th...
293	292	Topic_0	0.790	0.179	0.032	Comparison ...
294	293	Topic_1	0.033	0.509	0.459	Frontiers in L...
295	294	Topic_2	0.052	0.138	0.811	Citation: Shi, ...
296	295	Topic_0	0.715	0.041	0.245	RESEARCH...
297	296	Topic_2	0.355	0.209	0.436	30
298	297	Topic_1	0.117	0.453	0.429	Journal of Alz...
299	298	Topic_2	0.023	0.322	0.655	RESEARCH...
300	299	Topic_2	0.032	0.381	0.587	Association o...
301	300	Topic_0	0.792	0.046	0.162	547
302	301	Topic_1	0.053	0.916	0.031	https://doi.org...
303	302	Topic_2	0.253	0.202	0.544	A randomized...
304	303	Topic_2	0.162	0.021	0.817	See correspo...
305	304	Topic_2	0.176	0.142	0.682	Intensive vers...
306	305	Topic_2	0.130	0.008	0.863	International ...
307	306	Topic_1	0.116	0.750	0.134	RESEARCH...

ExampleSet (307 examples, 5 special attributes, 1 regular attribute)

Εικόνα 4.113: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε  $k = 3$  θέματα.

Αποτελέσματα για αριθμό θεμάτων  $k = 4$ :

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

ExampleSet (Extract Topics from Documents (LDA))

PerformanceVector (Extract Topics from Documents (LDA))

LDAModel (Extract Topics from Documents (LDA))

Criterion

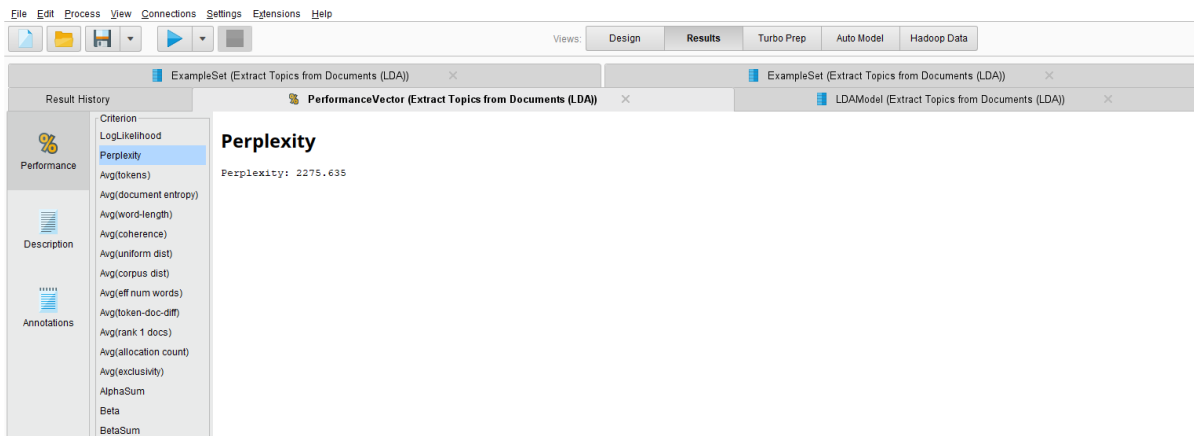
- LogLikelihood
- Perplexity
- Avg(tokens)
- Avg(document entropy)
- Avg(word-length)
- Avg(coherence)
- Avg(uniform dist)
- Avg(corpus dist)
- Avg(eff num words)
- Avg(token-doc-diff)
- Avg(rank 1 docs)
- Avg(allocation count)
- Avg(exclusivity)
- AlphaSum
- Beta
- BetaSum

LogLikelihood

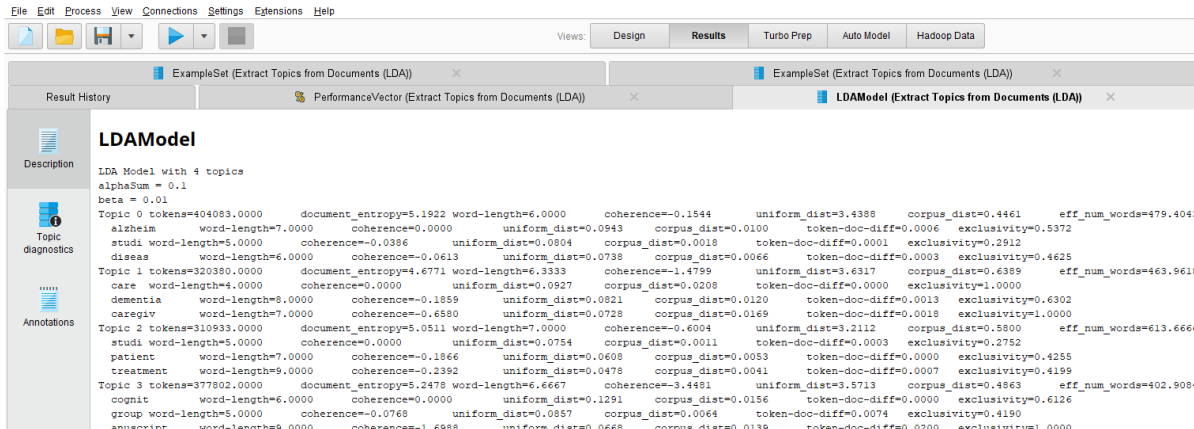
LogLikelihood: -11913865.771

Εικόνα 4.114: Η τιμή Loglikelihood για  $k = 4$  θέματα, Controlled Clinical Trial – πλήρη κείμενα.

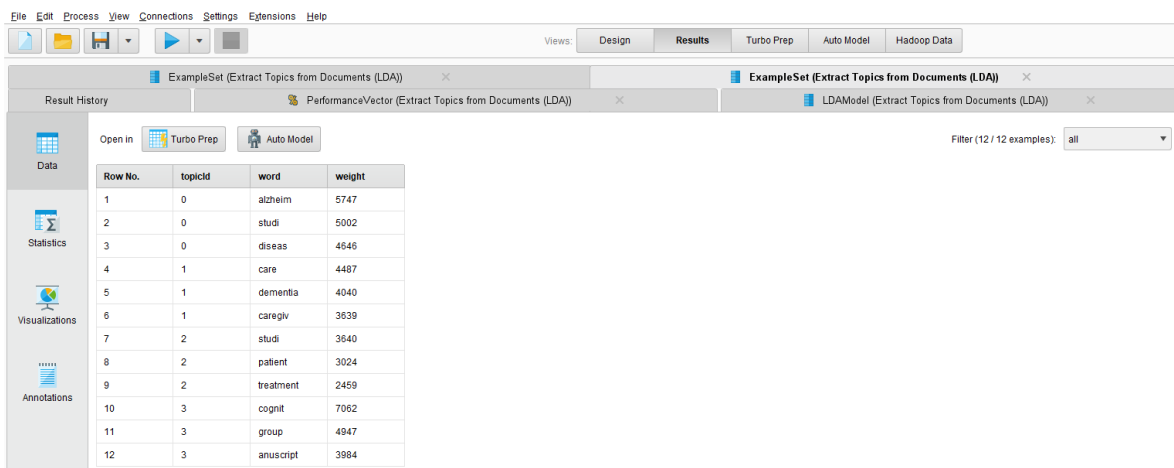
«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»



Εικόνα 4.115: Η τιμή Perplexity για k = 4 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.116: Το μοντέλο LDA για k = 4 θέματα, Controlled Clinical Trial – πλήρη κείμενα.



Εικόνα 4.117: Οι 3 πιο συχνές λέξεις κάθε θέματος και βάρη για k = 4 θέματα, Controlled Clinical Trial – πλήρη κείμενα.

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model Hadoop Data

Result History ExampleSet (Extract Topics from Documents (LDA)) ExampleSet (Extract Topics from Documents (LDA)) LDAModel (Extract Topics from Documents (LDA))

Open in Turbo Prep Auto Model Filter (307 / 307 examples): all

Row No.	documentid	prediction(Topic)	confidence(Topic_0)	confidence(Topic_1)	confidence(Topic_2)	confidence(Topic_3)	text
1	0	Topic_0	0.878	0.041	0.081	0.000	Alzheimer's
2	1	Topic_3	0.002	0.378	0.035	0.586	Research Arti...
3	2	Topic_3	0.200	0.273	0.000	0.527	The relative c...
4	3	Topic_0	0.687	0.133	0.171	0.008	780
5	4	Topic_2	0.247	0.151	0.516	0.086	Citation: Yu, X...
6	5	Topic_0	0.734	0.003	0.026	0.238	Intranasal Ins...
7	6	Topic_3	0.074	0.102	0.126	0.697	Journal of Alz...
8	7	Topic_2	0.012	0.379	0.597	0.013	233
9	8	Topic_0	0.655	0.041	0.016	0.288	Journal of Alz...
10	9	Topic_1	0.006	0.697	0.015	0.282	Original Paper
11	10	Topic_0	0.384	0.227	0.370	0.019	1Kondo T, et ...
12	11	Topic_2	0.371	0.030	0.588	0.012	Lerner et al. A...
13	12	Topic_0	0.751	0.017	0.132	0.100	Received: 7 ...
14	13	Topic_2	0.261	0.040	0.685	0.014	Nutrición
15	14	Topic_2	0.365	0.082	0.397	0.156	R E S E A R ...
16	15	Topic_0	0.822	0.022	0.099	0.057	This article is ...
17	16	Topic_0	0.576	0.157	0.260	0.007	Wunderlich et...
18	17	Topic_3	0.025	0.026	0.272	0.677	lable at Scien...
19	18	Topic_2	0.135	0.035	0.803	0.027	Journal of Alz...
20	19	Topic_1	0.026	0.601	0.018	0.355	Citation: Muñ...
21	20	Topic_1	0.099	0.636	0.005	0.260	RESEARCH ...
22	21	Topic_2	0.342	0.007	0.622	0.029	Journal of Alz...
23	22	Topic_0	0.599	0.172	0.110	0.119	Quantifying th...
24	23	Topic_0	0.925	0.048	0.022	0.005	Amyloid-Rela...
25	24	Topic_0	0.797	0.011	0.185	0.006	RESEARCH ...
26	25	Topic_0	0.553	0.065	0.349	0.032	Received: 7 J...
27	26	Topic_2	0.144	0.090	0.610	0.156	Received: 27 ...
28	27	Topic_3	0.004	0.481	0.001	0.514	Attention Con...
29	28	Topic_1	0.027	0.680	0.074	0.219	TYPE Origina...
30	29	Topic_2	0.098	0.056	0.749	0.097	OR I G I N A L...
31	30	Topic_1	0.063	0.482	0.043	0.412	Current Topic...
32	31	Topic_3	0.054	0.025	0.402	0.518	Vol.: (012345...
33	32	Topic_3	0.029	0.007	0.227	0.736	lable at Scien...
34	33	Topic_1	0.268	0.361	0.024	0.347	Baseline prev...
35	34	Topic_2	0.312	0.260	0.395	0.034	Am J of Geria...
36	35	Topic_0	0.497	0.019	0.369	0.115	Journal of Alz...
37	36	Topic_1	0.000	0.740	0.000	0.260	RESEARCH ...
38	37	Topic_2	0.039	0.023	0.937	0.000	635
39	38	Topic_3	0.297	0.135	0.266	0.301	Rationale, st...
40	39	Topic_0	0.742	0.055	0.059	0.144	Received: 12...
41	40	Topic_3	0.009	0.371	0.000	0.619	A Dyadic Appr...
42	41	Topic_2	0.133	0.018	0.439	0.411	J. Integr. Neu...
43	42	Topic_0	0.474	0.167	0.132	0.227	2169
44	43	Topic_2	0.081	0.103	0.811	0.005	Journal of Alz...
45	44	Topic_1	0.226	0.738	0.031	0.005	Public opinio...
46	45	Topic_3	0.346	0.090	0.037	0.527	Mediterranea...
47	46	Topic_3	0.043	0.085	0.234	0.638	RESEARCH ...
48	47	Topic_1	0.159	0.378	0.098	0.366	Research Arti...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

49	48	Topic_1	0.329	0.490	0.075	0.106	1
50	49	Topic_3	0.091	0.149	0.040	0.720	Available onli...
51	50	Topic_3	0.040	0.059	0.089	0.812	1
52	51	Topic_1	0.000	0.684	0.000	0.316	"It's Like a Cy...
53	52	Topic_3	0.164	0.173	0.007	0.656	Measuring ap...
54	53	Topic_1	0.274	0.453	0.061	0.211	646
55	54	Topic_1	0.000	0.800	0.000	0.200	Implications ...
56	55	Topic_1	0.005	0.769	0.013	0.213	Gaugler et al...
57	56	Topic_3	0.000	0.485	0.021	0.494	International ...
58	57	Topic_0	0.606	0.008	0.335	0.050	Journal of Alz...
59	58	Topic_1	0.000	0.825	0.000	0.175	The design a...
60	59	Topic_2	0.333	0.039	0.517	0.111	Whole-Brain ...
61	60	Topic_0	0.795	0.027	0.089	0.089	Vol. (012345...
62	61	Topic_0	0.720	0.036	0.067	0.177	RESEARCH ...
63	62	Topic_3	0.069	0.290	0.099	0.542	CLINICAL ST...
64	63	Topic_2	0.381	0.028	0.548	0.043	RESEARCH ...
65	64	Topic_0	0.641	0.013	0.005	0.340	Optimizing qu...
66	65	Topic_0	0.658	0.041	0.301	0.001	Nature Medici...
67	66	Topic_0	0.912	0.027	0.061	0.000	McDade et al...
68	67	Topic_1	0.000	0.665	0.000	0.335	One Call Mak...
69	68	Topic_2	0.088	0.106	0.627	0.180	Citation: LI, B...
70	69	Topic_0	0.518	0.184	0.015	0.282	Potashman e...
71	70	Topic_2	0.052	0.057	0.482	0.410	ARTICLE OP...
72	71	Topic_2	0.168	0.052	0.730	0.050	
73	72	Topic_2	0.313	0.104	0.374	0.209	OR I G I N A L...
74	73	Topic_0	0.555	0.016	0.426	0.003	Journal of Alz...
75	74	Topic_2	0.275	0.109	0.346	0.269	Parkinsons ...
76	75	Topic_1	0.012	0.808	0.080	0.101	Bentley et al. ...
77	76	Topic_1	0.000	0.718	0.000	0.282	Implementati...
78	77	Topic_0	0.552	0.014	0.111	0.323	Received: 21...
79	78	Topic_2	0.168	0.049	0.667	0.116	REGULAR A...
80	79	Topic_1	0.005	0.815	0.018	0.163	Cheung et al...
81	80	Topic_3	0.122	0.039	0.325	0.515	717www.eym...
82	81	Topic_0	0.656	0.006	0.289	0.048	Journal of Alz...
83	82	Topic_1	0.000	0.671	0.000	0.329	S H O R T R ...
84	83	Topic_2	0.319	0.030	0.629	0.022	190
85	84	Topic_3	0.143	0.016	0.055	0.787	Journal of Alz...
86	85	Topic_2	0.251	0.063	0.395	0.291	The MARBLE ...
87	86	Topic_3	0.157	0.108	0.015	0.719	Journal of Alz...
88	87	Topic_2	0.098	0.012	0.584	0.306	Citation: Choi...
89	88	Topic_0	0.782	0.132	0.031	0.055	Received: 9 J...
90	89	Topic_0	0.484	0.081	0.378	0.057	RESEARCH ...
91	90	Topic_0	0.792	0.040	0.118	0.049	Diaz-Galvan ...
92	91	Topic_1	0.033	0.728	0.000	0.239	Lessons fro...
93	92	Topic_0	0.802	0.031	0.016	0.151	RESEARCH ...
94	93	Topic_2	0.123	0.055	0.818	0.004	Biomedicine ...
95	94	Topic_3	0.111	0.144	0.024	0.721	Journal of Alz...
96	95	Topic_2	0.015	0.105	0.686	0.194	TaggedEndG...
97	96	Topic_3	0.088	0.412	0.041	0.459	RESEARCH ...
98	97	Topic_0	0.729	0.032	0.000	0.240	Prospective a...
99	98	Topic_0	0.521	0.150	0.040	0.289	Effect of Meth...
100	99	Topic_0	0.513	0.279	0.141	0.068	RESEARCH ...
101	100	Topic_3	0.284	0.004	0.049	0.663	[18F]FDNP ...
102	101	Topic_2	0.125	0.014	0.589	0.272	Received: De...
103	102	Topic_3	0.055	0.182	0.035	0.728	©Journal of S...
104	103	Topic_0	0.902	0.034	0.046	0.018	Efficacy and ...
105	104	Topic_0	0.513	0.163	0.000	0.323	Received: 25 ...
106	105	Topic_0	0.824	0.054	0.098	0.024	Kile et al. BM...
107	106	Topic_3	0.078	0.168	0.055	0.698	Rationale an...
108	107	Topic_3	0.013	0.045	0.025	0.917	Mnemonic str...
109	108	Topic_1	0.000	0.710	0.069	0.220	Vol. (012345...
110	109	Topic_2	0.049	0.105	0.584	0.262	32
111	110	Topic_0	0.550	0.019	0.123	0.308	Benfotamine...
112	111	Topic_2	0.013	0.121	0.866	0.000	
113	112	Topic_3	0.038	0.008	0.276	0.678	Estradiol trea...
114	113	Topic_2	0.028	0.022	0.927	0.022	OR I G I N A L...
115	114	Topic_3	0.369	0.110	0.000	0.521	A Randomize...
116	115	Topic_2	0.034	0.020	0.886	0.059	OR I G I N A L...
117	116	Topic_3	0.028	0.005	0.226	0.740	Vol.: (012345...
118	117	Topic_1	0.101	0.609	0.201	0.089	Journal of Alz...
119	118	Topic_1	0.003	0.540	0.075	0.383	International ...
120	119	Topic_3	0.013	0.213	0.000	0.774	Using Smart...

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

Statistics	121	120	Topic_3	0.330	0.034	0.031	0.605	Preliminary e...	
	122	121	Topic_0	0.498	0.100	0.395	0.006	RESEARCH ...	
	Visualizations	123	122	Topic_2	0.061	0.065	0.873	0.000	RESEARCH ...
		124	123	Topic_3	0.029	0.426	0.018	0.526	REGULAR A...
		125	124	Topic_1	0.047	0.520	0.139	0.294	Nighttime Agi...
		126	125	Topic_0	0.740	0.011	0.242	0.007	RESEARCH ...
		127	126	Topic_2	0.048	0.179	0.687	0.087	ORIGINAL IN...
		128	127	Topic_3	0.043	0.147	0.174	0.636	RESEARCH ...
		129	128	Topic_3	0.161	0.032	0.021	0.786	Few Structura...
		130	129	Topic_2	0.257	0.035	0.709	0.000	Vol. (012345...
		131	130	Topic_3	0.218	0.165	0.086	0.530	STUDY PRO...
		132	131	Topic_3	0.000	0.440	0.000	0.559	The Effect of...
		133	132	Topic_0	0.651	0.058	0.269	0.022	EBioMedicine...
		134	133	Topic_1	0.058	0.609	0.014	0.319	1017
		135	134	Topic_0	0.491	0.277	0.222	0.009	
		136	135	Topic_1	0.217	0.421	0.000	0.362	Received: 10 ...
		137	136	Topic_0	0.782	0.032	0.064	0.122	RESEARCH ...
		138	137	Topic_3	0.000	0.067	0.392	0.541	Received: 19 ...
		139	138	Topic_1	0.028	0.554	0.080	0.338	1412   ...
		140	139	Topic_0	0.418	0.037	0.340	0.205	ARTICLE OP...
141		140	Topic_1	0.019	0.698	0.000	0.282	https://doi.org...	
142		141	Topic_3	0.162	0.378	0.052	0.409	Courtois-Ami...	
143	142	Topic_1	0.000	0.679	0.000	0.321	Citation: Xu, L...		
144	143	Topic_0	0.550	0.061	0.330	0.058	22		
Statistics	145	144	Topic_1	0.001	0.788	0.000	0.211	Teles et al. B...	
	146	145	Topic_3	0.053	0.007	0.468	0.472	RESEARCH ...	
	147	146	Topic_0	0.632	0.023	0.345	0.000	RESEARCH ...	
	148	147	Topic_2	0.111	0.042	0.845	0.002	O R I G I N A ...	
	149	148	Topic_0	0.491	0.070	0.429	0.010	Clinical Ther...	
	150	149	Topic_3	0.158	0.027	0.325	0.489	Shin et al. BM...	
	151	150	Topic_0	0.670	0.029	0.042	0.259	Journal of Alz...	
	152	151	Topic_2	0.040	0.035	0.899	0.026	RESEARCH ...	
	153	152	Topic_0	0.564	0.076	0.186	0.175	Neuropsych...	
	154	153	Topic_3	0.000	0.276	0.003	0.721	Partnering a ...	
	155	154	Topic_2	0.167	0.024	0.803	0.006	Received: 20 ...	
	156	155	Topic_3	0.030	0.006	0.315	0.650		
	157	156	Topic_2	0.045	0.017	0.918	0.019	O R I G I N A L...	
	158	157	Topic_0	0.584	0.094	0.004	0.319	Received: 28 ...	
	159	158	Topic_3	0.249	0.272	0.001	0.479	Whitfield et al.	
	160	159	Topic_2	0.046	0.017	0.525	0.412	Journal of Alz...	
	161	160	Topic_2	0.114	0.047	0.708	0.132	Journal of Ps...	
	162	161	Topic_0	0.790	0.048	0.064	0.097	1	
	163	162	Topic_1	0.000	0.935	0.014	0.051	1Morgan- Tri...	
	164	163	Topic_0	0.832	0.042	0.000	0.126	Original Inves...	
165	164	Topic_3	0.044	0.473	0.001	0.482	Current Topic...		
166	165	Topic_0	0.484	0.051	0.097	0.369	Effect of Level...		
167	166	Topic_1	0.005	0.618	0.000	0.377	616		
168	167	Topic_0	0.465	0.022	0.232	0.281	Ca/Mg Ratio. ...		
Statistics	169	168	Topic_0	0.462	0.076	0.406	0.056	10/7/23, 1:36 ...	
	170	169	Topic_1	0.032	0.668	0.003	0.297	STUDY PRO...	
	171	170	Topic_3	0.071	0.172	0.197	0.500	Study Protoco...	
	172	171	Topic_2	0.169	0.015	0.547	0.269	nutrients	
	173	172	Topic_3	0.062	0.033	0.125	0.780	Journal of Alz...	
	174	173	Topic_2	0.251	0.027	0.437	0.284	1	
	175	174	Topic_2	0.167	0.042	0.773	0.018	ORIGINAL R...	
	176	175	Topic_1	0.250	0.604	0.000	0.146	RESEARCH ...	
	177	176	Topic_0	0.822	0.024	0.196	0.158	JAMDA 23 (2...	
	178	177	Topic_1	0.132	0.769	0.066	0.033	Smith et al. B...	
	179	178	Topic_0	0.453	0.053	0.392	0.102	Vol. (012345...	
	180	179	Topic_1	0.014	0.684	0.000	0.302	Protocol for a ...	
	181	180	Topic_0	0.645	0.026	0.156	0.173	Association o...	
	182	181	Topic_1	0.000	0.931	0.000	0.069	Clare et al. B...	
	183	182	Topic_1	0.000	0.728	0.000	0.272	Satisfaction, ...	
	184	183	Topic_3	0.149	0.005	0.098	0.748	Combined tre...	
	185	184	Topic_0	0.473	0.000	0.106	0.421	Retrospective...	
	186	185	Topic_1	0.055	0.756	0.011	0.178	Australas J A...	
	187	186	Topic_3	0.074	0.078	0.336	0.512	Citation: Oga...	
	188	187	Topic_0	0.980	0.008	0.012	0.000	Association o...	
189	188	Topic_2	0.389	0.037	0.549	0.024	Vol. (123456...		
190	189	Topic_2	0.208	0.046	0.656	0.091	Received: 20...		
191	190	Topic_1	0.007	0.881	0.008	0.104	STUDY PRO...		
192	191	Topic_0	0.814	0.027	0.001	0.158	Original Inves...		
Annotations									

«Εξόρυξη κειμένου και εφαρμογές στην Βιοϊατρική Μηχανική»

193	192	Topic_0	0.690	0.032	0.278	0.000	Alzheimer DL...
194	193	Topic_1	0.002	0.978	0.015	0.005	Aging & Ment...
195	194	Topic_0	0.730	0.070	0.008	0.193	1407
196	195	Topic_3	0.331	0.072	0.127	0.470	BRIEF COMM...
197	196	Topic_1	0.053	0.471	0.018	0.458	RESEARCH ...
198	197	Topic_1	0.095	0.807	0.000	0.099	RESIDENT & ...
199	198	Topic_2	0.422	0.010	0.482	0.086	A phase II stu...
200	199	Topic_1	0.003	0.581	0.000	0.416	Effectiveness...
201	200	Topic_2	0.029	0.019	0.951	0.000	Vol.:(012345...
202	201	Topic_3	0.242	0.031	0.230	0.497	The Journal o...
203	202	Topic_2	0.039	0.018	0.763	0.180	nutrients
204	203	Topic_1	0.005	0.546	0.001	0.447	Risks and be...
205	204	Topic_3	0.000	0.113	0.089	0.797	O R I G I N A ...
206	205	Topic_3	0.000	0.168	0.014	0.818	RESEARCH ...
207	206	Topic_3	0.347	0.008	0.295	0.349	Original Inves...
208	207	Topic_1	0.000	0.995	0.002	0.003	Journals Libr...
209	208	Topic_0	0.823	0.029	0.138	0.010	Original Inves...
210	209	Topic_0	0.617	0.313	0.059	0.012	www.thelanc...
211	210	Topic_0	0.811	0.150	0.001	0.039	Clinical Effect...
212	211	Topic_2	0.243	0.056	0.533	0.168	Original Inves...
213	212	Topic_1	0.009	0.577	0.000	0.414	Advance Car...
214	213	Topic_2	0.392	0.073	0.518	0.017	ARTICLE OP...
215	214	Topic_3	0.123	0.018	0.050	0.808	Cognitive co...
216	215	Topic_0	0.448	0.093	0.387	0.072	RESEARCH ...
217	216	Topic_3	0.004	0.301	0.190	0.506	Citation: Che...
218	217	Topic_0	0.899	0.017	0.069	0.016	Findings of Ef...
219	218	Topic_2	0.315	0.098	0.372	0.215	1
220	219	Topic_0	0.730	0.128	0.000	0.142	Short-term P...
221	220	Topic_3	0.153	0.165	0.004	0.677	Journal of Alz...
222	221	Topic_0	0.927	0.060	0.013	0.000	Dhadda et al...
223	222	Topic_2	0.028	0.149	0.447	0.376	Study Protoco...
224	223	Topic_1	0.122	0.833	0.020	0.025	The Trial to R...
225	224	Topic_1	0.000	0.653	0.000	0.347	Effects of a Vi...
226	225	Topic_3	0.022	0.227	0.268	0.483	Study Protoco...
227	226	Topic_1	0.113	0.677	0.001	0.210	Original Paper
228	227	Topic_2	0.468	0.047	0.485	0.000	ARTICLE CL...
229	228	Topic_3	0.036	0.368	0.000	0.596	Original Paper
230	229	Topic_1	0.029	0.914	0.000	0.057	RESEARCH ...
231	230	Topic_0	0.464	0.062	0.227	0.247	Levakov, Kapl...
232	231	Topic_0	0.823	0.173	0.005	0.000	Original Inves...
233	232	Topic_3	0.110	0.015	0.192	0.683	RESEARCH ...
234	233	Topic_0	0.488	0.256	0.232	0.023	Minocycline a...
235	234	Topic_0	0.857	0.023	0.109	0.011	Reactions to ...
236	235	Topic_0	0.848	0.052	0.053	0.047	Safety, Efficac...
237	236	Topic_3	0.024	0.275	0.025	0.676	International ...
238	237	Topic_2	0.233	0.111	0.462	0.194	lable at Scien...
239	238	Topic_3	0.292	0.015	0.055	0.639	NeuroImage ...
240	239	Topic_0	0.477	0.072	0.336	0.115	Viña et al. Alz...
241	240	Topic_0	0.825	0.040	0.102	0.033	Received: 23...
242	241	Topic_1	0.127	0.529	0.286	0.057	STUDY PRO...
243	242	Topic_2	0.295	0.016	0.648	0.041	Yulug et al. Tr...
244	243	Topic_0	0.785	0.038	0.005	0.172	A trial of gant...
245	244	Topic_2	0.289	0.036	0.607	0.068	nutrients
246	245	Topic_3	0.167	0.080	0.042	0.711	Journal of Alz...
247	246	Topic_0	0.667	0.010	0.248	0.075	Riluzole, a gl...
248	247	Topic_1	0.171	0.692	0.106	0.032	Articles
249	248	Topic_3	0.047	0.081	0.031	0.840	Journal of Alz...
250	249	Topic_3	0.028	0.021	0.159	0.792	lable at Scien...
251	250	Topic_3	0.180	0.055	0.041	0.724	Contents list...
252	251	Topic_0	0.702	0.058	0.240	0.000	Dubois et al. ...
253	252	Topic_3	0.046	0.178	0.032	0.745	Clinical Trial ...
254	253	Topic_3	0.058	0.004	0.161	0.777	lable at Scien...
255	254	Topic_2	0.029	0.047	0.884	0.040	ArticleFrom st...
256	255	Topic_0	0.505	0.075	0.413	0.007	RESEARCH ...
257	256	Topic_2	0.458	0.056	0.484	0.002	1
258	257	Topic_3	0.080	0.004	0.261	0.655	lable at Scien...
259	258	Topic_0	0.393	0.177	0.041	0.389	Low dose lith...
260	259	Topic_2	0.199	0.000	0.549	0.252	Sleep depriv...
261	260	Topic_3	0.113	0.041	0.000	0.846	Properties of ...
262	261	Topic_0	0.960	0.012	0.016	0.012	BACE inhibiti...
263	262	Topic_2	0.259	0.133	0.570	0.038	Received: 24...
264	263	Topic_3	0.087	0.029	0.255	0.629	Vol.:(012345...

265	264	Topic_2	0.121	0.052	0.820	0.007	Research Arti...
266	265	Topic_0	0.607	0.215	0.088	0.091	RESEARCH ...
267	266	Topic_0	0.614	0.222	0.015	0.149	NULL HYPOT...
268	267	Topic_0	0.953	0.016	0.031	0.000	Association o...
269	268	Topic_1	0.000	0.797	0.000	0.203	Transitioning ...
270	269	Topic_0	0.594	0.013	0.353	0.039	1
271	270	Topic_2	0.220	0.021	0.611	0.147	Journal of Alz...
272	271	Topic_0	0.796	0.001	0.050	0.153	Effect of APO...
273	272	Topic_3	0.136	0.023	0.209	0.632	RESEARCH ...
274	273	Topic_0	0.902	0.036	0.062	0.000	Safety and Eff...
275	274	Topic_3	0.381	0.061	0.029	0.530	RESEARCH ...
276	275	Topic_3	0.245	0.011	0.144	0.599	https://doi.org...
277	276	Topic_0	0.490	0.014	0.410	0.087	RESEARCH ...
278	277	Topic_2	0.000	0.100	0.731	0.169	Received: 10...
279	278	Topic_2	0.024	0.073	0.454	0.449	ORIGINAL AR...
280	279	Topic_2	0.131	0.043	0.825	0.000	2010   Clin ...
281	280	Topic_3	0.044	0.013	0.246	0.698	lable at Scien...
282	281	Topic_3	0.117	0.261	0.005	0.617	Mindfulness L...
283	282	Topic_2	0.257	0.010	0.502	0.231	Citation: Kad...
284	283	Topic_2	0.048	0.217	0.543	0.191	ARTICLE
285	284	Topic_3	0.209	0.132	0.000	0.659	Original Inves...
286	285	Topic_3	0.109	0.258	0.000	0.633	The Preventin...
287	286	Topic_2	0.132	0.005	0.522	0.341	Frontiers in E...
288	287	Topic_3	0.054	0.179	0.263	0.504	10   Ne...
289	288	Topic_0	0.704	0.009	0.127	0.161	Journal of Alz...
290	289	Topic_0	0.482	0.110	0.122	0.287	Original Inves...
291	290	Topic_3	0.178	0.009	0.030	0.782	Brain Perfusi...
292	291	Topic_0	0.909	0.056	0.036	0.000	Evaluating th...
293	292	Topic_1	0.201	0.588	0.008	0.203	Comparison ...
294	293	Topic_2	0.129	0.059	0.809	0.004	Frontiers in L...
295	294	Topic_2	0.032	0.051	0.620	0.296	Citation: Shi, ...
296	295	Topic_1	0.033	0.622	0.072	0.273	RESEARCH ...
297	296	Topic_0	0.422	0.277	0.010	0.291	30
298	297	Topic_0	0.448	0.085	0.280	0.187	Journal of Alz...
299	298	Topic_0	0.391	0.019	0.318	0.272	RESEARCH ...
300	299	Topic_0	0.829	0.025	0.000	0.145	Association o...
301	300	Topic_1	0.080	0.636	0.003	0.281	547
302	301	Topic_0	0.777	0.045	0.171	0.007	https://doi.org...
303	302	Topic_3	0.052	0.030	0.096	0.822	A randomized...
304	303	Topic_3	0.256	0.078	0.010	0.655	See correspo...
305	304	Topic_3	0.468	0.049	0.000	0.484	Intensive vers...
306	305	Topic_3	0.006	0.081	0.045	0.867	International ...
307	306	Topic_0	0.874	0.062	0.012	0.053	RESEARCH ...

Εικόνα 4.118: Η κατηγοριοποίηση των πλήρων κειμένων Controlled Clinical Trial σε  $k = 4$  θέματα.

## 4.5 Σύγκριση αποτελεσμάτων

Στην τελευταία ενότητα της διπλωματικής εργασίας, αναλύονται και συγκρίνονται τα αποτελέσματα που παρατέθηκαν στις ενότητες 4.3 και 4.4. Η ανάλυση και η σύγκριση θα γίνει ανά σώμα κειμένων (Clinical Trial I, Clinical Trial II, Controlled Clinical Trial), ανά κατηγορία άρθρων (περιλήψεων ή πλήρων κειμένων) και ανά αριθμό αναζητούμενων θεμάτων, έστω  $k$  ( $k = 2, 3, 4$  αναζητούμενα θέματα). Βάσει αυτών, θα δοθεί η απάντηση στο βασικό ερώτημα που έχει τεθεί στην ενότητα 3.3.

#### 4.5.1 Clinical Trial I

Αρχικά στον επόμενο Πίνακα 4.2 συγκεντρώνονται οι τιμές των παραμέτρων loglikelihood και perplexity για τα διάφορα μοντέλα ( $k=2, 3, 4$ ) για τα πλήρη κείμενα και για τις περιλήψεις:

Πίνακας 4.2: Τιμές παραμέτρων του μοντέλου των πλήρων κειμένων και των περιλήψεών τους, Clinical Trial I (L = Loglikelihood, P = Perplexity).

	<b>k = 2</b>	<b>k = 3</b>	<b>k = 4</b>
<b>Full Texts</b>	L = - 631412.699 P = 1729.352	L = - 631988.759 P = 1568.041	L = - 634818.019 P = 1457.686
<b>Abstracts</b>	L = - 19960.477 P = 451.1214	L = - 20001.790 P = 411.727	L = - 19949.457 P = 380.873

Στον διαχωρισμό των περιλήψεων και των αντίστοιχων κειμένων σε 2 θέματα, αρχικά ελέγχονται οι τιμές Loglikelihood και Perplexity. Όπως αναφέρθηκε στην ενότητα 4.2, οι τιμές αυτές πρέπει να είναι αντίστροφης λογικής, η πρώτη μεγάλη και η δεύτερη μικρή. Στην συγκεκριμένη περίπτωση, το βέλτιστο μοντέλο είναι αυτό των 2 θεμάτων των περιλήψεων καθώς η τιμή Loglikelihood είναι ίση με -19960.477 και η τιμή Perplexity ισούται με 451.121 ενώ το μοντέλο 2 θεμάτων των πλήρων κειμένων είναι -631412.699 και 1729.352 αντίστοιχα.

Στις εικόνες 4.27 και 4.75 εμφανίζονται οι 3 επικρατούσες λέξεις κάθε θέματος και για τις δύο κατηγορίες (περιλήψεις και πλήρη κείμενα). Συγκριτικά παρουσιάζονται και στον Πίνακα 4.3. Παρατηρείται ότι το Topic\_0 των κειμένων με το Topic\_1 των περιλήψεων έχουν τις ίδιες ακριβώς λέξεις αλλά με διαφορετικά βάρη λόγω της μικρής και της μεγάλης έκτασης των περιλήψεων και των κειμένων, αντιστοίχως. Αντίθετα, το Topic\_1 των κειμένων σε σύγκριση με το Topic\_0 των περιλήψεων, αποτελούνται από τελείως διαφορετικές λέξεις. Για τις περιλήψεις, οι λέξεις είναι “clinic”, “particip” και “patient”, ενώ για τα κείμενα είναι “cell”, “use” και “alzheim”. Γίνεται αντιληπτό ότι οι λέξεις του πρώτου θέματος των περιλήψεων είναι πιο γενικές σε σχέση με αυτές των κειμένων που είναι πιο ειδικές.

Εν συνεχεία, η κατηγοριοποίηση των κειμένων και των περιλήψεων παρουσιάζονται στις εικόνες 4.76 και 4.28 αντίστοιχα. Στον Πίνακα 4.4.a παρουσιάζεται η αναλογία της αντιστοίχισης των κειμένων και των περιλήψεων τους στο ίδιο θέμα. Όπως φαίνεται το Topic\_0 των κειμένων με τις περιλήψεις τους έχει συμφωνία 25% και το Topic\_1 έχει συμφωνία 20%. Τα ποσοστά συμφωνίας θεμάτων είναι μικρά λόγω του μικρού πλήθους των κειμένων και των περιλήψεων της κατηγορίας Clinical Trial I.



Αν όμως αντιμεταθέσουμε αμοιβαία τα ονόματα των Topic\_0 και Topic\_1 των περιλήψεων, όπως υποδεικνύει ο Πίνακας 4.3 με τις επικρατούσες λέξεις, έχουμε τον Πίνακα 4.4.β, όπου τα ποσοστά κοινής κατηγοριοποίησης άρθρων (διαγώνια στοιχεία του Πίνακα) αυξάνουν έναντι του Πίνακα 4.4.α.

Πίνακας 4.3: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 2$  (Clinical Trial I).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	studi	972	Topic_0	clinic	27
Topic_0	dose	966	Topic_0	particip	25
Topic_0	inject	415	Topic_0	patient	25
Topic_1	cell	502	Topic_1	dose	49
Topic_1	use	329	Topic_1	studi	37
Topic_1	alzheim	278	Topic_1	placebo	27

Πίνακας 4.4.α: Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεών τους για  $k = 2$ , Clinical Trial I.

Full text \ Abstract	Topic_0	Topic_1	Total
Topic_0	3 (25%)	9	<b>12</b>
Topic_1	4	1 (20%)	<b>5</b>
Total	<b>7</b>	<b>10</b>	<b>17</b>

Πίνακας 4.4.β: Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεών τους για  $k = 2$ , Clinical Trial I (με αντιμετάθεση ονομάτων topic\_0 και topic\_1).

Full text \ Abstract	Topic_1	Topic_0	Total
Topic_0	9 (75%)	3	<b>12</b>
Topic_1	1	4 (80%)	<b>5</b>
Total	<b>10</b>	<b>7</b>	<b>17</b>

Τέλος, από τα στοιχεία που παρατίθενται στα LDA μοντέλα και των δύο κατηγοριών (εικόνες 4.26 και 4.74), αξίζει να σχολιαστούν οι παράμετροι “document\_entropy”, “coherence”, “uniform\_dist” και “corpus\_dist”. Η παράμετρος “document\_entropy” και στις δύο περιπτώσεις, έχει παρόμοιες τιμές που σημαίνει ότι τόσο τα κείμενα όσο και οι περιλήψεις τους έχουν τον ίδιο βαθμό περιπλοκότητας. Συνεχίζοντας με την παράμετρο “coherence”, τα θέματα των κειμένων έχουν τιμές που πλησιάζουν την τιμή 0 σε σχέση με τις περιλήψεις, άρα η ποιότητα των θεμάτων που έχουν εξαχθεί και αντιστοιχηθεί με τα έγγραφα είναι καλύτερη από αυτή των περιλήψεων. Η παράμετρος “uniform\_dist” των περιλήψεων και των κειμένων κυμαίνονται περίπου στο ίδιο εύρος τιμών και αυτό δηλώνει ότι τα θέματα κατανέμονται με παρόμοια συχνότητα και στις δύο περιπτώσεις. Τέλος, οι τιμές παραμέτρου “corpus\_dist” των περιλήψεων είναι μεγαλύτερες από τις τιμές της παραμέτρου των κειμένων, συνεπώς τα θέματα των κειμένων είναι εξειδικευμένα και κατανέμονται πιο συγκεκριμένα στα έγγραφα.

Στα μοντέλα των  $k = 3$  θεμάτων θα συγκριθούν πρώτα οι τιμές Loglikelihood και Perplexity των περιλήψεων και των κειμένων. Όπως και στα μοντέλα  $k = 2$  θεμάτων, οι περιλήψεις, όπου η Loglikelihood ισούται με -20001.790 και η τιμή Perplexity ισούται με 411.727, είναι καλύτερες τιμές από αυτές των πλήρων κειμένων, όπου η Loglikelihood ισούται με -631988.759 και η τιμή Perplexity ισούται με 1568.041 καθώς η τιμή Loglikelihood πρέπει να είναι μεγάλη και η τιμή Perplexity να είναι μικρή.

Πίνακας 4.5 : Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 3$  (Clinical Trial I).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	dose	896	Topic_0	inject	25
Topic_0	studi	891	Topic_0	dose	19
Topic_0	day	371	Topic_0	pharmacokinet	15
Topic_1	inject	415	Topic_1	placebo	27
Topic_1	cell	350	Topic_1	studi	27
Topic_1	use	214	Topic_1	particip	25
Topic_2	alzheim	299	Topic_2	subject	20
Topic_2	week	204	Topic_2	clinic	16
Topic_2	trial	189	Topic_2	model	14

Στις εικόνες 4.32 και 4.80 εμφανίζονται οι 3 επικρατούσες λέξεις κάθε θέματος με τα βάρη τους. Συγκριτικά επαναλαμβάνονται στον Πίνακα 4.5. Παρατηρείται ότι τα θέματα των περιλήψεων σε σύγκριση με των κειμένων, που έχουν εξαχθεί, είναι τελείως διαφορετικά μεταξύ τους. Αυτό συμβαίνει διότι, η αύξηση αριθμού των θεμάτων ανάγκασε τον αλγόριθμο LDA να ανακαλύψει περισσότερα θέματα στις περιλήψεις που έχουν κρυμμένη σημασιολογία, καθώς οι περιλήψεις όχι μόνο εστιάζουν στα κυριότερα σημεία του κειμένου αλλά είναι και αποτέλεσμα υποκειμενικότητας, σε σχέση με τον LDA που επιδιώκει να είναι αντικειμενικός. Συνεπώς, είναι αναμενόμενο να διαφέρουν μεταξύ τους. Η διαφορά των θεμάτων θα έχει ως επακόλουθο αποτέλεσμα να μην μπορεί να γίνει η σύγκριση στις κατηγοριοποιήσεις των περιλήψεων και των κειμένων που έχουν γίνει, όπως αναπαρίστανται στις εικόνες 4.33 και 4.81 αντίστοιχα. Αντιθέτως, αξίζει να σχολιαστεί η συμφωνία των θεμάτων των κειμένων με τις περιλήψεις τους. Στον πίνακα 4.6 παρατηρούνται στην διαγώνιο οι ποσοστιαίες συμφωνίες των θεμάτων. Η αύξηση του αριθμού των θεμάτων, από  $k = 2$  σε  $k = 3$  δεν επηρέασε τα ποσοστά συμφωνίας θεμάτων καθώς το σύνολο των κειμένων παρέμεινε το ίδιο.

Πίνακας 4.6 : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους, για  $k = 3$ , Clinical Trial I.

Full Text\ Abstarct	Topic_0	Topic_1	Topic_2	Total
Topic_0	2 (25%)	3	3	<b>8</b>
Topic_1	1	1 (20%)	3	<b>5</b>
Topic_2	1	2	1 (25%)	<b>4</b>
Total	<b>4</b>	<b>6</b>	<b>7</b>	<b>17</b>

Γι' αυτό το λόγο θα εξεταστούν οι παράμετροι “document\_entropy”, “coherence”, “uniform\_dist” και “corpus\_dist” των LDA μοντέλων (εικόνες 4.31 και 4.79). Όσον αφορά τις τιμές της παραμέτρου “document\_entropy” διαπιστώνεται ότι οι περιλήψεις και τα κείμενα έχουν παρόμοια περιπλοκότητα, καθώς οι τιμές είναι παραπλήσιες. Εν συνεχεία, οι τιμές της παραμέτρου “coherence” των κειμένων είναι μεγαλύτερες από αυτές των περιλήψεων, άρα έχει γίνει καλύτερη αντιστοίχιση θέματος-κειμένου από ότι θέματος-περιλήψεως. Η παράμετρος “corpus\_dist” του εκάστοτε θέματος των κειμένων είναι μεγαλύτερες από τις αντίστοιχες τιμές της παραμέτρου του εκάστοτε θέματος των περιλήψεων. Επομένως, τα θέματα των κειμένων έχουν μεγαλύτερη εξειδίκευση συγκριτικά με τα θέματα των περιλήψεων που είναι πιο γενικευμένα. Οι τιμές της παραμέτρου “uniform\_dist” των περιλήψεων είναι μικρότερες από εκείνες των κειμένων και αυτό σημαίνει ότι υπάρχει η κατανομή θεμάτων στα κείμενα έχει γίνει ισότιμα.

Τέλος, τα μοντέλα των  $k = 4$  θεμάτων των περιλήψεων και των πλήρων κειμένων έχουν τιμή Loglikelihood -19949.457 και -634818.019 και Perplexity 380.873 και 1457.686 αντίστοιχα. Όπως και στα προηγούμενα είδη μοντέλων, οι τιμές αυτές των περιλήψεων δείχνουν ότι το μοντέλο αυτό είναι βέλτιστο συγκριτικά με το μοντέλο των κειμένων.

Πίνακας 4.7: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 4$  (Clinical Trial I).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	dose	306	Topic_0	studi	26
Topic_0	model	214	Topic_0	particip	25
Topic_0	patient	197	Topic_0	administr	20
Topic_1	cell	335	Topic_1	clinic	21
Topic_1	trial	142	Topic_1	model	14
Topic_1	inject	140	Topic_1	ngf	13
Topic_2	inject	217	Topic_2	patient	22
Topic_2	org	201	Topic_2	dose	20
Topic_2	group	198	Topic_2	result	19
Topic_3	studi	796	Topic_3	dose	29
Topic_3	dose	660	Topic_3	inject	25
Topic_3	day	372	Topic_3	subject	20

Στις εικόνες 4.37 και 4.85 παρουσιάζονται οι λέξεις του κάθε θέματος με τα βάρη τους. Σε αυτά τα μοντέλα, μπορεί να ειπωθεί ότι τα θέματα των περιλήψεων και των κειμένων είναι σχεδόν σχετικά μεταξύ τους, όμως δεν μπορεί να θεωρηθεί απόλυτη η ταύτιση των κατηγοριοποιήσεων που έχουν γίνει.

Ομοίως, στον πίνακα 4.8 εμφανίζονται τα ποσοστά συμφωνίας για  $k = 4$ . Η συμφωνία θεμάτων εμφανίζεται συγκεκριμένα στην διαγώνιο του πίνακα, όπου το Topic\_1 και το Topic\_2 έχουν 0% συμφωνία θεμάτων καθώς η αύξηση του αριθμού θεμάτων από  $k = 3$  σε  $k = 4$ , επηρέασε την κατανομή των κειμένων και των περιλήψεων και όπως φαίνεται από το ποσοστό 66,66%, τα περισσότερα κείμενα με τις αντίστοιχες περιλήψεις τους κατηγοριοποιήθηκαν στο Topic\_3.

Πίνακας 4.8 : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους, για  $k = 4$ , Clinical Trial I.

FullText\Abstarct	Topic_0	Topic_1	Topic_2	Topic_3	Total
Topic_0	1 (20%)	1	1	2	<b>5</b>
Topic_1	0	0 (0%)	2	0	<b>2</b>
Topic_2	1	2	0 (0%)	1	<b>4</b>
Topic_3	1	0	1	4 (66,66%)	<b>6</b>
Total	<b>3</b>	<b>3</b>	<b>4</b>	<b>7</b>	<b>17</b>

Στην σύγκριση των δύο αυτών μοντέλων θα βοηθήσουν οι τιμές των παραμέτρων “document\_entropy”, “coherence”, “uniform\_dist” και “corpus\_dist” των LDA μοντέλων (εικόνες 4.36 και 4.84). Όπως και στα δύο προηγούμενα μοντέλων με αριθμούς θεμάτων 2 και 3, οι τιμές “document\_entropy” των περιλήψεων και των κειμένων κυμαίνονται στο ίδιο εύρος, οπότε ισχύει ότι και τα δύο είδη εγγράφων είναι εξίσου περίπλοκα. Επιπλέον, οι τιμές των θεμάτων της παραμέτρου “coherence” των κειμένων είναι μεγαλύτερες από τις τιμές της ίδιας παραμέτρου των περιλήψεων, άρα προτιμάται η αντιστοίχιση που έχει πραγματοποιηθεί σχετικά με θέματα-κείμενα παρά με θέματα-περιλήψεις. Το μοντέλο LDA των κειμένων έχει παρόμοια συχνότητα κατανομής θεμάτων σε όλα τα κείμενα με αυτό των περιλήψεων καθώς οι αποκλίσεις των τιμών της παραμέτρου “uniform\_dist” των κειμένων δεν απέχουν πολύ από τις τιμές των περιλήψεων. Άρα, έχει πραγματοποιηθεί σχεδόν ισάξια η κατανομή των 4 θεμάτων σε όλο το «σώμα» των κειμένων και σε όλο το «σώμα» των περιλήψεων. Τέλος, η παράμετρος “corpus\_dist”, που αφορά την κατανομή θεμάτων σε όλο το «σώμα» των κειμένων έχει μικρότερες τιμές από ότι αυτή των περιλήψεων καθώς όπως φαίνεται το κάθε θέμα που έχει αναδυθεί, είναι πιο σχετικό με το αντίστοιχο του κείμενο, συγκριτικά με το εκάστοτε θέμα που έχει ανατεθεί σε κάθε περίληψη.

#### 4.5.2 Clinical Trial II

Η κατηγορία αυτή αποτελείται από 38 κείμενα με τις αντίστοιχες περιλήψεις τους, που σημαίνει ότι το σύνολο των εγγράφων είναι σχεδόν το διπλάσιο από την προηγούμενη κατηγορία, οπότε οι τιμές που εξετάστηκαν προηγουμένως θα έχουν αυξηθεί. Στον ακόλουθο Πίνακα 4.9 συγκεντρώνονται οι τιμές των δύο αυτών παραμέτρων για τα 3 μοντέλα του σώματος Clinical Trial II.

Πίνακας 4.9: Τιμές παραμέτρων του μοντέλου των πλήρων κειμένων και των περιλήψεών τους, Clinical Trial II (L = Loglikelihood, P = Perplexity).

	<b>k = 2</b>	<b>k = 3</b>	<b>k = 4</b>
<b>Full Texts</b>	L = - 1477233.728 P = 1957.332	L = - 1489951.811 P = 1830.781	L = - 1495413.343 P = 1717.269
<b>Abstracts</b>	L = - 50269.419 P = 531.882	L = - 50222.594 P = 501.253	L = - 50507.018 P = 482.481

Ξεκινώντας, θα γίνει η σύγκριση των μοντέλων  $k = 2$  θεμάτων, μεταξύ των περιλήψεων και των πλήρων κειμένων. Στις εικόνες 4.40-41 και 4.88-89 η τιμή Loglikelihood είναι ίση με -50269.419 και η τιμή Perplexity ίση με 531.882 των περιλήψεων και -1477233.728 και 1957.332 των πλήρων κειμένων, αντιστοίχως. Όπως αποδεικνύεται και σε αυτή την περίπτωση, το μοντέλο των περιλήψεων είναι προτιμότερο από εκείνο των κειμένων, καθώς έχει την μεγαλύτερη τιμή Loglikelihood και την μικρότερη τιμή Perplexity.

Στον Πίνακα 4.10 γίνεται η σύγκριση ανάμεσα στις 3 επικρατέστερες λέξεις κάθε θέματος, όπως προέκυψαν από τις περιλήψεις αφενός και τα πλήρη κείμενα αφετέρου. Σε αυτά τα μοντέλα (εικόνες 4.43 και 4.91) παρατηρείται ότι τα Topic\_1 και των πλήρων κειμένων και των περιλήψεων είναι σχεδόν όμοια μεταξύ τους. Η μόνη διαφορετική λέξη είναι “studi” για τα κείμενα και η λέξη “group” για τις περιλήψεις.

Πίνακας 4.10: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 2$  (Clinical Trial II).

<b>Full Texts</b>			<b>Abstracts</b>		
	<b>Λέξεις</b>	<b>Βάρη</b>		<b>Λέξεις</b>	<b>Βάρη</b>
Topic_0	cognit	867	Topic_0	placebo	62
Topic_0	alzheim	789	Topic_0	clinic	61
Topic_0	diseas	744	Topic_0	month	53
Topic_1	studi	1529	Topic_1	diseas	89
Topic_1	alzheim	1206	Topic_1	alzheim	67
Topic_1	diseas	1092	Topic_1	group	64

Σε συνδυασμό με την παράμετρο “corpus\_dist”, γίνεται αντιληπτό ότι η μικρή τιμή της στο Topic\_1 των κειμένων υποδηλώνει ότι το θέμα είναι πιο εξειδικευμένο σε

σχέση με την τιμή που δίνεται στο Topic\_1 των περιλήψεων, που θεωρείται πιο γενικευμένο. Τα Topic\_0 και των δύο ειδών, όπως φαίνεται στις εικόνες 4.43 και 4.91 είναι τελείως διαφορετικά. Η ομοιότητα ενός από τα δύο θέματα εξυπηρετεί στην σύγκριση της κατηγοριοποίησης των εγγράφων. Σύμφωνα με τις εικόνες 4.44 και 4.92, τα 13 κείμενα από τα 38 και οι αντίστοιχες 13 περιλήψεις από τις 38 έχουν κατηγοριοποιηθεί στο Topic\_1.

Στον πίνακα 4.11 παρουσιάζεται η συμφωνία κατηγοριοποίησης άρθρων σε κοινό θέματα, με βάση τα πλήρη κείμενα αφενός και τις περιλήψεις τους αφετέρου. Στο Clinical Trial II σώμα, το σύνολο των κειμένων και των περιλήψεων αυξήθηκε σε 38 και γι' αυτό το λόγο και τα ποσοστά -των σύμφωνων θεμάτων- είναι υψηλότερα συγκριτικά με το προηγούμενο «σώμα» Clinical Trial I.

Πίνακας 4.11 : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους, για  $k = 2$ , Clinical Trial II.

Full Text \ Abstract	Topic_0	Topic_1	Total
Topic_0	6 (43%)	8	<b>14</b>
Topic_1	11	13 (54%)	<b>24</b>
Total	<b>17</b>	<b>21</b>	<b>38</b>

Συνεχίζοντας με την σύγκριση των LDA μοντέλων, η παράμετρος “document\_entropy”, έχει αυξηθεί κατά μία μονάδα συγκριτικά με την κατηγορία Clinical Trial I αλλά εξακολουθεί να έχει παρόμοιες τιμές στα θέματα των περιλήψεων και των κειμένων. Η αύξηση της παραμέτρου αυτής δηλώνει ότι τα έγγραφα της κατηγορίας Clinical Trial II είναι ακόμα πιο περίπλοκα. Οι τιμές της παραμέτρου “coherence” των κειμένων, όπως φαίνονται στην εικόνα 4.90, είναι μεγαλύτερες από τις τιμές της ίδιας παραμέτρου των περιλήψεων που εμφανίζονται στην εικόνα 4.42. Οι χαμηλές τιμές της “uniform\_dist” παραμέτρου των περιλήψεων δηλώνουν ότι τα θέματα δεν κατανομούνται με την ίδια συχνότητα στις περιλήψεις όπως γίνεται στο μοντέλο των κειμένων.

Εξετάζοντας το μοντέλο των  $k = 3$  θεμάτων συγκριτικά μεταξύ των περιλήψεων και των πλήρων κειμένων, τιμή Loglikelihood ισούται με -50222.594 και η τιμή Perplexity ισούται με 501.253 για τις περιλήψεις (εικόνες 4.45-4.46) ενώ είναι -



1489951.811 και 1830.781 για τα πλήρη κείμενα (εικόνες 4.93-4.94), αντιστοίχως. Άρα το μοντέλο 3 θεμάτων των περιλήψεων αποδεικνύεται πάλι το βέλτιστο.

Στις εικόνες 4.48 και 4.96 φαίνονται συγκριτικά οι 3 επικρατούσες λέξεις κάθε θέματος, όπως προέκυψαν από ανάλυση των περιλήψεων αφενός και των πλήρων κειμένων αφετέρου. Από τον Πίνακα 4.12 διαπιστώνεται ότι είναι τελείως διαφορετικές μεταξύ τους. Αφού αυξήθηκε ο αριθμός των θεμάτων και στα δύο είδη κειμένων, ο αλγόριθμος LDA έκανε εκτενέστερη αναζήτηση κρυμμένων πληροφοριών και για αυτόν τον λόγο προκύπτει η διαφοροποίηση των λέξεων.

Πίνακας 4.12: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 3$  (Clinical Trial II).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	test	535	Topic_0	cognit	54
Topic_0	brain	462	Topic_0	particip	49
Topic_0	age	383	Topic_0	improvement	40
Topic_1	patient	877	Topic_1	cognit	51
Topic_1	group	750	Topic_1	trial	32
Topic_1	alzheim	730	Topic_1	efficacy	26
Topic_2	study	1419	Topic_2	treatment	87
Topic_2	alzheim	1024	Topic_2	placebo	79
Topic_2	diseases	947	Topic_2	clinic	76

Έτσι, δεν μπορεί να πραγματοποιηθεί και η σύγκριση των κατηγοριοποιήσεων που εμφανίζονται αναλυτικά στις εικόνες 4.49 και 4.97. Όμως αξίζει να σχολιαστεί η συμφωνία θεμάτων που παρουσιάζεται στον πίνακα 4.13. Διαπιστώνεται ότι τα περισσότερα κείμενα με τις αντίστοιχες περιλήψεις τους έχουν κατηγοριοποιηθεί στο Topic\_2, με ποσοστό 63%. Η αύξηση του αριθμού των θεμάτων από  $k = 2$  σε  $k = 3$  επηρέασε την κατηγοριοποίηση των θεμάτων καθώς αυτά απέκτησαν μία πιο εξειδικευμένη μορφή με βάση τα «σώματα», όπως άλλωστε φαίνεται και στις λέξεις των θεμάτων τόσο των κειμένων όσο και των περιλήψεών τους.

Πίνακας 4.13 : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους για  $k = 3$ , Clinical Trial II.

Full Text\ Abstarct	Topic_0	Topic_1	Topic_2	Total
Topic_0	3 (33%)	4	2	<b>9</b>
Topic_1	4	3 (23%)	6	<b>13</b>
Topic_2	3	3	10 (63%)	<b>16</b>
Total	<b>10</b>	<b>10</b>	<b>18</b>	<b>38</b>

Οι τιμές της παραμέτρου “document\_entropy”, ενώ μεταβλήθηκαν ελάχιστα, η πολυπλοκότητα των κειμένων και των περιλήψεων παραμένει σταθερή. Η παράμετρος “coherence” των κειμένων και για τα 3 θέματα είναι μεγαλύτερη των αντίστοιχων θεμάτων των περιλήψεων, άρα το μοντέλο των κειμένων έχει καλύτερη αντιστοίχιση θέματος-κειμένου. Η κατανομή των θεμάτων σε όλα τα κείμενα έχει μεγαλύτερη ομοιομορφία από ότι στις περιλήψεις σύμφωνα με τις τιμές της παραμέτρου “uniform\_dist”. Τέλος, η παράμετρος “corpus\_dist” των περιλήψεων έχει υψηλότερες τιμές στο εκάστοτε θέμα αλλά αυτό υποδεικνύει ότι τα θέματα είναι πιο γενικά σε σχέση με τις περιλήψεις. (εικόνες 4.47 και 4.95)

Στα μοντέλα  $k = 4$  θεμάτων των περιλήψεων και των πλήρων κειμένων πρέπει να γίνει η απαραίτητη σύγκριση των τιμών Loglikelihood και Perplexity που εμφανίζονται στις εικόνες 4.50-51 και 4.98-99 ώστε να αποφασιστεί ποιο από τα δύο είναι το καταλληλότερο μοντέλο. Για τις περιλήψεις, η τιμή Loglikelihood είναι -50507.018 και η τιμή Perplexity είναι 482.481 ενώ για τα κείμενα η τιμή Loglikelihood ισούται με -1495413.343 και η τιμή Perplexity ισούται 1717.269. Άρα το μοντέλο των περιλήψεων είναι το βέλτιστο.

Συνεχίζοντας, στον Πίνακα 4.14 συγκρίνονται οι τρεις επικρατούσες λέξεις και τα βάρη τους (εικόνα 4.53 και 4.101), για τις περιλήψεις και τα πλήρη κείμενα. Είναι αντιληπτό ότι οι επικρατούσες λέξεις των θεμάτων των πλήρων κειμένων είναι παρόμοιες με αυτές των περιλήψεων αλλά και συγχρόνως διαφορετικές, διότι οι τα κείμενα παρέχουν περισσότερες πληροφορίες από ότι οι περιλήψεις, οπότε οι λέξεις των θεμάτων είναι πιο εξειδικευμένες.

Πίνακας 4.14: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 4$  (Clinical Trial II).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	test	558	Topic_0	diseas	32
Topic_0	cell	354	Topic_0	model	27
Topic_0	cognit	353	Topic_0	use	27
Topic_1	alzheim	638	Topic_1	cognit	66
Topic_1	diseas	547	Topic_1	particip	62
Topic_1	brain	522	Topic_1	group	36
Topic_2	studi	948	Topic_2	placebo	79
Topic_2	dose	873	Topic_2	studi	71
Topic_2	amyloid	633	Topic_2	clinic	71
Topic_3	alzheim	879	Topic_3	diseas	77
Topic_3	diseas	875	Topic_3	month	53
Topic_3	group	748	Topic_3	alzheim	51

Στον πίνακα 4.15, το μεγαλύτερο ποσοστό συμφωνίας θεμάτων των κειμένων και των περιλήψεών τους συγκεντρώνεται στο Topic\_2 (64%). Γενικότερα, παρατηρείται ότι με την αύξηση του αριθμού θεμάτων από  $k = 3$  σε  $k = 4$ , οι περισσότερες περιλήψεις έχουν κατηγοριοποιηθεί στο Topic\_2 και στο Topic\_3 συγκριτικά με τα άλλα δύο θέματα διότι οι περιλήψεις έχουν γραφτεί με αντικειμενικότητα και εμπεριέχουν όλες τις βασικές πληροφορίες των κειμένων, από τα οποία εξήχθησαν.

Πίνακας 4.15 : Ποσοστιαία συμφωνία θεμάτων των κειμένων και των περιλήψεων τους για  $k = 4$ , Clinical Trial II.

Full Text\ Abstract	Topic_0	Topic_1	Topic_2	Topic_3	Total
Topic_0	1 (20%)	2	1	1	<b>5</b>
Topic_1	4	2 (20%)	0	4	<b>10</b>
Topic_2	2	2	7 (64%)	0	<b>11</b>
Topic_3	0	1	9	2 (17%)	<b>12</b>
<b>Total</b>	<b>7</b>	<b>7</b>	<b>17</b>	<b>7</b>	<b>38</b>

Αυτό επιβεβαιώνεται και από τις τιμές της παραμέτρου “corpus\_dist” των κειμένων καθώς είναι μικρότερες από τις τιμές της παραμέτρου των περιλήψεων. Όπως αναφέρθηκε και στα προηγούμενα μοντέλα, οι τιμές της παραμέτρου “document\_entropy”, που καθορίζει την πολυπλοκότητα των κειμένων και των περιλήψεων παραμένουν στο ίδιο εύρος τιμών. Οι λιγότερο αρνητικές τιμές της παραμέτρου “coherence” δηλώνουν την καλύτερη αντιστοίχιση θεμάτων-κειμένων συγκριτικά με τις τιμές της παραμέτρου των περιλήψεων που είναι μικρότερες. Τέλος, η παράμετρος “uniform\_dist” των περιλήψεων έχει μικρότερες τιμές από αυτές των κειμένων, δηλαδή υπάρχει παρόμοια συχνότητα κατανομής θεμάτων σε όλα τα κείμενα σε σχέση με τις περιλήψεις (εικόνες 4.52 και 4.100).

### 4.5.3 Controlled Clinical Trial

Η κατηγορία Controlled Clinical Trial αποτελείται από το μεγαλύτερο σύνολο κειμένων και περιλήψεων (307 κείμενα και οι αντίστοιχες περιλήψεις). Γι’ αυτό το λόγο, οι τιμές Loglikelihood και Perplexity θα είναι πολύ μεγάλες συγκριτικά με τις δύο προηγούμενες κατηγορίες. Στον ακόλουθο Πίνακα 4.16 συγκεντρώνονται οι τιμές των δύο αυτών παραμέτρων για τα 3 μοντέλα του σώματος Controlled Clinical Trial.

Πίνακας 4.16: Τιμές παραμέτρων του μοντέλου των πλήρων κειμένων και των περιλήψεών τους, Controlled Clinical Trial (L = Loglikelihood, P = Perplexity).

	<b>k = 2</b>	<b>k = 3</b>	<b>k = 4</b>
<b>Full Texts</b>	L = - 11660725.770 P = 2505.411	L = - 11796788.272 P = 2354.580	L = - 11913865.771 P = 2275.635
<b>Abstracts</b>	L = - 372392.136 P = 821.082	L = - 374954.813 P = 769.141	L = - 374006.756 P = 717.697

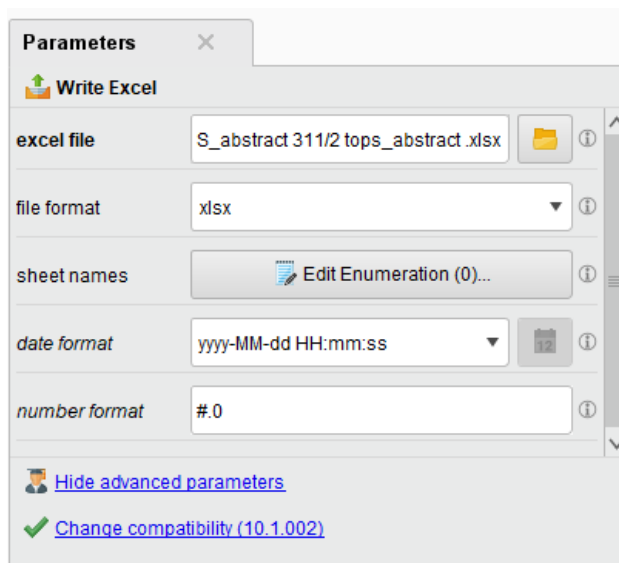
Ξεκινώντας από το μοντέλο των k = 2 θεμάτων, για τις περιλήψεις (εικόνες 4.56-57) η τιμή Loglikelihood είναι ίση με -372392.136 και η τιμή Perplexity είναι ίση με 821.082 και για τα κείμενα (εικόνες 4.104-105) η τιμή Loglikelihood ισούται με -11660725.770 και η τιμή Perplexity ισούται με 2505.411. Το μοντέλο 2 θεμάτων των περιλήψεων είναι καλύτερο από το μοντέλο των κειμένων διότι η τιμή Loglikelihood είναι η μεγαλύτερη και η τιμή Perplexity η μικρότερη.

Στις εικόνες 4.59 και 4.107 παρουσιάζονται οι 3 επικρατούσες λέξεις των  $k = 2$  θεμάτων. Όπως φαίνεται συγκριτικά στον Πίνακα 4.17, το Topic\_1 των κειμένων αντιστοιχεί στο Topic\_0 των περιλήψεων με την μοναδική διαφορά να είναι στις λέξεις “studi”, που είναι πιο συγκεκριμένη, σε σχέση με την λέξη “group”, που είναι πιο γενική. Έτσι, μπορεί να γίνει η σύγκριση των κατηγοριοποιήσεων των κειμένων και των περιλήψεων σε αυτά τα θέματα (εικόνες 4.60 και 4.92).

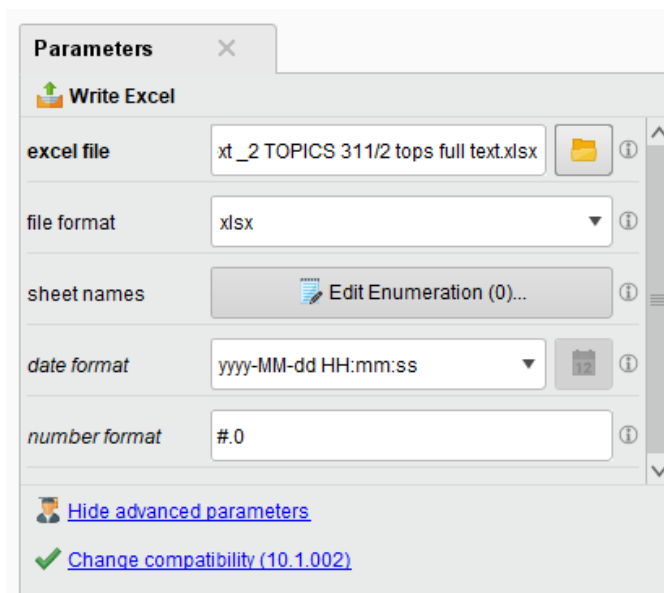
Πίνακας 4.17: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 2$  (Controlled Clinical Trial).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	studi	9253	Topic_0	group	480
Topic_0	alzheim	7682	Topic_0	intervent	405
Topic_0	cognit	7562	Topic_0	dementia	369
Topic_1	intervent	5918	Topic_1	cognit	518
Topic_1	studi	5822	Topic_1	placebo	395
Topic_1	dementia	5613	Topic_1	diseas	331

Επειδή το σύνολο των εγγράφων είναι αρκετά μεγάλο για να συγκριθούν ένα προς ένα, στο πρόγραμμα RapidMiner, συνδέεται η έξοδος “exa” του χειριστή “Extract Topics from Documents (LDA)” με την είσοδο “inp” του χειριστή “Write Excel” ώστε να εξαχθούν σε δύο αρχεία Excel οι κατηγοριοποιήσεις των κειμένων και των περιλήψεων και να μπορέσουν να συγκριθούν κάνοντας αναζήτηση στα αρχεία το Topic\_1 των κειμένων και το Topic\_0 των περιλήψεων. Στις παραμέτρους του χειριστή “Write Excel”, ονοματίζεται το αρχείο και πατώντας το διπλανό εικονίδιο του φακέλου αποθηκεύεται στον επιθυμητό φάκελο. Στις εικόνες 4.119 και 4.120 φαίνονται τα ονόματα των αρχείων.



Εικόνα 4.119: Ονομασία αρχείου Excel “2 tops\_abstract”.



Εικόνα 4.120: Ονομασία αρχείου Excel “2 tops full text”.

Στον Πίνακα 4.18.α εξετάζεται η κατηγοριοποίηση των άρθρων σε topics συγκριτικά, με βάση τις περιλήψεις και με βάση τα πλήρη κείμενα. Σύμφωνα με την διαγώνιο του Πίνακα 4.18.α, τα ποσοστά συμφωνίας των θεμάτων δεν είναι αρκετά υψηλά, καθώς οι πλειονότητα των κειμένων και των αντίστοιχων περιλήψεων τους έχουν κατηγοριοποιηθεί στο ίδιο θέμα, δηλαδή Topic\_0 για τις περιλήψεις και Topic\_1 για τα κείμενα. Οι εξαγόμενες περιλήψεις είναι αντιπροσωπευτικές των αντίστοιχων κειμένων τους και για αυτό τον λόγο βλέπουμε την υψηλότερη συγκέντρωση περιλήψεων (105/198, 53%). Αν όμως γίνει αμοιβαία αντιμετάθεση μεταξύ των

ονομάτων Topic\_0 και Topic\_1 των περιλήψεων, όπως υπονοούν οι επικρατούσες λέξεις (Πίνακας 4.17), τότε ο Πίνακας 4.18.α παίρνει τη μορφή του Πίνακα 4.18.β, οπότε αυξάνεται σημαντικά το ποσοστό κοινής κατηγοριοποίησης.

Πίνακας 4.18.α : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους για  $k = 2$ , Controlled Clinical Trial.

Full Text \ Abstarct	Topic_0	Topic_1	Total
Topic_0	42 (39%)	67	<b>109</b>
Topic_1	105	93 (47%)	<b>198</b>
Total	<b>147</b>	<b>160</b>	<b>307</b>

Πίνακας 4.18.β : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεων τους για  $k = 2$ , Controlled Clinical Trial (μετά από αντιμετάθεση Topic\_0 και Topic\_1).

Full Text \ Abstarct	Topic_1	Topic_0	Total
Topic_0	67 (62.5%)	42	<b>109</b>
Topic_1	93	105 (53.0%)	<b>198</b>
Total	<b>160</b>	<b>147</b>	<b>307</b>

Στις εικόνες 4.58 και 4.106 οι τιμές της παραμέτρου “document\_entropy” και των δύο ειδών εγγράφων, έχουν αυξηθεί καθώς αυξήθηκε ο συνολικός αριθμός των κειμένων και των περιλήψεων και άρα η περιπλοκότητά τους. Συνεχίζοντας με την παράμετρο “coherence”, οι τιμές της στο μοντέλο των κειμένων είναι μεγαλύτερες από τις τιμές του μοντέλου των περιλήψεων, συνεπώς έχουν αντιστοιχηθεί καταλληλότερα τα θέματα στα κείμενα από ότι στις περιλήψεις. Οι τιμές της “uniform\_dist” των κειμένων είναι σχεδόν οι διπλάσιες των τιμών των περιλήψεων, με αυτό να υποδηλώνεται ότι στο μοντέλο 2 θεμάτων των κειμένων τα θέματα έχουν ανατεθεί με παρόμοια συχνότητα σε όλα τα κείμενα. Τέλος, η παράμετρος “corpus\_dist” τόσο των κειμένων όσο και των περιλήψεων έχουν παρόμοιες τιμές. Οι χαμηλές τιμές και στα δύο μοντέλα σημαίνουν ότι τα θέματα που έχουν εξαχθεί είναι εξειδικευμένα και κατανέμονται πιο συγκεκριμένα πάνω στα κείμενα και στις περιλήψεις.

Στο μοντέλο των  $k = 3$  θεμάτων, για τις περιλήψεις η τιμή της Loglikelihood είναι ίση με -374954.813 και η τιμή Perplexity ίση με 769.141 (εικόνες 4.61-62) ενώ στα πλήρη κείμενα η τιμή Loglikelihood ισούται με -11796788.272 και η τιμή Perplexity ισούται με 2354.580 (εικόνες 4.109-110). Για ακόμη μία φορά αποδεικνύεται ότι το μοντέλο  $k = 3$  θεμάτων των περιλήψεων είναι καλύτερο από το μοντέλο των κειμένων.

Οι τρεις πιο συχνές λέξεις και τα βάρη τους που εμφανίζονται στις εικόνες 4.64 και 4.112 και συγκριτικά στον Πίνακα 4.19. Διαπιστώνεται ότι το Topic\_2 των κειμένων είναι αρκετά παρόμοιο με το Topic\_0 των περιλήψεων όπου διαφοροποιούνται στις λέξεις “studi” και “particip” αντίστοιχα.

Πίνακας 4.19: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 3$  (Controlled Clinical Trial).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	dementia	4791	Topic_0	cognit	468
Topic_0	care	4487	Topic_0	group	306
Topic_0	intervent	4189	Topic_0	particip	261
Topic_1	studi	6722	Topic_1	placebo	434
Topic_1	alzheim	6283	Topic_1	cognit	330
Topic_1	diseas	5639	Topic_1	treatment	296
Topic_2	cognit	9016	Topic_2	group	219
Topic_2	group	5323	Topic_2	dementia	206
Topic_2	studi	4285	Topic_2	caregiv	197

Ακολουθώντας την διαδικασία που αναλύθηκε παραπάνω, δημιουργούνται δύο αρχεία Excel με ονόματα “3 tops\_abstract” και “3 tops\_full text”. Η εύρεση του συνδυασμού Topic\_2 των κειμένων με το Topic\_0 των περιλήψεων ανέδειξε ότι 5 κείμενα και οι 5 -αντίστοιχες- περιλήψεις τους κατανεμήθηκαν στα παρόμοια θέματα, όπως παρουσιάζεται στον Πίνακα 4.20. Το σύνολο των κειμένων και των περιλήψεών τους αντίστοιχα έχει αυξηθεί στα 307 και με την αύξηση του αριθμού θεμάτων από  $k = 2$  σε  $k = 3$ , παρατηρείται μία γενικότερη συνολική ομοιόμορφη κατανομή τόσο των κειμένων όσο και των περιλήψεων, όπως παρουσιάζεται στην τελευταία γραμμή και τελευταία στήλη του Πίνακα 4.20. Το μεγαλύτερο ποσοστό συμφωνίας είναι το 75% του Topic\_0, όπου 61 άρθρα κατηγοριοποιήθηκαν στο ίδιο θέμα είτε με βάση τις περιλήψεις είτε με βάση τα πλήρη κείμενά τους.



Πίνακας 4.20 : Ποσοστιαία συμφωνία θεμάτων των πλήρων κειμένων και των περιλήψεών τους, για  $k = 3$ , Controlled Clinical Trial.

Full Text\ Abstract	Topic_0	Topic_1	Topic_2	Total
Topic_0	61 (75%)	2	18	<b>81</b>
Topic_1	18	38 (33%)	58	<b>114</b>
Topic_2	5	86	21 (19%)	<b>112</b>
Total	<b>84</b>	<b>126</b>	<b>97</b>	<b>307</b>

Η αύξηση του πλήθους των θεμάτων σε  $k = 3$  ευθύνεται για τον μικρό αριθμό των κειμένων και των περιλήψεών τους καθώς ο αλγόριθμος LDA εμβάθυνε περισσότερο μέσα στα έγγραφα και εξήγαγε περισσότερες πληροφορίες, οι οποίες επιβεβαιώνονται και από τις επικρατούσες λέξεις των θεμάτων.

Οι τιμές της παραμέτρου “document\_entropy” αυξήθηκε κι άλλο με την αύξηση του αριθμού των θεμάτων και στις δύο περιπτώσεις εγγράφων, δηλαδή τα κείμενα και οι περιλήψεις τους γίνονται ακόμα πιο περίπλοκα. Η παράμετρος “coherence” των κειμένων παραμένει λιγότερο αρνητική από την τιμή της παραμέτρου των περιλήψεων. Οι μεγάλες τιμές της “uniform\_dist” του μοντέλου των κειμένων υποδηλώνει ότι έχουν κατανομηθεί ομοιόμορφα τα θέματα στα κείμενα σε σχέση με τις περιλήψεις. Η παράμετρος “corpus\_dist” και των δύο ειδών εγγράφων έχει σχεδόν τις ίδιες, χαμηλές τιμές και αυτό συνεπάγεται με το γεγονός ότι τα θέματα που έχουν αναδυθεί είναι εξειδικευμένα πάνω στα κείμενα και στις περιλήψεις. (εικόνες 4.63 και 4.111)

Προχωρώντας σε  $k = 4$  θέματα, συγκρίνοντας τις εικόνες 4.66-4.67 και 4.114-4.115 διαπιστώνεται ότι ο συνδυασμός των τιμών Loglikelihood (-37406.6) και Perplexity (717.697) των περιλήψεων είναι καλύτερος από τον αντίστοιχο συνδυασμό τιμών Loglikelihood = -11913865.771 και Perplexity = 2275.635 των πλήρων κειμένων.

Επιπλέον, στα μοντέλα των  $k = 4$  παρατηρούνται περισσότερα παρόμοια θέματα μεταξύ των πλήρων κειμένων και των περιλήψεων (εικόνες 4.69 και 4.117). Όπως φαίνεται στον Πίνακα 4.21, το Topic\_2 των περιλήψεων σχετίζεται με το Topic\_3 των πλήρων κειμένων και το Topic\_3 των περιλήψεων σχετίζεται με το Topic\_2 των πλήρων κειμένων. Με τον ίδιο τρόπο φαίνεται το Topic\_0 των περιλήψεων να έχει κοινά με το Topic\_1 των πλήρων κειμένων και αντίστροφα το Topic\_1 των περιλήψεων

να έχει κοινά με το Topic\_0 των πλήρων κειμένων. Επειδή η αρίθμηση των topics σε Topic\_0, \_1, \_2, \_3 γίνεται από τον αλγόριθμο LDA με τυχαίο τρόπο, η ταυτότητα κάθε topic καθορίζεται κυρίως από τις 3 επικρατούσες λέξεις του που δηλώνουν το περιεχόμενό του, και όχι από τον αριθμό του που δίνεται τυχαία.

Πίνακας 4.21: Σύγκριση των 3 επικρατέστερων λέξεων ανά θέμα, μεταξύ πλήρων κειμένων και περιλήψεων, στο μοντέλο  $k = 4$  (Controlled Clinical Trial).

Full Texts			Abstracts		
	Λέξεις	Βάρη		Λέξεις	Βάρη
Topic_0	alzheim	5747	Topic_0	intervent	266
Topic_0	studi	5002	Topic_0	dementia	214
Topic_0	diseas	4646	Topic_0	group	198
Topic_1	care	4487	Topic_1	cognit	153
Topic_1	dementia	4040	Topic_1	level	125
Topic_1	caregiv	3639	Topic_1	amyloid	125
Topic_2	studi	3640	Topic_2	cognit	544
Topic_2	patient	3024	Topic_2	group	247
Topic_2	treatment	2459	Topic_2	function	209
Topic_3	cognit	7062	Topic_3	placebo	386
Topic_3	group	4947	Topic_3	treatment	337
Topic_3	anuscript	3984	Topic_3	patient	299

Βάσει της παραπάνω διαδικασίας που υλοποιήθηκε και εξάγοντας δύο αρχεία Excel με ονόματα “4 tops\_abstarct” και “4 tops full text” προκύπτει ότι 46 κείμενα και οι αντίστοιχες περιλήψεις τους ανήκουν στα Topic\_2 των περιλήψεων και Topic\_3 των κειμένων και 32 κείμενα με τις περιλήψεις του κατηγοριοποιήθηκαν στο ίδιο θέμα, δηλαδή Topic\_3 των περιλήψεων και Topic\_2 των κειμένων.

Όπως φαίνεται στον Πίνακα 4.22.α, η ποσοστιαία συμφωνία θεμάτων για  $k = 4$  είναι πολύ μικρή. Το μεγάλο σύνολο κειμένων σε συνδυασμό με την αύξηση του αριθμού θεμάτων από  $k = 3$  σε  $k = 4$ , δείχνει ότι τόσο οι λέξεις των θεμάτων των κειμένων όσο και των περιλήψεων είναι πολύ εξειδικευμένες στο εκάστοτε «σώμα». Αν παρόλα αυτά παρατηρήσουμε προσεκτικά τα 4 θέματα, διαπιστώνεται ότι η αμοιβαία αντιμετάθεση των topic\_0 και topic\_1 καθώς και των topic\_2 και topic\_3, ώστε να συμφωνούν και με τις επικρατούσες λέξεις τους, οδηγεί σε σημαντική αύξηση

του ποσοστού κοινής κατηγοριοποίησης των άρθρων (σύγκριση των διαγώνιων τιμών Πίνακα 4.22.α και Πίνακα 4.22.β).

Πίνακας 4.22.α : Ποσοστιαία συμφωνία θεμάτων των κειμένων και των περιλήψεων τους για  $k = 4$ , Controlled Clinical Trial.

Full Text \ Abstract	Topic_0	Topic_1	Topic_2	Topic_3	Total
Topic_0	2 (2%)	22	26	47	<b>97</b>
Topic_1	51	1 (2%)	0	6	<b>58</b>
Topic_2	2	16	20 (28%)	32	<b>70</b>
Topic_3	17	14	46	5 (6%)	<b>82</b>
Total	<b>72</b>	<b>53</b>	<b>92</b>	<b>90</b>	<b>307</b>

Πίνακας 4.22.β : Ποσοστιαία συμφωνία θεμάτων των κειμένων και των περιλήψεων τους για  $k = 4$ , Controlled Clinical Trial – μετά από αντιμετάθεση των topics σύμφωνα με τις επικρατούσες λέξεις.

Full Text \ Abstract	Topic_1	Topic_0	Topic_3	Topic_2	Total
Topic_0	22 (22.7%)	2	47	26	<b>97</b>
Topic_1	1	51 (88.0%)	6	0	<b>58</b>
Topic_2	16	2	32 (45.7%)	20	<b>70</b>
Topic_3	14	17	5	46 (56.1%)	<b>82</b>
Total	<b>53</b>	<b>72</b>	<b>90</b>	<b>92</b>	<b>307</b>

Στα μοντέλα LDA των εικόνων 4.68 και 4.116 η παράμετρος “document\_entropy” των κειμένων και των περιλήψεων παρέμεινε σχεδόν στις ίδιες τιμές με τα μοντέλα 3 θεμάτων. Αυτό σημαίνει ότι η περιπλοκότητα των εγγράφων παρέμεινε σταθερή. Οι τιμές της παραμέτρου “coherence” των κειμένων έχουν κάποιες παραπλήσιες τιμές με αυτές των περιλήψεων αλλά παραμένει καλύτερη η αντιστοίχιση θεμάτων-κειμένων. Επιπλέον, οι τιμές της παραμέτρου “uniform\_dist” των περιλήψεων είναι μικρές συγκριτικά με τις τιμές της παραμέτρου των κειμένων, άρα τα κείμενα έχουν καλύτερη

ομοιόμορφη κατανομή θεμάτων. Τέλος, οι τιμές της παραμέτρου “corpus\_dist” των κειμένων είναι μικρότερες από τις αντίστοιχες τιμές των περιλήψεων, οπότε τα θέματα των κειμένων καταλήγουν να έχουν ειδικευτεί πάνω στα κείμενα.

## ΚΕΦΑΛΑΙΟ 5: Συμπεράσματα – Περαιτέρω έρευνα

Φτάνοντας στο σημείο της ολοκλήρωσης της παρούσας διπλωματικής εργασίας, μπορούν να διατυπωθούν τα εξής συμπεράσματα:

- Ο αλγόριθμος LDA κατάφερε με αρκετή επιτυχία να προ-επεξεργαστεί τα έγγραφα και να τα κατατάξει σε κατηγορίες.
- Σύμφωνα με την ανάλυση που προηγήθηκε, ένας μεγάλος αριθμός θεμάτων είναι καλύτερος από έναν μικρότερο καθώς ο αλγόριθμος εξορύσσει παραπάνω πληροφορίες.
- Όπως φάνηκε και στις εικόνες των αποτελεσμάτων στις ενότητες 4.2, 4.3, 4.4 και 4.5, οι τιμές των παραμέτρων των μοντέλων LDA των κειμένων και αντίστοιχα των περιλήψεων δεν είχαν πολύ μεγάλες αριθμητικές αποκλίσεις.
- Οι παράμετροι που επιλέχθηκαν να σχολιαστούν, θεωρήθηκαν οι καταλληλότερες για την απάντηση στο ερώτημα που έχει τεθεί.

Συμπερασματικά, με βάση την παραπάνω σύγκριση και ανάλυση των αποτελεσμάτων των πειραμάτων που διεξάχθηκαν στο Κεφάλαιο 4, μπορεί να υποστηριχθεί ότι η απάντηση είναι θετική, δηλαδή ότι είναι επαρκές να αναγνωσθούν οι περιλήψεις των επιστημονικών άρθρων βιοϊατρικού περιεχομένου από τους ερευνητές και να αντλήσουν τις βασικές πληροφορίες διαλογής και κατηγοριοποίησης των άρθρων. Αυτό το συμπέρασμα βέβαια αναδεικνύει και τον κρίσιμο ρόλο που έχουν οι περιλήψεις, και την ανάγκη της πληρότητας και αντιπροσωπευτικότητας κατά την συγγραφή των περιλήψεων, ώστε να μην παραλειφθεί καμία σημαντική πληροφορία του κειμένου. Και φυσικά, μετά την διαλογή, εννοείται ότι ο ερευνητής που επιθυμεί να εμβαθύνει περισσότερο σε ένα βιοϊατρικό θέμα, θα πρέπει να έχει την δυνατότητα να ανατρέξει στο πλήρες κείμενο του άρθρου, ώστε να ενημερωθεί.

Η δομημένη διεργασία που δημιουργήθηκε στο RapidMiner, η οποία χρησιμοποίησε τον αλγόριθμο LDA για την εξαγωγή θεμάτων από βιοϊατρικά κείμενα και από τις περιλήψεις τους, έχει την δυναμική να εξελιχθεί περαιτέρω. Όσον αφορά τον βιοϊατρικό τομέα, εκτός από τις περιλήψεις βιοϊατρικών κειμένων που σχετίζονται με την νόσο Alzheimer, μπορούν να εισαχθούν και περιλήψεις και από άλλες νευροεκφυλιστικές νόσους, όπως είναι η νόσος Parkinson, η νόσος Huntington, η Αμυοτροφική Πλάγια Σκλήρυνση (ALS). κ.ά. Συγχρόνως, η χρήση της γλώσσας προγραμματισμού Python μπορεί να χρησιμοποιηθεί για την επεξεργασία των ακτινογραφιών CT του εγκεφάλου, παρέχοντας λεπτομερείς δομικές πληροφορίες που είναι ουσιώδεις για την διάγνωση και την παρακολούθηση. Ο συνδυασμός

της ανάλυσης των κειμενικών δεδομένων με τα δεδομένα νευροεικονογραφίας μπορεί να παράσχει μια ολοκληρωμένη άποψη για τις νευροεκφυλιστικές νόσους, βοηθώντας τους ιατρικούς ερευνητές, τους επαγγελματίες υγείας και τους κλινικούς στην καλύτερη κατανόηση αυτών των πολύπλοκων νοσημάτων. Η προσέγγιση αυτή επιτρέπει όχι μόνο την ανακάλυψη νέας γνώσης, αλλά συμβάλλει σημαντικά στις προσπάθειες της ιατρικής κοινότητας για την δημιουργία ή/και εξέλιξη των τρόπων διάγνωσης, της πρόληψης και των θεραπειών των νευροεκφυλιστικών νόσων.

Όπως αναφέρθηκε στην ενότητα 2.2, ο χειριστής “Filter Stopwords (English)” και γενικότερα το πρόγραμμα RapidMiner, περιορίζουν την προ-επεξεργασία κειμένων σε συγκεκριμένες γλώσσες. Αυτό αποτελεί ένα βασικό εμπόδιο για την επιστημονική κοινότητα καθώς οι επιστήμονες, οι μελετητές και το ενδιαφερόμενο ευρύ κοινό δεν έχουν την δυνατότητα να εισάγουν κείμενα της μητρικής τους γλώσσας. Ειδικότερα, η γλώσσα των Ελληνικών θα μας ενδιέφερε να περιλαμβάνεται στο πρόγραμμα. Η ελληνική επιστημονική κοινότητα -διαρκώς- διεξάγει σημαντικές έρευνες σε πολλούς επιστημονικούς τομείς, οι οποίες είναι άξιες προς περαιτέρω επεξεργασία και ανάλυση. Η εισαγωγή της Ελληνικής Γλώσσας στο λογισμικό RapidMiner θα αποτελέσει ένα σημαντικό σημείο λόγω του γλωσσικού της πλούτου.

Εν κατακλείδι, η εξέλιξη και η εύκολη προσαρμογή της κάθε δομημένης διεργασίας εντός του RapidMiner σε αλλαγές ανοίγει τον δρόμο για πολλές άλλες καινοτόμες ιδέες, που θα βάλουν τα θεμέλια για ποικίλες εφαρμογές και λύσεις.

## Βιβλιογραφία – Αναφορές - Διαδικτυακές Πηγές

- [1.] Twin, A. (2023, April 15). What Is Data Mining? How It Works, Benefits, Techniques, and Examples. Investopedia.  
Ανάκτηση από: <https://www.investopedia.com/terms/d/datamining.asp>
- [2.] Stedman, C., & Hughes, A. (n.d.). Data Mining. TechTarget.  
Ανάκτηση από: <https://www.techtarget.com/searchbusinessanalytics/definition/data-mining#:~:text=Data%20mining%20is%20the%20process,make%20more%2Dinform ed%20business%20decisions>
- [3.] Lutkevich, B., & Burns, E. (n.d.). Natural Language Processing (NLP). TechTarget.  
Ανάκτηση από: <https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP>
- [4.] Nadif, M., & Role, F. (2021). Unsupervised and Self-supervised Deep Learning Approaches for Biomedical Text Mining.
- [5.] Kulshrestha, R. (2019, July 19). A Beginner's Guide to Latent Dirichlet Allocation (LDA). Towards Data Science.  
Ανάκτηση από: <https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2>
- [6.] Naushan, H. (2020). Topic Modeling with Latent Dirichlet Allocation. Towards Data Science.  
Ανάκτηση από: <https://towardsdatascience.com/topic-modeling-with-latent-dirichlet-allocation-e7ff75290f8>
- [7.] Zankadi, H. (2023). Topic Modeling with Latent Dirichlet Allocation (LDA) using Gensim and NLP techniques (Part I). Medium.  
Ανάκτηση από: <https://medium.com/@hajar.zankadi/using-latent-dirichlet-allocation-lda-and-nlp-techniques-to-predict-interest-tags-from-tweets-e30e9e4d83ec>
- [8.] European Commission. (n.d.). Causes of Climate Change. Ανάκτηση από: [https://climate.ec.europa.eu/climate-change/causes-climate-change\\_en](https://climate.ec.europa.eu/climate-change/causes-climate-change_en)
- [9.] United Nations. (n.d.). Take Action for the Sustainable Development Goals. Ανάκτηση από: <https://www.un.org/en/actnow/ten-actions>
- [10.] NASA. (n.d.). Climate Change: Vital Signs of the Planet. Effects of Climate Change.  
Ανάκτηση από: <https://climate.nasa.gov/effects/>
- [11.] World Health Organization. (n.d.). Climate Change and Health.

- Ανάκτηση από: <https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health>
- [12.] World Wide Fund for Nature (WWF). (n.d.). Climate Change and Energy.  
Ανάκτηση από: <https://www.wwf.org.uk/what-we-do/climate-change-and-energy>
- [13.] Progressive Insurance. (n.d.). The Future of Electric Cars.  
Ανάκτηση από: <https://www.progressive.com/lifelanes/on-the-road/future-of-electric-cars/>
- [14.] Hawley, D. (2022, September 21). How Long Do Electric Car Batteries Last? JD Power.  
Ανάκτηση από: <https://www.jdpower.com/cars/shopping-guides/how-long-do-electric-car-batteries-last>
- [15.] Threewitt, C. (n.d.). Gas-powered vs. Electric Cars: Which Is Faster? HowStuffWorks.  
Ανάκτηση από: <https://auto.howstuffworks.com/gas-powered-vs-electric-cars-which-is-faster.htm>
- [16.] Conserve Energy Future. (n.d.). Advantages and Disadvantages of Electric Cars.  
Ανάκτηση από: <https://www.conserve-energy-future.com/advantages-and-disadvantages-of-electric-cars.php>
- [17.] Barry, K. (2023, March 2). Why Hybrid Vehicles Are a Smart Choice Right Now. Consumer Reports.  
Ανάκτηση από: <https://www.consumerreports.org/cars/hybrids-evs/why-hybrid-vehicles-are-a-smart-choice-right-now-a2736240282/>
- [18.] Homes & Gardens. (n.d.). When to Start Planting Flowers for Spring.  
Ανάκτηση από: <https://www.homesandgardens.com/gardens/when-to-start-planting-flowers-for-spring>
- [19.] SanSone, A. E. (2020, May 29). How to Plant Flowers in Your Garden. House Beautiful.  
Ανάκτηση από: <https://www.housebeautiful.com/lifestyle/gardening/a32702349/how-to-plant-flowers/>
- [20.] Longfield Gardens. (n.d.). New to Flower Gardening? Here's Where to Start.  
Ανάκτηση από: <https://www.longfield-gardens.com/article/new-to-flower-gardening-heres-where-to-start>
- [21.] Wolfe, M. (2022, June 24). How do you plant flowers? Here are the steps to cultivating a beautiful backyard flower garden. Happy Sprout.



Ανάκτηση από: <https://www.happysprout.com/gardening/start-backyard-flower-garden/>

[22.] Ebert, J. (2022, July 15). How to plant a flower bed — the essential guide to planting beautiful flowers. Homes & Gardens.

Ανάκτηση από: <https://www.homesandgardens.com/advice/how-to-plant-a-flower-bed>

## ΠΑΡΑΡΤΗΜΑ Ι Άρθρα της ομάδας Clinical Trial I

- [1.] Wang S, Mustafa M, Yuede CM, Salazar SV, Kong P, Long H, Ward M, Siddiqui O, Paul R, Gilfillan S, Ibrahim A, Rhinn H, Tassi I, Rosenthal A, Schwabe T, Colonna M. Anti-human TREM2 induces microglia proliferation and reduces pathology in an Alzheimer's disease model. *J Exp Med.* 2020 Sep 7;217(9):e20200785. doi: 10.1084/jem.20200785. PMID: 32579671; PMCID: PMC7478730.
- [2.] Kim HJ, Cho KR, Jang H, Lee NK, Jung YH, Kim JP, Lee JI, Chang JW, Park S, Kim ST, Moon SW, Seo SW, Choi SJ, Na DL. Intracerebroventricular injection of human umbilical cord blood mesenchymal stem cells in patients with Alzheimer's disease dementia: a phase I clinical trial. *Alzheimers Res Ther.* 2021 Sep 14;13(1):154. doi: 10.1186/s13195-021-00897-2. PMID: 34521461; PMCID: PMC8439008.
- [3.] Kandadi Muralidharan K, Tong X, Kowalski KG, Rajagovindan R, Lin L, Budd Haberlain S, Nestorov I. Population pharmacokinetics and standard uptake value ratio of aducanumab, an amyloid plaque-removing agent, in patients with Alzheimer's disease. *CPT Pharmacometrics Syst Pharmacol.* 2022 Jan;11(1):7-19. doi: 10.1002/psp4.12728. Epub 2021 Oct 26. PMID: 34697913; PMCID: PMC8752104.
- [4.] Vissers MFJM, Heuberger JAAC, Groeneveld GJ, Oude Nijhuis J, De Deyn PP, Hadi S, Harris J, Tsai RM, Cruz-Herranz A, Huang F, Tong V, Erickson R, Zhu Y, Scearce-Levie K, Hsiao-Nakamoto J, Tang X, Chang M, Fox BM, Estrada AA, Pomponio RJ, Alonso-Alonso M, Zilberstein M, Atassi N, Troyer MD, Ho C. Safety, pharmacokinetics and target engagement of novel RIPK1 inhibitor SAR443060 (DNL747) for neurodegenerative disorders: Randomized, placebo-controlled, double-blind phase I/Ib studies in healthy subjects and patients. *Clin Transl Sci.* 2022 Aug;15(8):2010-2023. doi: 10.1111/cts.13317. Epub 2022 Jun 1. PMID: 35649245; PMCID: PMC9372423.
- [5.] Vidoni ED, Choi IY, Lee P, Reed G, Zhang N, Pleen J, Mahnken JD, Clutton J, Becker A, Sherry E, Bothwell R, Anderson H, Harris RA, Brooks W, Wilkins HM, Mosconi L, Burns JM, Swerdlow RH. Safety and target engagement profile of two oxaloacetate doses in Alzheimer's patients. *Alzheimers Dement.* 2021 Jan;17(1):7-17. doi: 10.1002/alz.12156. Epub 2020 Jul 27. PMID: 32715609; PMCID: PMC8084114.
- [6.] Hua X, Church K, Walker W, L'Hostis P, Viardot G, Danjou P, Hendrix S, Moebius HJ. Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of the Positive Modulator of HGF/MET, Fosgonimeton, in Healthy Volunteers and Subjects with Alzheimer's Disease: Randomized, Placebo-Controlled, Double-Blind, Phase I Clinical Trial. *J Alzheimers Dis.* 2022;86(3):1399-1413. doi: 10.3233/JAD-215511. PMID: 35180125; PMCID: PMC9108585.
- [7.] Brody M, Agronin M, Herskowitz BJ, Bookheimer SY, Small GW, Hitchinson B, Ramdas K, Wishard T, McInerney KF, Vellas B, Sierra F, Jiang Z, McClain-Moss L, Perez C, Fuquay A, Rodriguez S, Hare JM, Oliva AA Jr, Baumel B. Results and insights from a phase I clinical trial of Lomecel-B for Alzheimer's disease. *Alzheimers Dement.* 2023 Jan;19(1):261-273. doi: 10.1002/alz.12651. Epub 2022 Mar 31. PMID: 35357079; PMCID: PMC10084163.

- [8.] Castle MJ, Baltanás FC, Kovacs I, Nagahara AH, Barba D, Tuszynski MH. Postmortem Analysis in a Clinical Trial of AAV2-NGF Gene Therapy for Alzheimer's Disease Identifies a Need for Improved Vector Delivery. *Hum Gene Ther.* 2020 Apr;31(7-8):415-422. doi: 10.1089/hum.2019.367. Epub 2020 Mar 30. PMID: 32126838; PMCID: PMC7194314.
- [9.] Valenzuela M, Duncan T, Abey A, Johnson A, Boulamatsis C, Dalton MA, Jacobson E, Brunel L, Child G, Simpson D, Buckland M, Lowe A, Siette J, Westbrook F, McGreevy P. Autologous skin-derived neural precursor cell therapy reverses canine Alzheimer dementia-like syndrome in a proof of concept veterinary trial. *Stem Cell Res Ther.* 2022 Jun 17;13(1):261. doi: 10.1186/s13287-022-02933-w. PMID: 35715872; PMCID: PMC9205057.
- [10.] Family N, Maillet EL, Williams LTJ, Krediet E, Carhart-Harris RL, Williams TM, Nichols CD, Goble DJ, Raz S. Safety, tolerability, pharmacokinetics, and pharmacodynamics of low dose lysergic acid diethylamide (LSD) in healthy older volunteers. *Psychopharmacology (Berl).* 2020 Mar;237(3):841-853. doi: 10.1007/s00213-019-05417-7. Epub 2019 Dec 18. PMID: 31853557; PMCID: PMC7036065.
- [11.] Breier A, Brannan SK, Paul SM, Miller AC. Evidence of trospium's ability to mitigate cholinergic adverse events related to xanomeline: phase 1 study results. *Psychopharmacology (Berl).* 2023 May;240(5):1191-1198. doi: 10.1007/s00213-023-06362-2. Epub 2023 Apr 10. PMID: 37036495; PMCID: PMC10102054.
- [12.] Yoshida K, Moein A, Bittner T, Ostrowitzki S, Lin H, Honigberg L, Jin JY, Quartino A. Pharmacokinetics and pharmacodynamic effect of crenezumab on plasma and cerebrospinal fluid beta-amyloid in patients with mild-to-moderate Alzheimer's disease. *Alzheimers Res Ther.* 2020 Jan 22;12(1):16. doi: 10.1186/s13195-020-0580-2. PMID: 31969177; PMCID: PMC6977279.
- [13.] Sheng L, Cao W, Lin P, Chen W, Xu H, Zhong C, Yuan F, Chen H, Li H, Liu C, Yang M, Li X. Safety, Tolerability and Pharmacokinetics of Single and Multiple Ascending Doses of Benfotiamine in Healthy Subjects. *Drug Des Devel Ther.* 2021 Mar 9; 15:1101-1110. doi: 10.2147/DDDT.S296197. PMID: 33727798; PMCID: PMC7955752.
- [14.] Portron A, Jordan P, Draper K, Muenzer C, Dickerson D, van Iersel T, Hofmann C. A Phase I Study to Assess the Effect of Speed of Injection on Pain, Tolerability, and Pharmacokinetics After High-volume Subcutaneous Administration of Gantenerumab in Healthy Volunteers. *Clin Ther.* 2020 Jan;42(1):108-120.e1. doi: 10.1016/j.clinthera.2019.11.015. Epub 2019 Dec 26. PMID: 31883703.
- [15.] Rafii MS, Sol O, Mobley WC, Delpretti S, Skotko BG, Burke AD, Sabbagh MN, Yuan SH, Rissman RA, Pulsifer M, Evans C, Evans AC, Beth G, Fournier N, Gray JA, Dos Santos AM, Hliva V, Vukicevic M, Kosco-Vilbois M, Streffer J, Pfeifer A, Feldman HH. Safety, Tolerability, and Immunogenicity of the ACI-24 Vaccine in Adults With Down Syndrome: A Phase 1b Randomized Clinical Trial. *JAMA Neurol.* 2022 Jun 1;79(6):565-574. doi: 10.1001/jamaneurol.2022.0983. PMID: 35532913; PMCID: PMC9086937.
- [16.] Wang Y, Hernandez G, Mack WJ, Schneider LS, Yin F, Brinton RD. Retrospective analysis of phytoSERM for management of menopause-associated vasomotor symptoms and cognitive decline: a pilot study on pharmacogenomic effects of mitochondrial haplogroup and APOE genotype on therapeutic

efficacy. Menopause. 2020 Jan;27(1):57-65. doi: 10.1097/GME.0000000000001418. PMID: 31567873; PMCID: PMC7100617.

- [17.] Qian H, Yu C, Zhu H, Ding Q, Cai Y, Jing J, Xu X, Guo R, Zhang H, Liu H, Chen X, Liu Y. Safety, tolerability, and pharmacokinetics of fluoropezil (DC20), a novel acetylcholinesterase inhibitor: A phase I study in healthy young and elderly Chinese subjects. Clin Transl Sci. 2023 May;16(5):810-822. doi: 10.1111/cts.13490. Epub 2023 Feb 26. PMID: 36798012; PMCID: PMC10175976.

## ΠΑΡΑΡΤΗΜΑ II Άρθρα της ομάδας Clinical Trial II

- [1.] 1: Swanson CJ, Zhang Y, Dhadda S, Wang J, Kaplow J, Lai RYK, Lannfelt L, Bradley H, Rabe M, Koyama A, Reyderman L, Berry DA, Berry S, Gordon R, Kramer LD, Cummings JL. A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-A $\beta$  protofibril antibody. *Alzheimers Res Ther.* 2021 Apr 17;13(1):80. doi: 10.1186/s13195-021-00813-8. Erratum in: *Alzheimers Res Ther.* 2022 May 21;14(1):70. PMID: 33865446; PMCID: PMC8053280.
- [2.] 2: Koch G, Casula EP, Bonni S, Borghi I, Assogna M, Minei M, Pellicciari MC, Motta C, D'Acunto A, Porraccini F, Maiella M, Ferrari C, Caltagirone C, Santarnecchi E, Bozzali M, Martorana A. Precuneus magnetic stimulation for Alzheimer's disease: a randomized, sham-controlled trial. *Brain.* 2022 Nov 21;145(11):3776-3786. doi: 10.1093/brain/awac285. PMID: 36281767; PMCID: PMC9679166.
- [3.] 3: McDade E, Cummings JL, Dhadda S, Swanson CJ, Reyderman L, Kanekiyo M, Koyama A, Irizarry M, Kramer LD, Bateman RJ. Lecanemab in patients with early Alzheimer's disease: detailed results on biomarker, cognitive, and clinical effects from the randomized and open-label extension of the phase 2 proof-of-concept study. *Alzheimers Res Ther.* 2022 Dec 21;14(1):191. doi: 10.1186/s13195-022-01124-2. PMID: 36544184; PMCID: PMC9768996.
- [4.] 4: Gibson GE, Luchsinger JA, Cirio R, Chen H, Franchino-Elder J, Hirsch JA, Bettendorff L, Chen Z, Flowers SA, Gerber LM, Grandville T, Schupf N, Xu H, Stern Y, Habeck C, Jordan B, Fonze P. Benfotiamine and Cognitive Decline in Alzheimer's Disease: Results of a Randomized Placebo-Controlled Phase IIa Clinical Trial. *J Alzheimers Dis.* 2020;78(3):989-1010. doi: 10.3233/JAD-200896. PMID: 33074237; PMCID: PMC7880246.
- [5.] 5: Rezai AR, Ranjan M, D'Haese PF, Haut MW, Carpenter J, Najib U, Mehta RI, Chazen JL, Zibly Z, Yates JR, Hodder SL, Kaplitt M. Noninvasive hippocampal blood-brain barrier opening in Alzheimer's disease with focused ultrasound. *Proc Natl Acad Sci U S A.* 2020 Apr 28;117(17):9180-9182. doi: 10.1073/pnas.2002571117. Epub 2020 Apr 13. PMID: 32284421; PMCID: PMC7196825.
- [6.] 6: Gonzales MM, Garbarino VR, Marques Zilli E, Petersen RC, Kirkland JL, Tchkonja T, Musi N, Seshadri S, Craft S, Orr ME. Senolytic Therapy to Modulate the Progression of Alzheimer's Disease (SToMP-AD): A Pilot Clinical Trial. *J Prev Alzheimers Dis.* 2022;9(1):22-29. doi: 10.14283/jpad.2021.62. PMID: 35098970; PMCID: PMC8612719.
- [7.] 7: Vossel K, Ranasinghe KG, Beagle AJ, La A, Ah Pook K, Castro M, Mizuiri D, Honma SM, Venkateswaran N, Koestler M, Zhang W, Mucke L, Howell MJ, Possin KL, Kramer JH, Boxer AL, Miller BL, Nagarajan SS, Kirsch HE. Effect of Levetiracetam on Cognition in Patients With Alzheimer Disease With and Without Epileptiform Activity: A Randomized Clinical Trial. *JAMA Neurol.* 2021 Nov 1;78(11):1345-1354. doi: 10.1001/jamaneurol.2021.3310. PMID: 34570177; PMCID: PMC8477304.
- [8.] 8: Shcherbinin S, Evans CD, Lu M, Andersen SW, Pontecorvo MJ, Willis BA, Gueorguieva I, Hauck PM, Brooks DA, Mintun MA, Sims JR. Association of Amyloid Reduction After Donanemab Treatment With Tau Pathology and Clinical Outcomes: The TRAILBLAZER-ALZ Randomized Clinical Trial. *JAMA Neurol.* 2022 Oct 1;79(10):1015-1024. doi: 10.1001/jamaneurol.2022.2793. PMID: 36094645; PMCID: PMC9468959.
- [9.] 9: Levey AI, Qiu D, Zhao L, Hu WT, Duong DM, Higginbotham L, Dammer EB, Seyfried NT, Wingo TS, Hales CM, Gámez Tansey M, Goldstein DS, Abrol A, Calhoun VD, Goldstein FC, Hajjar I, Fagan AM, Galasko D, Edland SD, Hanfelt J, Lah JJ, Weinschenker D. A phase II study repurposing atomoxetine for neuroprotection in mild cognitive impairment. *Brain.* 2022 Jun 30;145(6):1924-1938. doi: 10.1093/brain/awab452. PMID: 34919634; PMCID: PMC9630662.
- [10.] 10: Kehoe PG, Turner N, Howden B, Jarutyte L, Clegg SL, Malone IB, Barnes J, Nielsen C, Sudre CH, Wilson A, Thai NJ, Blair PS, Coulthard E, Lane JA, Passmore P, Taylor J, Mutsaerts HJ, Thomas DL, Fox NC, Wilkinson I, Ben-Shlomo Y; RADAR investigators. Safety and efficacy of losartan for the reduction of brain atrophy in clinically diagnosed Alzheimer's disease (the RADAR trial): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Neurol.* 2021 Nov;20(11):895-906. doi: 10.1016/S1474-4422(21)00263-5. PMID: 34687634; PMCID: PMC8528717.
- [11.] 11: Dhadda S, Kanekiyo M, Li D, Swanson CJ, Irizarry M, Berry S, Kramer LD, Berry DA. Consistency of efficacy results across various clinical measures and statistical methods in the lecanemab phase 2 trial of early Alzheimer's disease. *Alzheimers Res Ther.* 2022 Dec 9;14(1):182. doi: 10.1186/s13195-022-01129-x. PMID: 36482412; PMCID: PMC9733166.

- [12.] 12: Boada M, López OL, Olazarán J, Núñez L, Pfeffer M, Paricio M, Lorites J, Piñol-Ripoll G, Gámez JE, Anaya F, Kiprof D, Lima J, Grifols C, Torres M, Costa M, Bozzo J, Szczepiorkowski ZM, Hendrix S, Páez A. A randomized, controlled clinical trial of plasma exchange with albumin replacement for Alzheimer's disease: Primary results of the AMBAR Study. *Alzheimers Dement*. 2020 Oct;16(10):1412-1425. doi: 10.1002/alz.12137. Epub 2020 Jul 27. PMID: 32715623; PMCID: PMC7984263.
- [13.] 13: Yulug B, Altay O, Li X, Hanoglu L, Cankaya S, Lam S, Velioglu HA, Yang H, Coskun E, Idil E, Nogaylar R, Ozsimsek A, Bayram C, Bolat I, Oner S, Tozlu OO, Arslan ME, Hacimuftuoglu A, Yildirim S, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M, Mardinoglu A. Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: a randomised, double- blinded, placebo-controlled phase-II trial. *Transl Neurodegener*. 2023 Jan 26;12(1):4. doi: 10.1186/s40035-023-00336-2. PMID: 36703196; PMCID: PMC9879258.
- [14.] 14: Matthews DC, Mao X, Dowd K, Tsakanikas D, Jiang CS, Meuser C, Andrews RD, Lukic AS, Lee J, Hampilos N, Shafii N, Sano M, David Mozley P, Fillit H, McEwen BS, Shungu DC, Pereira AC. Riluzole, a glutamate modulator, slows cerebral glucose metabolism decline in patients with Alzheimer's disease. *Brain*. 2021 Dec 31;144(12):3742-3755. doi: 10.1093/brain/awab222. PMID: 34145880; PMCID: PMC8719848.
- [15.] 15: Wang T, Kuang W, Chen W, Xu W, Zhang L, Li Y, Li H, Peng Y, Chen Y, Wang B, Xiao J, Li H, Yan C, Du Y, Tang M, He Z, Chen H, Li W, Lin H, Shi S, Bi J, Zhou H, Cheng Y, Gao X, Guan Y, Huang Q, Chen K, Xin X, Ding J, Geng M, Xiao S. A phase II randomized trial of sodium oligomannate in Alzheimer's dementia. *Alzheimers Res Ther*. 2020 Sep 14;12(1):110. doi: 10.1186/s13195-020-00678-3. PMID: 32928279; PMCID: PMC7489025.
- [16.] 16: Kandadi Muralidharan K, Tong X, Kowalski KG, Rajagovindan R, Lin L, Budd Haberlain S, Nestorov I. Population pharmacokinetics and standard uptake value ratio of aducanumab, an amyloid plaque-removing agent, in patients with Alzheimer's disease. *CPT Pharmacometrics Syst Pharmacol*. 2022 Jan;11(1):7-19. doi: 10.1002/psp4.12728. Epub 2021 Oct 26. PMID: 34697913; PMCID: PMC8752104.
- [17.] 17: Teng E, Manser PT, Pickthorn K, Brunstein F, Blendstrup M, Sanabria Bohorquez S, Wildsmith KR, Toth B, Dolton M, Ramakrishnan V, Bobbala A, Sikkas SAM, Ward M, Fuji RN, Kerchner GA; Tauriel Investigators. Safety and Efficacy of Semorinemab in Individuals With Prodromal to Mild Alzheimer Disease: A Randomized Clinical Trial. *JAMA Neurol*. 2022 Aug 1;79(8):758-767. doi: 10.1001/jamaneurol.2022.1375. PMID: 35696185; PMCID: PMC9194753.
- [18.] 18: Vissers MFJM, Heuberger JAAC, Groeneveld GJ, Oude Nijhuis J, De Deyn PP, Hadi S, Harris J, Tsai RM, Cruz-Herranz A, Huang F, Tong V, Erickson R, Zhu Y, Scearce-Levie K, Hsiao-Nakamoto J, Tang X, Chang M, Fox BM, Estrada AA, Pomponio RJ, Alonso-Alonso M, Zilberstein M, Atassi N, Troyer MD, Ho C. Safety, pharmacokinetics and target engagement of novel RIPK1 inhibitor SAR443060 (DNL747) for neurodegenerative disorders: Randomized, placebo-controlled, double-blind phase I/Ib studies in healthy subjects and patients. *Clin Transl Sci*. 2022 Aug;15(8):2010-2023. doi: 10.1111/cts.13317. Epub 2022 Jun 1. PMID: 35649245; PMCID: PMC9372423.
- [19.] 19: Florian H, Wang D, Arnold SE, Boada M, Guo Q, Jin Z, Zheng H, Fisseha N, Kalluri HV, Rendenbach-Mueller B, Budur K, Gold M. Tilavonemab in early Alzheimer's disease: results from a phase 2, randomized, double-blind study. *Brain*. 2023 Jun 1;146(6):2275-2284. doi: 10.1093/brain/awad024. PMID: 36730056; PMCID: PMC10232284.
- [20.] 20: Kellar D, Lockhart SN, Aisen P, Raman R, Rissman RA, Brewer J, Craft S. Intranasal Insulin Reduces White Matter Hyperintensity Progression in Association with Improvements in Cognition and CSF Biomarker Profiles in Mild Cognitive Impairment and Alzheimer's Disease. *J Prev Alzheimers Dis*. 2021;8(3):240-248. doi: 10.14283/jpad.2021.14. PMID: 34101779; PMCID: PMC10233712.
- [21.] 21: Biglan K, Munsie L, Svensson KA, Ardayfio P, Pugh M, Sims J, Brys M. Safety and Efficacy of Mevidalen in Lewy Body Dementia: A Phase 2, Randomized, Placebo- Controlled Trial. *Mov Disord*. 2022 Mar;37(3):513-524. doi: 10.1002/mds.28879. Epub 2021 Dec 2. PMID: 34859493; PMCID: PMC9300146.
- [22.] 22: Boada M, López OL, Olazarán J, Núñez L, Pfeffer M, Puente O, Piñol-Ripoll G, Gámez JE, Anaya F, Kiprof D, Alegret M, Grifols C, Barceló M, Bozzo J, Szczepiorkowski ZM, Páez A; AMBAR Trial Group. Neuropsychological, neuropsychiatric, and quality-of-life assessments in Alzheimer's disease patients treated with plasma exchange with albumin replacement from the randomized AMBAR study. *Alzheimers Dement*. 2022 Jul;18(7):1314-1324. doi: 10.1002/alz.12477. Epub 2021 Nov 2. PMID: 34726348; PMCID: PMC9540900.

- [23.] 23: Prins ND, Harrison JE, Chu HM, Blackburn K, Alam JJ, Scheltens P; REVERSE-SD Study Investigators. A phase 2 double-blind placebo-controlled 24-week treatment clinical study of the p38 alpha kinase inhibitor neflamapimod in mild Alzheimer's disease. *Alzheimers Res Ther.* 2021 May 27;13(1):106. doi: 10.1186/s13195-021-00843-2. PMID: 34044875; PMCID: PMC8157623.
- [24.] 24: Chamberlain S, Gabriel H, Strittmatter W, Didsbury J. An Exploratory Phase IIa Study of the PPAR delta/gamma Agonist T3D-959 Assessing Metabolic and Cognitive Function in Subjects with Mild to Moderate Alzheimer's Disease. *J Alzheimers Dis.* 2020;73(3):1085-1103. doi: 10.3233/JAD-190864. PMID: 31884472; PMCID: PMC7081093.
- [25.] 25: Chen C, Kowahl NR, Rainaldi E, Burq M, Munsie LM, Battioui C, Wang J, Biglan K, Marks WJ Jr, Kapur R. Wrist-worn sensor-based measurements for drug effect detection with small samples in people with Lewy Body Dementia. *Parkinsonism Relat Disord.* 2023 Apr; 109:105355. doi: 10.1016/j.parkreldis.2023.105355. Epub 2023 Mar 4. PMID: 36905719.
- [26.] 26: VanDusen KW, Eleswarpu S, Moretti EW, Devinney MJ, Crabtree DM, Laskowitz DT, Woldorff MG, Roberts KC, Whittle J, Browndyke JN, Cooter M, Rockhold FW, Anakwenze O, Bolognesi MP, Easley ME, Ferrandino MN, Jiranek WA, Berger M; MARBLE Study Investigators. The MARBLE Study Protocol: Modulating ApoE Signaling to Reduce Brain Inflammation, DeLirium, and Postoperative Cognitive Dysfunction. *J Alzheimers Dis.* 2020;75(4):1319-1328. doi: 10.3233/JAD-191185. PMID: 32417770; PMCID: PMC7923142.
- [27.] 27: Koh SH, Kwon HS, Choi SH, Jeong JH, Na HR, Lee CN, Yang Y, Lee AY, Lee JH, Park KW, Han HJ, Kim BC, Park JS, Lee JY, Kim S, Lee KY. Efficacy and safety of GV1001 in patients with moderate-to-severe Alzheimer's disease already receiving donepezil: a phase 2 randomized, double-blind, placebo-controlled, multicenter clinical trial. *Alzheimers Res Ther.* 2021 Mar 26;13(1):66. doi: 10.1186/s13195-021-00803-w. PMID: 33771205; PMCID: PMC7995588.
- [28.] 28: Kim M, Kim SJ, Park JE, Yun J, Shim WH, Oh JS, Oh M, Roh JH, Seo SW, Oh SJ, Kim JS. Combination of automated brain volumetry on MRI and quantitative tau deposition on THK-5351 PET to support diagnosis of Alzheimer's disease. *Sci Rep.* 2021 May 14;11(1):10343. doi: 10.1038/s41598-021-89797-x. PMID: 33990649; PMCID: PMC8121780.
- [29.] 29: Cao K, Bay AA, Hajjar I, Wharton W, Goldstein F, Qiu D, Prusin T, McKay JL, Perkins MM, Hackney ME. Rationale and Design of the PARTNER Trial: Partnered Rhythmic Rehabilitation for Enhanced Motor-Cognition in Prodromal Alzheimer's Disease. *J Alzheimers Dis.* 2023;91(3):1019-1033. doi: 10.3233/JAD-220783. PMID: 36530084; PMCID: PMC10105523.
- [30.] 30: Valenzuela M, Duncan T, Abey A, Johnson A, Boulamatsis C, Dalton MA, Jacobson E, Brunel L, Child G, Simpson D, Buckland M, Lowe A, Siette J, Westbrook F, McGreevy P. Autologous skin-derived neural precursor cell therapy reverses canine Alzheimer dementia-like syndrome in a proof of concept veterinary trial. *Stem Cell Res Ther.* 2022 Jun 17;13(1):261. doi: 10.1186/s13287-022-02933-w. PMID: 35715872; PMCID: PMC9205057. 31: Park KS, Ganesh AB, Berry NT, Mobley YP, Karper WB, Labban JD, Wahlheim CN, Williams TM, Wideman L, Etnier JL. The effect of physical activity on cognition relative to APOE genotype (PAAD-2): study protocol for a phase II randomized control trial. *BMC Neurol.* 2020 Jun 6;20(1):231. doi: 10.1186/s12883-020-01732-1. PMID: 32503473; PMCID: PMC7274941.
- [31.] 32: Park KS, Etnier JL. An innovative protocol for the artificial speech- directed, contactless administration of laboratory-based comprehensive cognitive assessments: PAAD-2 trial management during the COVID-19 pandemic. *Contemp Clin Trials.* 2021 Aug;107:106500. doi: 10.1016/j.cct.2021.106500. Epub 2021 Jul 2. PMID: 34217888; PMCID: PMC8249106.
- [32.] 33: Sauder C, Allen LA, Baker E, Miller AC, Paul SM, Brannan SK. Effectiveness of KarXT (xanomeline-tropium) for cognitive impairment in schizophrenia: post hoc analyses from a randomised, double-blind, placebo-controlled phase 2 study. *Transl Psychiatry.* 2022 Nov 21;12(1):491. doi: 10.1038/s41398-022-02254-9. PMID: 36414626; PMCID: PMC9681874.
- [33.] 34: Yoshida K, Moein A, Bittner T, Ostrowitzki S, Lin H, Honigberg L, Jin JY, Quartino A. Pharmacokinetics and pharmacodynamic effect of crenezumab on plasma and cerebrospinal fluid beta-amyloid in patients with mild-to-moderate Alzheimer's disease. *Alzheimers Res Ther.* 2020 Jan 22;12(1):16. doi: 10.1186/s13195-020-0580-2. PMID: 31969177; PMCID: PMC6977279.
- [34.] 35: Rosenbloom M, Barclay T, Johnsen J, Erickson L, Svitak A, Pyle M, Frey W, Hanson LR. Double-Blind Placebo-Controlled Pilot Investigation of the Safety of a Single Dose of Rapid-Acting Intranasal Insulin in Down Syndrome. *Drugs R D.* 2020 Mar;20(1):11-15. doi: 10.1007/s40268-020-00296-2. PMID: 32077057; PMCID: PMC7067723.

- [35.] 36: Wang Y, Hernandez G, Mack WJ, Schneider LS, Yin F, Brinton RD. Retrospective analysis of phytoSERM for management of menopause-associated vasomotor symptoms and cognitive decline: a pilot study on pharmacogenomic effects of mitochondrial haplogroup and APOE genotype on therapeutic efficacy. *Menopause*. 2020 Jan;27(1):57-65. doi: 10.1097/GME.0000000000001418. PMID: 31567873; PMCID: PMC7100617.
- [36.] 37: Sperling R, Henley D, Aisen PS, Raman R, Donohue MC, Ernstrom K, Rafii MS, Streffer J, Shi Y, Karcher K, Raghavan N, Tymofyeyev Y, Bogert J, Brashear HR, Novak G, Thippahawong J, Saad ZS, Kolb H, Rofael H, Sanga P, Romano G. Findings of Efficacy, Safety, and Biomarker Outcomes of Atabecestat in Preclinical Alzheimer Disease: A Truncated Randomized Phase 2b/3 Clinical Trial. *JAMA Neurol*. 2021 Mar 1;78(3):293-301. doi: 10.1001/jamaneurol.2020.4857. PMID: 33464300; PMCID: PMC7816119.
- [37.] 38: Mehta RI, Carpenter JS, Mehta RI, Haut MW, Ranjan M, Najib U, Lockman P, Wang P, D'haese PF, Rezai AR. Blood-Brain Barrier Opening with MRI-guided Focused Ultrasound Elicits Meningeal Venous Permeability in Humans with Early Alzheimer Disease. *Radiology*. 2021 Mar;298(3):654-662. doi: 10.1148/radiol.2021200643. Epub 2021 Jan 5. PMID: 33399511; PMCID: PMC7924515.



## ΠΑΡΑΡΤΗΜΑ ΙΙΙ Άρθρα της ομάδας **Controlled Clinical Trial**

1: Swanson CJ, Zhang Y, Dhadda S, Wang J, Kaplow J, Lai RYK, Lannfelt L, Bradley H, Rabe M, Koyama A, Reyderman L, Berry DA, Berry S, Gordon R, Kramer LD, Cummings JL. A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-A $\beta$  protofibril antibody. *Alzheimers Res Ther.* 2021 Apr 17;13(1):80. doi: 10.1186/s13195-021-00813-8. Erratum in: *Alzheimers Res Ther.* 2022 May 21;14(1):70. PMID: 33865446; PMCID: PMC8053280.

2: Kim CS, Cha L, Sim M, Jung S, Chun WY, Baik HW, Shin DM. Probiotic Supplementation Improves Cognitive Function and Mood with Changes in Gut Microbiota in Community-Dwelling Older Adults: A Randomized, Double-Blind, Placebo-Controlled, Multicenter Trial. *J Gerontol A Biol Sci Med Sci.* 2021 Jan 1;76(1):32-40. doi: 10.1093/gerona/glaa090. PMID: 32300799; PMCID: PMC7861012.

3: Yu F, Vock DM, Zhang L, Salisbury D, Nelson NW, Chow LS, Smith G, Barclay TR, Dysken M, Wyman JF. Cognitive Effects of Aerobic Exercise in Alzheimer's Disease: A Pilot Randomized Controlled Trial. *J Alzheimers Dis.* 2021;80(1):233-244. doi: 10.3233/JAD-201100. PMID: 33523004; PMCID: PMC8075384.

4: Salloway S, Farlow M, McDade E, Clifford DB, Wang G, Llibre-Guerra JJ, Hitchcock JM, Mills SL, Santacruz AM, Aschenbrenner AJ, Hassenstab J, Benzinger TLS, Gordon BA, Fagan AM, Coalier KA, Cruchaga C, Goate AA, Perrin RJ, Xiong C, Li Y, Morris JC, Snider BJ, Mummery C, Surti GM, Hannequin D, Wallon D, Berman SB, Lah JJ, Jimenez-Velazquez IZ, Roberson ED, van Dyck CH, Honig LS, Sánchez-Valle R, Brooks WS, Gauthier S, Galasko DR, Masters CL, Brosch JR, Hsiung GR, Jayadev S, Formaglio M, Masellis M, Clarnette R, Pariente J, Dubois B, Pasquier F, Jack CR Jr, Koeppe R, Snyder PJ, Aisen PS, Thomas RG, Berry SM, Wendelberger BA, Andersen SW, Holdridge KC, Mintun MA, Yaari R, Sims JR, Baudler M, Delmar P, Doody RS, Fontoura P, Giacobino C, Kerchner GA, Bateman RJ; Dominantly Inherited Alzheimer Network–Trials Unit. A trial of gantenerumab or solanezumab in dominantly inherited Alzheimer's disease. *Nat Med.* 2021 Jul;27(7):1187-1196. doi: 10.1038/s41591-021-01369-8. Epub 2021 Jun 21. PMID: 34155411; PMCID: PMC8988051.

5: Phillips MCL, Deprez LM, Mortimer GMN, Murtagh DKJ, McCoy S, Mylchreest R, Gilbertson LJ, Clark KM, Simpson PV, McManus EJ, Oh JE, Yadavaraj S, King VM, Pillai A, Romero-Ferrando B, Brinkhuis M, Copeland BM, Samad S, Liao S, Schepel JAC. Randomized crossover trial of a modified ketogenic diet in Alzheimer's disease. *Alzheimers Res Ther.* 2021 Feb 23;13(1):51. doi: 10.1186/s13195-021-00783-x. PMID: 33622392; PMCID: PMC7901512.

6: Brown AJH, Bradley SJ, Marshall FH, Brown GA, Bennett KA, Brown J, Cansfield JE, Cross DM, de Graaf C, Hudson BD, Dwomoh L, Dias JM, Errey JC, Hurrell E, Liptrot J, Mattedi G, Molloy C, Nathan PJ, Okrasa K, Osborne G, Patel JC, Pickworth M, Robertson N, Shahabi S, Bundgaard C, Phillips K, Broad LM, Goonawardena AV, Morairty SR, Browning M, Perini F, Dawson GR, Deakin JFW, Smith RT, Sexton PM, Warneck J, Vinson M, Tasker T, Tehan BG, Teobald B, Christopoulos A, Langmead CJ, Jazayeri A, Cooke RM, Rucktooa P, Congreve MS, Weir M, Tobin AB. From structure to clinic: Design of a muscarinic M1 receptor agonist with potential to treatment of Alzheimer's disease. *Cell.* 2021 Nov 24;184(24):5886-5901.e22. doi: 10.1016/j.cell.2021.11.001. PMID: 34822784.

- 7: Koch G, Casula EP, Bonni S, Borghi I, Assogna M, Minei M, Pellicciari MC, Motta C, D'Acunto A, Porraccini F, Maiella M, Ferrari C, Caltagirone C, Santarnecchi E, Bozzali M, Martorana A. Precuneus magnetic stimulation for Alzheimer's disease: a randomized, sham-controlled trial. *Brain*. 2022 Nov 21;145(11):3776-3786. doi: 10.1093/brain/awac285. PMID: 36281767; PMCID: PMC9679166.
- 8: Gaitán JM, Moon HY, Stremlau M, Dubal DB, Cook DB, Okonkwo OC, van Praag H. Effects of Aerobic Exercise Training on Systemic Biomarkers and Cognition in Late Middle-Aged Adults at Risk for Alzheimer's Disease. *Front Endocrinol (Lausanne)*. 2021 May 20;12:660181. doi: 10.3389/fendo.2021.660181. PMID: 34093436; PMCID: PMC8173166.
- 9: Touns K, Hathaway A, Gordon D, Chung H, Raji C, Boyd A, Hill BD, Hausman-Cohen S, Attarha M, Chwa WJ, Jarrett M, Bredesen DE. Precision Medicine Approach to Alzheimer's Disease: Successful Pilot Project. *J Alzheimers Dis*. 2022;88(4):1411-1421. doi: 10.3233/JAD-215707. PMID: 35811518; PMCID: PMC9484109.
- 10: Kim D. The Effects of a Recollection-Based Occupational Therapy Program of Alzheimer's Disease: A Randomized Controlled Trial. *Occup Ther Int*. 2020 Jul 30;2020:6305727. doi: 10.1155/2020/6305727. PMID: 32821251; PMCID: PMC7416254.
- 11: Salloway S, Chalkias S, Barkhof F, Burkett P, Barakos J, Purcell D, Suhy J, Forrestal F, Tian Y, Umans K, Wang G, Singhal P, Budd Haeberlein S, Smirnakis K. Amyloid-Related Imaging Abnormalities in 2 Phase 3 Studies Evaluating Aducanumab in Patients With Early Alzheimer Disease. *JAMA Neurol*. 2022 Jan 1;79(1):13-21. doi: 10.1001/jamaneurol.2021.4161. PMID: 34807243; PMCID: PMC8609465.
- 12: Grossberg GT, Kohegyi E, Mergel V, Josiassen MK, Meulien D, Hobart M, Slomkowski M, Baker RA, McQuade RD, Cummings JL. Efficacy and Safety of Brexpiprazole for the Treatment of Agitation in Alzheimer's Dementia: Two 12-Week, Randomized, Double-Blind, Placebo-Controlled Trials. *Am J Geriatr Psychiatry*. 2020 Apr;28(4):383-400. doi: 10.1016/j.jagp.2019.09.009. Epub 2019 Oct 1. PMID: 31708380.
- 13: Liu X, Morris MC, Dhana K, Ventrelle J, Johnson K, Bishop L, Hollings CS, Boulin A, Laranjo N, Stubbs BJ, Reilly X, Carey VJ, Wang Y, Furtado JD, Marcovina SM, Tangney C, Aggarwal NT, Arfanakis K, Sacks FM, Barnes LL. Mediterranean-DASH Intervention for Neurodegenerative Delay (MIND) study: Rationale, design and baseline characteristics of a randomized control trial of the MIND diet on cognitive decline. *Contemp Clin Trials*. 2021 Mar;102:106270. doi: 10.1016/j.cct.2021.106270. Epub 2021 Jan 9. PMID: 33434704; PMCID: PMC8042655.
- 14: Gómez-Gallego M, Gómez-Gallego JC, Gallego-Mellado M, García-García J. Comparative Efficacy of Active Group Music Intervention versus Group Music Listening in Alzheimer's Disease. *Int J Environ Res Public Health*. 2021 Jul 30;18(15):8067. doi: 10.3390/ijerph18158067. PMID: 34360360; PMCID: PMC8345612.
- 15: McDade E, Cummings JL, Dhadda S, Swanson CJ, Reyderman L, Kanekiyo M, Koyama A, Irizarry M, Kramer LD, Bateman RJ. Lecanemab in patients with early Alzheimer's disease: detailed results on biomarker, cognitive, and clinical effects from the randomized and open-label extension of the phase 2 proof-of-concept study. *Alzheimers Res Ther*. 2022 Dec 21;14(1):191. doi: 10.1186/s13195-022-01124-2. PMID: 36544184; PMCID: PMC9768996.

16: Fortier M, Castellano CA, St-Pierre V, Myette-Côté É, Langlois F, Roy M, Morin MC, Bocti C, Fulop T, Godin JP, Delannoy C, Cuenoud B, Cunnane SC. A ketogenic drink improves cognition in mild cognitive impairment: Results of a 6-month RCT. *Alzheimers Dement*. 2021 Mar;17(3):543-552. doi: 10.1002/alz.12206. Epub 2020 Oct 26. PMID: 33103819; PMCID: PMC8048678.

17: Soininen H, Solomon A, Visser PJ, Hendrix SB, Blennow K, Kivipelto M, Hartmann T; LipiDiDiet clinical study group. 36-month LipiDiDiet multinutrient clinical trial in prodromal Alzheimer's disease. *Alzheimers Dement*. 2021 Jan;17(1):29-40. doi: 10.1002/alz.12172. Epub 2020 Sep 13. Erratum in: *Alzheimers Dement*. 2021 May;17(5):909. PMID: 32920957; PMCID: PMC7821311.

18: Mintzer J, Lanctôt KL, Scherer RW, Rosenberg PB, Herrmann N, van Dyck CH, Padala PR, Brawman-Mintzer O, Porsteinsson AP, Lerner AJ, Craft S, Levey AI, Burke W, Perin J, Shade D; ADMET 2 Research Group. Effect of Methylphenidate on Apathy in Patients With Alzheimer Disease: The ADMET 2 Randomized Clinical Trial. *JAMA Neurol*. 2021 Nov 1;78(11):1324-1332. doi: 10.1001/jamaneurol.2021.3356. PMID: 34570180; PMCID: PMC8477302.

19: Gibson GE, Luchsinger JA, Cirio R, Chen H, Franchino-Elder J, Hirsch JA, Bettendorff L, Chen Z, Flowers SA, Gerber LM, Grandville T, Schupf N, Xu H, Stern Y, Habeck C, Jordan B, Fonzetti P. Benfotiamine and Cognitive Decline in Alzheimer's Disease: Results of a Randomized Placebo-Controlled Phase IIa Clinical Trial. *J Alzheimers Dis*. 2020;78(3):989-1010. doi: 10.3233/JAD-200896. PMID: 33074237; PMCID: PMC7880246.

20: Xiao S, Chan P, Wang T, Hong Z, Wang S, Kuang W, He J, Pan X, Zhou Y, Ji Y, Wang L, Cheng Y, Peng Y, Ye Q, Wang X, Wu Y, Qu Q, Chen S, Li S, Chen W, Xu J, Peng D, Zhao Z, Li Y, Zhang J, Du Y, Chen W, Fan D, Yan Y, Liu X, Zhang W, Luo B, Wu W, Shen L, Liu C, Mao P, Wang Q, Zhao Q, Guo Q, Zhou Y, Li Y, Jiang L, Ren W, Ouyang Y, Wang Y, Liu S, Jia J, Zhang N, Liu Z, He R, Feng T, Lu W, Tang H, Gao P, Zhang Y, Chen L, Wang L, Yin Y, Xu Q, Xiao J, Cong L, Cheng X, Zhang H, Gao D, Xia M, Lian T, Peng G, Zhang X, Jiao B, Hu H, Chen X, Guan Y, Cui R, Huang Q, Xin X, Chen H, Ding Y, Zhang J, Feng T, Cantillon M, Chen K, Cummings JL, Ding J, Geng M, Zhang Z. A 36-week multicenter, randomized, double-blind, placebo-controlled, parallel-group, phase 3 clinical trial of sodium oligomannate for mild-to-moderate Alzheimer's dementia. *Alzheimers Res Ther*. 2021 Mar 17;13(1):62. doi: 10.1186/s13195-021-00795-7. PMID: 33731209; PMCID: PMC7967962.

21: Hoscheidt S, Sanderlin AH, Baker LD, Jung Y, Lockhart S, Kellar D, Whitlow CT, Hanson AJ, Friedman S, Register T, Leverenz JB, Craft S. Mediterranean and Western diet effects on Alzheimer's disease biomarkers, cerebral perfusion, and cognition in mid-life: A randomized trial. *Alzheimers Dement*. 2022 Mar;18(3):457-468. doi: 10.1002/alz.12421. Epub 2021 Jul 26. PMID: 34310044; PMCID: PMC9207984.

22: Arellanes IC, Choe N, Solomon V, He X, Kavin B, Martinez AE, Kono N, Buennagel DP, Hazra N, Kim G, D'Orazio LM, McCleary C, Sagare A, Zlokovic BV, Hodis HN, Mack WJ, Chui HC, Harrington MG, Braskie MN, Schneider LS, Yassine HN. Brain delivery of supplemental docosahexaenoic acid (DHA): A randomized placebo-controlled clinical trial. *EBioMedicine*. 2020 Sep;59:102883. doi: 10.1016/j.ebiom.2020.102883. Epub 2020 Jul 17. PMID: 32690472; PMCID: PMC7502665.

- 23: Gonzales MM, Garbarino VR, Marques Zilli E, Petersen RC, Kirkland JL, Tchkonja T, Musi N, Seshadri S, Craft S, Orr ME. Senolytic Therapy to Modulate the Progression of Alzheimer's Disease (SToMP-AD): A Pilot Clinical Trial. *J Prev Alzheimers Dis*. 2022;9(1):22-29. doi: 10.14283/jpad.2021.62. PMID: 35098970; PMCID: PMC8612719.
- 24: Vossel K, Ranasinghe KG, Beagle AJ, La A, Ah Pook K, Castro M, Mizuiri D, Honma SM, Venkateswaran N, Koestler M, Zhang W, Mucke L, Howell MJ, Possin KL, Kramer JH, Boxer AL, Miller BL, Nagarajan SS, Kirsch HE. Effect of Levetiracetam on Cognition in Patients With Alzheimer Disease With and Without Epileptiform Activity: A Randomized Clinical Trial. *JAMA Neurol*. 2021 Nov 1;78(11):1345-1354. doi: 10.1001/jamaneurol.2021.3310. PMID: 34570177; PMCID: PMC8477304.
- 25: Izzo NJ, Yuede CM, LaBarbera KM, Limegrover CS, Rehak C, Yurko R, Waybright L, Look G, Rishton G, Safferstein H, Hamby ME, Williams C, Sadlek K, Edwards HM, Davis CS, Grundman M, Schneider LS, DeKosky ST, Chelsky D, Pike I, Henstridge C, Blennow K, Zetterberg H, LeVine H 3rd, Spires-Jones TL, Cirrito JR, Catalano SM. Preclinical and clinical biomarker studies of CT1812: A novel approach to Alzheimer's disease modification. *Alzheimers Dement*. 2021 Aug;17(8):1365-1382. doi: 10.1002/alz.12302. Epub 2021 Feb 8. PMID: 33559354; PMCID: PMC8349378.
- 26: Chen CLH, Lu Q, Moorakonda RB, Kandiah N, Tan BY, Villaraza SG, Cano J, Venketasubramanian N. Alzheimer's Disease Therapy With NeuroAid (ATHENE): A Randomized Double-Blind Delayed-Start Trial. *J Am Med Dir Assoc*. 2022 Mar;23(3):379-386.e3. doi: 10.1016/j.jamda.2021.10.018. Epub 2021 Nov 29. PMID: 34856171.
- 27: Shcherbinin S, Evans CD, Lu M, Andersen SW, Pontecorvo MJ, Willis BA, Gueorguieva I, Hauck PM, Brooks DA, Mintun MA, Sims JR. Association of Amyloid Reduction After Donanemab Treatment With Tau Pathology and Clinical Outcomes: The TRAILBLAZER-ALZ Randomized Clinical Trial. *JAMA Neurol*. 2022 Oct 1;79(10):1015-1024. doi: 10.1001/jamaneurol.2022.2793. PMID: 36094645; PMCID: PMC9468959.
- 28: Levey AI, Qiu D, Zhao L, Hu WT, Duong DM, Higginbotham L, Dammer EB, Seyfried NT, Wingo TS, Hales CM, Gámez Tansey M, Goldstein DS, Abrol A, Calhoun VD, Goldstein FC, Hajjar I, Fagan AM, Galasko D, Edland SD, Hanfelt J, Lah JJ, Weinschenker D. A phase II study repurposing atomoxetine for neuroprotection in mild cognitive impairment. *Brain*. 2022 Jun 30;145(6):1924-1938. doi: 10.1093/brain/awab452. PMID: 34919634; PMCID: PMC9630662.
- 29: Kehoe PG, Turner N, Howden B, Jarutyte L, Clegg SL, Malone IB, Barnes J, Nielsen C, Sudre CH, Wilson A, Thai NJ, Blair PS, Coulthard E, Lane JA, Passmore P, Taylor J, Mutsaerts HJ, Thomas DL, Fox NC, Wilkinson I, Ben-Shlomo Y; RADAR investigators. Safety and efficacy of losartan for the reduction of brain atrophy in clinically diagnosed Alzheimer's disease (the RADAR trial): a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Neurol*. 2021 Nov;20(11):895-906. doi: 10.1016/S1474-4422(21)00263-5. PMID: 34687634; PMCID: PMC8528717.
- 30: Dhadda S, Kanekiyo M, Li D, Swanson CJ, Irizarry M, Berry S, Kramer LD, Berry DA. Consistency of efficacy results across various clinical measures and statistical methods in the lecanemab phase 2 trial of early Alzheimer's disease. *Alzheimers Res Ther*. 2022 Dec 9;14(1):182. doi: 10.1186/s13195-022-01129-x. PMID: 36482412; PMCID: PMC9733166.

- 31: Benussi A, Cantoni V, Grassi M, Brechet L, Michel CM, Datta A, Thomas C, Gazzina S, Cotelli MS, Bianchi M, Premi E, Gadola Y, Cotelli M, Pengo M, Perrone F, Sclaro M, Archetti S, Solje E, Padovani A, Pascual-Leone A, Borroni B. Increasing Brain Gamma Activity Improves Episodic Memory and Restores Cholinergic Dysfunction in Alzheimer's Disease. *Ann Neurol.* 2022 Aug;92(2):322-334. doi: 10.1002/ana.26411. Epub 2022 Jun 6. PMID: 35607946; PMCID: PMC9546168.
- 32: Craft S, Raman R, Chow TW, Rafii MS, Sun CK, Rissman RA, Donohue MC, Brewer JB, Jenkins C, Harless K, Gessert D, Aisen PS. Safety, Efficacy, and Feasibility of Intranasal Insulin for the Treatment of Mild Cognitive Impairment and Alzheimer Disease Dementia: A Randomized Clinical Trial. *JAMA Neurol.* 2020 Sep 1;77(9):1099-1109. doi: 10.1001/jamaneurol.2020.1840. PMID: 32568367; PMCID: PMC7309571.
- 33: Oliveira J, Gamito P, Souto T, Conde R, Ferreira M, Corotnean T, Fernandes A, Silva H, Neto T. Virtual Reality-Based Cognitive Stimulation on People with Mild to Moderate Dementia due to Alzheimer's Disease: A Pilot Randomized Controlled Trial. *Int J Environ Res Public Health.* 2021 May 16;18(10):5290. doi: 10.3390/ijerph18105290. PMID: 34065698; PMCID: PMC8156930.
- 34: Power R, Nolan JM, Prado-Cabrero A, Roche W, Coen R, Power T, Mulcahy R. Omega-3 fatty acid, carotenoid and vitamin E supplementation improves working memory in older adults: A randomised clinical trial. *Clin Nutr.* 2022 Feb;41(2):405-414. doi: 10.1016/j.clnu.2021.12.004. Epub 2021 Dec 7. PMID: 34999335.
- 35: Mendez Colmenares A, Voss MW, Fanning J, Salerno EA, Gothe NP, Thomas ML, McAuley E, Kramer AF, Burzynska AZ. White matter plasticity in healthy older adults: The effects of aerobic exercise. *Neuroimage.* 2021 Oct 1;239:118305. doi: 10.1016/j.neuroimage.2021.118305. Epub 2021 Jun 24. PMID: 34174392.
- 36: Viña J, Escudero J, Baquero M, Cebrián M, Carbonell-Asíns JA, Muñoz JE, Satorres E, Meléndez JC, Ferrer-Rebolleda J, Cózar-Santiago MDP, Santabárbara-Gómez JM, Jové M, Pamplona R, Tarazona-Santabalbina FJ, Borrás C. Genistein effect on cognition in prodromal Alzheimer's disease patients. The GENIAL clinical trial. *Alzheimers Res Ther.* 2022 Nov 4;14(1):164. doi: 10.1186/s13195-022-01097-2. PMID: 36329553; PMCID: PMC9635167.
- 37: Krause-Sorio B, Siddarth P, Kilpatrick L, Milillo MM, Aguilar-Faustino Y, Ercoli L, Narr KL, Khalsa DS, Lavretsky H. Yoga Prevents Gray Matter Atrophy in Women at Risk for Alzheimer's Disease: A Randomized Controlled Trial. *J Alzheimers Dis.* 2022;87(2):569-581. doi: 10.3233/JAD-215563. PMID: 35275541; PMCID: PMC9198760.
- 38: Boada M, López OL, Olazarán J, Núñez L, Pfeffer M, Paricio M, Lorites J, Piñol-Ripoll G, Gámez JE, Anaya F, Kiproff D, Lima J, Grifols C, Torres M, Costa M, Bozzo J, Szczepiorkowski ZM, Hendrix S, Páez A. A randomized, controlled clinical trial of plasma exchange with albumin replacement for Alzheimer's disease: Primary results of the AMBAR Study. *Alzheimers Dement.* 2020 Oct;16(10):1412-1425. doi: 10.1002/alz.12137. Epub 2020 Jul 27. PMID: 32715623; PMCID: PMC7984263.
- 39: Sen A, Akinola M, Tai XY, Symmonds M, Davis Jones G, Mura S, Galloway J, Hallam A, Chan JYC, Koychev I, Butler C, Geddes J, Van Der Putt R, Thompson S, Manohar SG, Frangou E, Love S, McShane R, Husain M. An Investigation of

Levetiracetam in Alzheimer's Disease (ILiAD): a double-blind, placebo-controlled, randomised crossover proof of concept study. *Trials*. 2021 Jul 31;22(1):508. doi: 10.1186/s13063-021-05404-4. PMID: 34332638; PMCID: PMC8325256.

40: Yulug B, Altay O, Li X, Hanoglu L, Cankaya S, Lam S, Velioglu HA, Yang H, Coskun E, Idil E, Nogaylar R, Ozsimsek A, Bayram C, Bolat I, Oner S, Tozlu OO, Arslan ME, Hacimuftuoglu A, Yildirim S, Arif M, Shoaie S, Zhang C, Nielsen J, Turkez H, Borén J, Uhlén M, Mardinoglu A. Combined metabolic activators improve cognitive functions in Alzheimer's disease patients: a randomised, double-blinded, placebo-controlled phase-II trial. *Transl Neurodegener*. 2023 Jan 26;12(1):4. doi: 10.1186/s40035-023-00336-2. PMID: 36703196; PMCID: PMC9879258.

41: Thota RN, Rosato JI, Dias CB, Burrows TL, Martins RN, Garg ML. Dietary Supplementation with Curcumin Reduce Circulating Levels of Glycogen Synthase Kinase-3 $\beta$  and Islet Amyloid Polypeptide in Adults with High Risk of Type 2 Diabetes and Alzheimer's Disease. *Nutrients*. 2020 Apr 9;12(4):1032. doi: 10.3390/nu12041032. PMID: 32283762; PMCID: PMC7230780.

42: Matthews DC, Mao X, Dowd K, Tsakanikas D, Jiang CS, Meuser C, Andrews RD, Lukic AS, Lee J, Hampilos N, Shafii N, Sano M, David Mozley P, Fillit H, McEwen BS, Shungu DC, Pereira AC. Riluzole, a glutamate modulator, slows cerebral glucose metabolism decline in patients with Alzheimer's disease. *Brain*. 2021 Dec 31;144(12):3742-3755. doi: 10.1093/brain/awab222. PMID: 34145880; PMCID: PMC8719848.

43: Banerjee S, High J, Stirling S, Shepstone L, Swart AM, Telling T, Henderson C, Ballard C, Bentham P, Burns A, Farina N, Fox C, Francis P, Howard R, Knapp M, Leroi I, Livingston G, Nilforooshan R, Nurock S, O'Brien J, Price A, Thomas AJ, Tabet N. Study of mirtazapine for agitated behaviours in dementia (SYMBAD): a randomised, double-blind, placebo-controlled trial. *Lancet*. 2021 Oct 23;398(10310):1487-1497. doi: 10.1016/S0140-6736(21)01210-1. PMID: 34688369; PMCID: PMC8546216.

44: Bisbe M, Fuente-Vidal A, López E, Moreno M, Naya M, de Benetti C, Milà R, Bruna O, Boada M, Alegret M. Comparative Cognitive Effects of Choreographed Exercise and Multimodal Physical Therapy in Older Adults with Amnesic Mild Cognitive Impairment: Randomized Clinical Trial. *J Alzheimers Dis*. 2020;73(2):769-783. doi: 10.3233/JAD-190552. PMID: 31868666; PMCID: PMC7029368.

45: Bagattini C, Zanni M, Barocco F, Caffarra P, Brignani D, Miniussi C, Defanti CA. Enhancing cognitive training effects in Alzheimer's disease: rTMS as an add-on treatment. *Brain Stimul*. 2020 Nov-Dec;13(6):1655-1664. doi: 10.1016/j.brs.2020.09.010. Epub 2020 Sep 28. PMID: 33002645.

46: Broadhouse KM, Singh MF, Suo C, Gates N, Wen W, Brodaty H, Jain N, Wilson GC, Meiklejohn J, Singh N, Baune BT, Baker M, Foroughi N, Wang Y, Kochan N, Ashton K, Brown M, Li Z, Mavros Y, Sachdev PS, Valenzuela MJ. Hippocampal plasticity underpins long-term cognitive gains from resistance exercise in MCI. *Neuroimage Clin*. 2020;25:102182. doi: 10.1016/j.nicl.2020.102182. Epub 2020 Jan 14. PMID: 31978826; PMCID: PMC6974789.

47: Dubois B, López-Arrieta J, Lipschitz S, Doskas T, Spuru L, Moroz S, Venger O, Vermersch P, Moussy A, Mansfield CD, Hermine O, Tsolaki M; AB09004 Study Group Investigators. Masitinib for mild-to-moderate Alzheimer's disease: results from a randomized, placebo-controlled, phase 3, clinical trial. *Alzheimers Res*

Ther. 2023 Feb 28;15(1):39. doi: 10.1186/s13195-023-01169-x. Erratum in: Alzheimers Res Ther. 2023 Apr 22;15(1):85. PMID: 36849969; PMCID: PMC9972756.

48: Blumen HM, Ayers E, Wang C, Ambrose AF, Verghese J. A social dancing pilot intervention for older adults at high risk for Alzheimer's disease and related dementias. Neurodegener Dis Manag. 2020 Aug;10(4):183-194. doi: 10.2217/nmt-2020-0002. Epub 2020 Aug 3. PMID: 32741240; PMCID: PMC7426754.

49: Benussi A, Cantoni V, Cotelli MS, Cotelli M, Brattini C, Datta A, Thomas C, Santarnecchi E, Pascual-Leone A, Borroni B. Exposure to gamma tACS in Alzheimer's disease: A randomized, double-blind, sham-controlled, crossover, pilot study. Brain Stimul. 2021 May-Jun;14(3):531-540. doi: 10.1016/j.brs.2021.03.007. Epub 2021 Mar 21. PMID: 33762220.

50: Barthélemy NR, Liu H, Lu W, Kotzbauer PT, Bateman RJ, Lucey BP. Sleep Deprivation Affects Tau Phosphorylation in Human Cerebrospinal Fluid. Ann Neurol. 2020 May;87(5):700-709. doi: 10.1002/ana.25702. Epub 2020 Feb 27. PMID: 32057125; PMCID: PMC7289061.

51: Wang T, Kuang W, Chen W, Xu W, Zhang L, Li Y, Li H, Peng Y, Chen Y, Wang B, Xiao J, Li H, Yan C, Du Y, Tang M, He Z, Chen H, Li W, Lin H, Shi S, Bi J, Zhou H, Cheng Y, Gao X, Guan Y, Huang Q, Chen K, Xin X, Ding J, Geng M, Xiao S. A phase II randomized trial of sodium oligomannate in Alzheimer's dementia. Alzheimers Res Ther. 2020 Sep 14;12(1):110. doi: 10.1186/s13195-020-00678-3. PMID: 32928279; PMCID: PMC7489025.

52: Noguchi-Shinohara M, Ono K, Hamaguchi T, Nagai T, Kobayashi S, Komatsu J, Samuraki-Yokohama M, Iwasa K, Yokoyama K, Nakamura H, Yamada M. Safety and efficacy of Melissa officinalis extract containing rosmarinic acid in the prevention of Alzheimer's disease progression. Sci Rep. 2020 Oct 29;10(1):18627. doi: 10.1038/s41598-020-73729-2. PMID: 33122694; PMCID: PMC7596544.

53: Li X, Qi G, Yu C, Lian G, Zheng H, Wu S, Yuan TF, Zhou D. Cortical plasticity is correlated with cognitive improvement in Alzheimer's disease patients after rTMS treatment. Brain Stimul. 2021 May-Jun;14(3):503-510. doi: 10.1016/j.brs.2021.01.012. Epub 2021 Feb 11. PMID: 33581283.

54: Devanand DP, Crocco E, Forester BP, Husain MM, Lee S, Vahia IV, Andrews H, Simon-Pearson L, Imran N, Luca L, Huey ED, Deliyannides DA, Pelton GH. Low Dose Lithium Treatment of Behavioral Complications in Alzheimer's Disease: Lit-AD Randomized Clinical Trial. Am J Geriatr Psychiatry. 2022 Jan;30(1):32-42. doi: 10.1016/j.jagp.2021.04.014. Epub 2021 May 12. PMID: 34059401; PMCID: PMC8586042.

55: Krueger KR, Dhana K, Aggarwal NT, Arfanakis K, Carey VJ, Sacks FM, Barnes LL. Properties of the Cognitive Function Battery for the MIND Diet Intervention to Prevent Alzheimer's Disease. J Int Neuropsychol Soc. 2022 Sep;28(8):790-797. doi: 10.1017/S1355617721001089. Epub 2021 Oct 6. PMID: 34612190; PMCID: PMC8983783.

56: Sur C, Kost J, Scott D, Adamczuk K, Fox NC, Cummings JL, Tariot PN, Aisen PS, Vellas B, Voss T, Mahoney E, Mukai Y, Kennedy ME, Lines C, Michelson D, Egan MF. BACE inhibition causes rapid, regional, and non-progressive volume reduction in Alzheimer's disease brain. Brain. 2020 Dec 1;143(12):3816-3826. doi: 10.1093/brain/awaa332. PMID: 33253354; PMCID: PMC8453290.

57: Asaoka D, Xiao J, Takeda T, Yanagisawa N, Yamazaki T, Matsubara Y, Sugiyama

H, Endo N, Higa M, Kasanuki K, Ichimiya Y, Koido S, Ohno K, Bernier F, Katsumata N, Nagahara A, Arai H, Ohkusa T, Sato N. Effect of Probiotic Bifidobacterium breve in Improving Cognitive Function and Preventing Brain Atrophy in Older Patients with Suspected Mild Cognitive Impairment: Results of a 24-Week Randomized, Double-Blind, Placebo-Controlled Trial. *J Alzheimers Dis.* 2022;88(1):75-95. doi: 10.3233/JAD-220148. PMID: 35570493; PMCID: PMC9277669.

58: Herring WJ, Ceesay P, Snyder E, Bliwise D, Budd K, Hutzelmann J, Stevens J, Lines C, Michelson D. Polysomnographic assessment of suvorexant in patients with probable Alzheimer's disease dementia and insomnia: a randomized trial. *Alzheimers Dement.* 2020 Mar;16(3):541-551. doi: 10.1002/alz.12035. Epub 2020 Jan 15. PMID: 31944580; PMCID: PMC7984350.

59: Pedrinolla A, Venturelli M, Fonte C, Tamburin S, Di Baldassarre A, Naro F, Varalta V, Giuriato G, Ghinassi B, Muti E, Smania N, Schena F. Exercise training improves vascular function in patients with Alzheimer's disease. *Eur J Appl Physiol.* 2020 Oct;120(10):2233-2245. doi: 10.1007/s00421-020-04447-w. Epub 2020 Jul 30. PMID: 32728820; PMCID: PMC7502067.

60: Liu CG, Zhao Y, Lu Y, Wang PC. ABCA1-Labeled Exosomes in Serum Contain Higher MicroRNA-193b Levels in Alzheimer's Disease. *Biomed Res Int.* 2021 Mar 8;2021:5450397. doi: 10.1155/2021/5450397. PMID: 33763470; PMCID: PMC7963893.

61: Ryan J, Storey E, Murray AM, Woods RL, Wolfe R, Reid CM, Nelson MR, Chong TTJ, Williamson JD, Ward SA, Lockery JE, Orchard SG, Trevaks R, Kirpach B, Newman AB, Ernst ME, McNeil JJ, Shah RC; ASPREE Investigator Group. Randomized placebo-controlled trial of the effects of aspirin on dementia and cognitive decline. *Neurology.* 2020 Jul 21;95(3):e320-e331. doi: 10.1212/WNL.0000000000009277. Epub 2020 Mar 25. PMID: 32213642; PMCID: PMC7455352.

62: Pontecorvo MJ, Lu M, Burnham SC, Schade AE, Dage JL, Shcherbinin S, Collins EC, Sims JR, Mintun MA. Association of Donanemab Treatment With Exploratory Plasma Biomarkers in Early Symptomatic Alzheimer Disease: A Secondary Analysis of the TRAILBLAZER-ALZ Randomized Clinical Trial. *JAMA Neurol.* 2022 Dec 1;79(12):1250-1259. doi: 10.1001/jamaneurol.2022.3392. PMID: 36251300; PMCID: PMC9577883.

63: Zmora R, Statz TL, Birkeland RW, McCarron HR, Finlay JM, Rosebush CE, Gaugler JE. Transitioning to Long-Term Care: Family Caregiver Experiences of Dementia, Communities, and Counseling. *J Aging Health.* 2021 Jan;33(1-2):133-146. doi: 10.1177/0898264320963588. Epub 2020 Sep 29. Erratum in: *J Aging Health.* 2021 Jun-Jul;33(5-6):NP1. PMID: 32990494; PMCID: PMC7891851.

64: Kellar D, Register T, Lockhart SN, Aisen P, Raman R, Rissman RA, Brewer J, Craft S. Intranasal insulin modulates cerebrospinal fluid markers of neuroinflammation in mild cognitive impairment and Alzheimer's disease: a randomized trial. *Sci Rep.* 2022 Jan 25;12(1):1346. doi: 10.1038/s41598-022-05165-3. PMID: 35079029; PMCID: PMC8789895.

65: Tomaszewski N, He X, Solomon V, Lee M, Mack WJ, Quinn JF, Braskie MN, Yassine HN. Effect of APOE Genotype on Plasma Docosahexaenoic Acid (DHA), Eicosapentaenoic Acid, Arachidonic Acid, and Hippocampal Volume in the Alzheimer's Disease Cooperative Study-Sponsored DHA Clinical Trial. *J Alzheimers Dis.* 2020;74(3):975-990. doi: 10.3233/JAD-191017. PMID: 32116250; PMCID: PMC7156328.



66: Fang X, Zhang J, Zhao J, Wang L. Effect of Resveratrol Combined with Donepezil Hydrochloride on Inflammatory Factor Level and Cognitive Function Level of Patients with Alzheimer's Disease. *J Healthc Eng.* 2022 Mar 25;2022:9148650. doi: 10.1155/2022/9148650. Retraction in: *J Healthc Eng.* 2023 Jan 25;2023:9834320. PMID: 35368930; PMCID: PMC8975642.

67: Chan D, Suk HJ, Jackson BL, Milman NP, Stark D, Klerman EB, Kitchener E, Fernandez Avalos VS, de Weck G, Banerjee A, Beach SD, Blanchard J, Stearns C, Boes AD, Uitermarkt B, Gander P, Howard M 3rd, Sternberg EJ, Nieto-Castanon A, Anteraper S, Whitfield-Gabrieli S, Brown EN, Boyden ES, Dickerson BC, Tsai LH. Gamma frequency sensory stimulation in mild probable Alzheimer's dementia patients: Results of feasibility and pilot studies. *PLoS One.* 2022 Dec 1;17(12):e0278412. doi: 10.1371/journal.pone.0278412. PMID: 36454969; PMCID: PMC9714926.

68: Teng E, Manser PT, Pickthorn K, Brunstein F, Blendstrup M, Sanabria Bohorquez S, Wildsmith KR, Toth B, Dolton M, Ramakrishnan V, Bobbala A, Sikkes SAM, Ward M, Fuji RN, Kerchner GA; Tauriel Investigators. Safety and Efficacy of Semorinemab in Individuals With Prodromal to Mild Alzheimer Disease: A Randomized Clinical Trial. *JAMA Neurol.* 2022 Aug 1;79(8):758-767. doi: 10.1001/jamaneurol.2022.1375. PMID: 35696185; PMCID: PMC9194753.

69: Vidoni ED, Morris JK, Watts A, Perry M, Clutton J, Van Sciver A, Kamat AS, Mahnken J, Hunt SL, Townley R, Honea R, Shaw AR, Johnson DK, Vacek J, Burns JM. Effect of aerobic exercise on amyloid accumulation in preclinical Alzheimer's: A 1-year randomized controlled trial. *PLoS One.* 2021 Jan 14;16(1):e0244893. doi: 10.1371/journal.pone.0244893. PMID: 33444359; PMCID: PMC7808620.

70: Berger M, Browndyke JN, Cooter Wright M, Nobuhara C, Reese M, Acker L, Bullock WM, Colin BJ, Devinney MJ, Moretti EW, Moul JW, Ohlendorf B, Laskowitz DT, Waligorska T, Shaw LM, Whitson HE, Cohen HJ, Mathew JP; MADCO-PC Investigators. Postoperative changes in cognition and cerebrospinal fluid neurodegenerative disease biomarkers. *Ann Clin Transl Neurol.* 2022 Feb;9(2):155-170. doi: 10.1002/acn3.51499. Epub 2022 Feb 1. PMID: 35104057; PMCID: PMC8862419.

71: Huo S, Cheng L, Li S, Xu F. Effects of eszopiclone on sleep quality and cognitive function in elderly patients with Alzheimer's disease and sleep disorder: A randomized controlled trial. *Brain Behav.* 2022 Feb;12(2):e2488. doi: 10.1002/brb3.2488. Epub 2022 Jan 18. PMID: 35041261; PMCID: PMC8865158.

72: Kouzuki M, Kitao S, Kaju T, Urakami K. Evaluation of the effect of aroma oil as a bath salt on cognitive function. *Psychogeriatrics.* 2020 Mar;20(2):163-171. doi: 10.1111/psyg.12481. Epub 2019 Oct 2. PMID: 31578055; PMCID: PMC7079117.

73: Vissers MFJM, Heuberger JAAC, Groeneveld GJ, Oude Nijhuis J, De Deyn PP, Hadi S, Harris J, Tsai RM, Cruz-Herranz A, Huang F, Tong V, Erickson R, Zhu Y, Scarce-Levie K, Hsiao-Nakamoto J, Tang X, Chang M, Fox BM, Estrada AA, Pomponio RJ, Alonso-Alonso M, Zilberstein M, Atassi N, Troyer MD, Ho C. Safety, pharmacokinetics and target engagement of novel RIPK1 inhibitor SAR443060 (DNL747) for neurodegenerative disorders: Randomized, placebo-controlled, double-blind phase I/Ib studies in healthy subjects and patients. *Clin Transl Sci.* 2022 Aug;15(8):2010-2023. doi: 10.1111/cts.13317. Epub 2022 Jun 1. PMID: 35649245; PMCID: PMC9372423.

74: Yao Q, Tang F, Wang Y, Yan Y, Dong L, Wang T, Zhu D, Tian M, Lin X, Shi J. Effect of cerebellum stimulation on cognitive recovery in patients with Alzheimer disease: A randomized clinical trial. *Brain Stimul.* 2022 Jul-Aug;15(4):910-920. doi: 10.1016/j.brs.2022.06.004. Epub 2022 Jun 11. PMID: 35700915.

75: Wright KD, Klatt MD, Adams IR, Nguyen CM, Mion LC, Tan A, Monroe TB, Rose KM, Scharre DW. Mindfulness in Motion and Dietary Approaches to Stop Hypertension (DASH) in Hypertensive African Americans. *J Am Geriatr Soc.* 2021 Mar;69(3):773-778. doi: 10.1111/jgs.16947. Epub 2020 Nov 23. PMID: 33227157; PMCID: PMC8329944.

76: Kaddoumi A, Denney TS Jr, Deshpande G, Robinson JL, Beyers RJ, Redden DT, Praticò D, Kyriakides TC, Lu B, Kirby AN, Beck DT, Merner ND. Extra-Virgin Olive Oil Enhances the Blood-Brain Barrier Function in Mild Cognitive Impairment: A Randomized Controlled Trial. *Nutrients.* 2022 Dec 1;14(23):5102. doi: 10.3390/nu14235102. PMID: 36501136; PMCID: PMC9736478.

77: Louzada LL, Machado FV, Quintas JL, Ribeiro GA, Silva MV, Mendonça-Silva DL, Gonçalves BSB, Nóbrega OT, Camargos EF. The efficacy and safety of zolpidem and zopiclone to treat insomnia in Alzheimer's disease: a randomized, triple-blind, placebo-controlled trial. *Neuropsychopharmacology.* 2022 Jan;47(2):570-579. doi: 10.1038/s41386-021-01191-3. Epub 2021 Oct 11. PMID: 34635802; PMCID: PMC8674235.

78: Liu-Ambrose T, Falck RS, Dao E, Best JR, Davis JC, Bennett K, Hall PA, Hsiung GR, Middleton LE, Goldsmith CH, Graf P, Eng JJ. Effect of Exercise Training or Complex Mental and Social Activities on Cognitive Function in Adults With Chronic Stroke: A Randomized Clinical Trial. *JAMA Netw Open.* 2022 Oct 3;5(10):e2236510. doi: 10.1001/jamanetworkopen.2022.36510. PMID: 36227593; PMCID: PMC9561961.

79: Takahashi Y, Shindo S, Kanbayashi T, Takeshima M, Imanishi A, Mishima K. Examination of the influence of cedar fragrance on cognitive function and behavioral and psychological symptoms of dementia in Alzheimer type dementia. *Neuropsychopharmacol Rep.* 2020 Mar;40(1):10-15. doi: 10.1002/npr2.12096. Epub 2020 Feb 9. PMID: 32037737; PMCID: PMC7292212.

80: Perła-Kaján J, Włoczkowska O, Ziola-Frankowska A, Frankowski M, Smith AD, de Jager CA, Refsum H, Jakubowski H. Paraoxonase 1, B Vitamins Supplementation, and Mild Cognitive Impairment. *J Alzheimers Dis.* 2021;81(3):1211-1229. doi: 10.3233/JAD-210137. PMID: 33935094; PMCID: PMC8293656.

81: Nasrallah IM, Gaussoin SA, Pomponio R, Dolui S, Erus G, Wright CB, Launer LJ, Detre JA, Wolk DA, Davatzikos C, Williamson JD, Pajewski NM, Bryan RN; SPRINT Research Group. Association of Intensive vs Standard Blood Pressure Control With Magnetic Resonance Imaging Biomarkers of Alzheimer Disease: Secondary Analysis of the SPRINT MIND Randomized Trial. *JAMA Neurol.* 2021 May 1;78(5):568-577. doi: 10.1001/jamaneurol.2021.0178. PMID: 33683313; PMCID: PMC7941253.

82: Maier F, Spottke A, Bach JP, Bartels C, Buerger K, Dodel R, Fellgiebel A, Fließbach K, Frölich L, Hausner L, Hellmich M, Klöppel S, Klostermann A, Kornhuber J, Laske C, Peters O, Priller J, Richter-Schmidinger T, Schneider A, Shah-Hosseini K, Teipel S, von Arnim CAF, Wiltfang J, Jessen F. Bupropion for the Treatment of Apathy in Alzheimer Disease: A Randomized Clinical Trial. *JAMA Netw Open.* 2020 May 1;3(5):e206027. doi: 10.1001/jamanetworkopen.2020.6027.

Erratum in: JAMA Netw Open. 2021 May 3;4(5):e2114100. PMID: 32463470; PMCID: PMC7256670.

83: Thomas BP, Tarumi T, Sheng M, Tseng B, Womack KB, Cullum CM, Rypma B, Zhang R, Lu H. Brain Perfusion Change in Patients with Mild Cognitive Impairment After 12 Months of Aerobic Exercise Training. J Alzheimers Dis. 2020;75(2):617-631. doi: 10.3233/JAD-190977. PMID: 32310162; PMCID: PMC8062932.

84: Ostrowitzki S, Bittner T, Sink KM, Mackey H, Rabe C, Honig LS, Cassetta E, Woodward M, Boada M, van Dyck CH, Grimmer T, Selkoe DJ, Schneider A, Blondeau K, Hu N, Quartino A, Clayton D, Dolton M, Dang Y, Ostaszewski B, Sanabria-Bohórquez SM, Rabbia M, Toth B, Eichenlaub U, Smith J, Honigberg LA, Doody RS. Evaluating the Safety and Efficacy of Crenezumab vs Placebo in Adults With Early Alzheimer Disease: Two Phase 3 Randomized Placebo-Controlled Trials. JAMA Neurol. 2022 Nov 1;79(11):1113-1121. doi: 10.1001/jamaneurol.2022.2909. PMID: 36121669; PMCID: PMC9486635.

85: Kluger BM, Miyasaki J, Katz M, Galifianakis N, Hall K, Pantilat S, Khan R, Friedman C, Cernik W, Goto Y, Long J, Fairclough D, Sillau S, Kutner JS. Comparison of Integrated Outpatient Palliative Care With Standard Care in Patients With Parkinson Disease and Related Disorders: A Randomized Clinical Trial. JAMA Neurol. 2020 May 1;77(5):551-560. doi: 10.1001/jamaneurol.2019.4992. PMID: 32040141; PMCID: PMC7042842.

86: Hasturk H, Schulte F, Martins M, Sherzai H, Floros C, Cugini M, Chiu CJ, Hardt M, Van Dyke T. Safety and Preliminary Efficacy of a Novel Host-Modulatory Therapy for Reducing Gingival Inflammation. Front Immunol. 2021 Sep 13;12:704163. doi: 10.3389/fimmu.2021.704163. PMID: 34589083; PMCID: PMC8475270.

87: Shi S, Zhang Q, Sang Y, Ge S, Wang Q, Wang R, He J. Probiotic *Bifidobacterium longum* BB68S Improves Cognitive Functions in Healthy Older Adults: A Randomized, Double-Blind, Placebo-Controlled Trial. Nutrients. 2022 Dec 22;15(1):51. doi: 10.3390/nu15010051. PMID: 36615708; PMCID: PMC9824790.

88: Bourdon E, Belmin J. Enriched gardens improve cognition and independence of nursing home residents with dementia: a pilot controlled trial. Alzheimers Res Ther. 2021 Jun 16;13(1):116. doi: 10.1186/s13195-021-00849-w. PMID: 34134758; PMCID: PMC8207740.

89: Sindi S, Thunborg C, Rosenberg A, Andersen P, Andrieu S, Broersen LM, Coley N, Couderc C, Duval CZ, Faxen-Irving G, Hagman G, Hallikainen M, Håkansson K, Lehtisalo J, Levak N, Mangialasche F, Pantel J, Kekkonen E, Rydström A, Stigsdotter-Neely A, Wimo A, Ngandu T, Soininen H, Hartmann T, Solomon A, Kivipelto M. Multimodal Preventive Trial for Alzheimer's Disease: MIND-ADmini Pilot Trial Study Design and Progress. J Prev Alzheimers Dis. 2022;9(1):30-39. doi: 10.14283/jpad.2022.4. PMID: 35098971; PMCID: PMC8783958.

90: Liu X, Dhana K, Furtado JD, Agarwal P, Aggarwal NT, Tangney C, Laranjo N, Carey V, Barnes LL, Sacks FM. Higher circulating  $\alpha$ -carotene was associated with better cognitive function: an evaluation among the MIND trial participants. J Nutr Sci. 2021 Aug 16;10:e64. doi: 10.1017/jns.2021.56. PMID: 34527222; PMCID: PMC8411267.

91: Greimel S, Wyman JF, Zhang L, Yu F. Recruitment and Screening Methods in Alzheimer's Disease Research: The FIT-AD Trial. J Gerontol A Biol Sci Med Sci.

2022 Mar 3;77(3):547-553. doi: 10.1093/gerona/glab092. PMID: 33780529; PMCID: PMC8893175.

92: Zhang J, Kuang X, Tang C, Xu N, Xiao S, Xiao L, Wang S, Dong Y, Lu L, Zhang L. Acupuncture for amnesic mild cognitive impairment: A pilot multicenter, randomized, parallel controlled trial. *Medicine (Baltimore)*. 2021 Nov 19;100(46):e27686. doi: 10.1097/MD.00000000000027686. PMID: 34797294; PMCID: PMC8601273.

93: Florian H, Wang D, Arnold SE, Boada M, Guo Q, Jin Z, Zheng H, Fisseha N, Kalluri HV, Rendenbach-Mueller B, Budur K, Gold M. Tilavonemab in early Alzheimer's disease: results from a phase 2, randomized, double-blind study. *Brain*. 2023 Jun 1;146(6):2275-2284. doi: 10.1093/brain/awad024. PMID: 36730056; PMCID: PMC10232284.

94: Lavretsky H, Laird KT, Krause-Sorio B, Heimberg BF, Yeargin J, Grzenda A, Wu P, Thana-Udom K, Ercoli LM, Siddarth P. A Randomized Double-Blind Placebo-Controlled Trial of Combined Escitalopram and Memantine for Older Adults With Major Depression and Subjective Memory Complaints. *Am J Geriatr Psychiatry*. 2020 Feb;28(2):178-190. doi: 10.1016/j.jagp.2019.08.011. Epub 2019 Aug 22. PMID: 31519517; PMCID: PMC6997044.

95: Komulainen P, Tuomilehto J, Savonen K, Männikkö R, Hassinen M, Lakka TA, Hänninen T, Kiviniemi V, Jacobs DR, Kivipelto M, Rauramaa R. Exercise, diet, and cognition in a 4-year randomized controlled trial: Dose-Responses to Exercise Training (DR's EXTRA). *Am J Clin Nutr*. 2021 Jun 1;113(6):1428-1439. doi: 10.1093/ajcn/nqab018. PMID: 33742194; PMCID: PMC8244125.

96: Rapp SR, Gaussoin SA, Sachs BC, Chelune G, Supiano MA, Lerner AJ, Wadley VG, Wilson VM, Fine LJ, Whittle JC, Auchus AP, Beddhu S, Berlowitz DR, Bress AP, Johnson KC, Krousel-Wood M, Martindale-Adams J, Miller EC, Rifkin DE, Snyder JK, Tamariz L, Wolfgram DF, Cleveland ML, Yang M, Nichols LO, Bryan RN, Reboussin DM, Williamson JD, Pajewski NM; SPRINT Research Group. Effects of intensive versus standard blood pressure control on domain-specific cognitive function: a substudy of the SPRINT randomised controlled trial. *Lancet Neurol*. 2020 Nov;19(11):899-907. doi: 10.1016/S1474-4422(20)30319-7. PMID: 33098800; PMCID: PMC7714000.

97: Jeong MK, Park KW, Ryu JK, Kim GM, Jung HH, Park H. Multi-Component Intervention Program on Habitual Physical Activity Parameters and Cognitive Function in Patients with Mild Cognitive Impairment: A Randomized Controlled Trial. *Int J Environ Res Public Health*. 2021 Jun 9;18(12):6240. doi: 10.3390/ijerph18126240. PMID: 34207701; PMCID: PMC8296099.

98: Rosenberg A, Solomon A, Soininen H, Visser PJ, Blennow K, Hartmann T, Kivipelto M; LipiDiDiet clinical study group. Research diagnostic criteria for Alzheimer's disease: findings from the LipiDiDiet randomized controlled trial. *Alzheimers Res Ther*. 2021 Mar 25;13(1):64. doi: 10.1186/s13195-021-00799-3. PMID: 33766132; PMCID: PMC7995792.

99: Lam LCW, Chan WC, Kwok TCY, Lee JSW, Yu BML, Lee S, Lee ATC, Ma SL, Cheng ST. Combined physical exercise-working memory training on slowing down cognitive decline in elders with mild clinical Alzheimer disease: a randomised controlled study (abridged secondary publication). *Hong Kong Med J*. 2022 Jun;28 Suppl 3(3):28-30. PMID: 35701227.

100: Stites SD, Gill J, Largent EA, Harkins K, Sankar P, Krieger A, Karlawish J. The relative contributions of biomarkers, disease modifying treatment, and dementia severity to Alzheimer's stigma: A vignette-based experiment. *Soc Sci Med*. 2022 Jan;292:114620. doi: 10.1016/j.socscimed.2021.114620. Epub 2021 Dec 1. PMID: 34883313; PMCID: PMC8748419.

101: Wischik CM, Bentham P, Gauthier S, Miller S, Kook K, Schelter BO. Oral Tau Aggregation Inhibitor for Alzheimer's Disease: Design, Progress and Basis for Selection of the 16 mg/day Dose in a Phase 3, Randomized, Placebo-Controlled Trial of Hydromethylthionine Mesylate. *J Prev Alzheimers Dis*. 2022;9(4):780-790. doi: 10.14283/jpad.2022.63. PMID: 36281683; PMCID: PMC9226274.

102: Yu X, Srivastava S, Huang S, Hayden EY, Teplow DB, Xie YH. The Feasibility of Early Alzheimer's Disease Diagnosis Using a Neural Network Hybrid Platform. *Biosensors (Basel)*. 2022 Sep 13;12(9):753. doi: 10.3390/bios12090753. PMID: 36140138; PMCID: PMC9496690.

103: Kellar D, Lockhart SN, Aisen P, Raman R, Rissman RA, Brewer J, Craft S. Intranasal Insulin Reduces White Matter Hyperintensity Progression in Association with Improvements in Cognition and CSF Biomarker Profiles in Mild Cognitive Impairment and Alzheimer's Disease. *J Prev Alzheimers Dis*. 2021;8(3):240-248. doi: 10.14283/jpad.2021.14. PMID: 34101779; PMCID: PMC10233712.

104: Padala PR, Boozer EM, Lensing SY, Parkes CM, Hunter CR, Dennis RA, Caceda R, Padala KP. Neuromodulation for Apathy in Alzheimer's Disease: A Double-Blind, Randomized, Sham-Controlled Pilot Study. *J Alzheimers Dis*. 2020;77(4):1483-1493. doi: 10.3233/JAD-200640. PMID: 32925060; PMCID: PMC7683089.

105: Shen X, Zhang Q, Shao W, Shi J, Liu B. Medication reconciliation role and value in Alzheimer's disease treatment. *Arq Neuropsiquiatr*. 2022 Mar;80(3):233-239. doi: 10.1590/0004-282X-ANP-2021-0147. PMID: 35239819; PMCID: PMC9648934.

106: Targum SD, Fosdick L, Drake KE, Rosenberg PB, Burke AD, Wolk DA, Foote KD, Asaad WF, Sabbagh M, Smith GS, Lozano AM, Lyketsos CG. Effect of Age on Clinical Trial Outcome in Participants with Probable Alzheimer's Disease. *J Alzheimers Dis*. 2021;82(3):1243-1257. doi: 10.3233/JAD-210530. PMID: 34151817; PMCID: PMC8461716.

107: Saredakis D, Keage HA, Corlis M, Ghezzi ES, Loffler H, Loetscher T. The Effect of Reminiscence Therapy Using Virtual Reality on Apathy in Residential Aged Care: Multisite Nonrandomized Controlled Trial. *J Med Internet Res*. 2021 Sep 20;23(9):e29210. doi: 10.2196/29210. PMID: 34542418; PMCID: PMC8491119.

108: Kondo T, Banno H, Okunomiya T, Amino Y, Endo K, Nakakura A, Uozumi R, Kinoshita A, Tada H, Morita S, Ishikawa H, Shindo A, Yasuda K, Taruno Y, Maki T, Suehiro T, Mori K, Ikeda M, Fujita K, Izumi Y, Kanemaru K, Ishii K, Shigenobu K, Kutoku Y, Sunada Y, Kawakatsu S, Shiota S, Watanabe T, Uchikawa O, Takahashi R, Tomimoto H, Inoue H. Repurposing bromocriptine for A $\beta$  metabolism in Alzheimer's disease (REBRANd) study: randomised placebo-controlled double-blind comparative trial and open-label extension trial to investigate the safety and efficacy of bromocriptine in Alzheimer's disease with presenilin 1 (PSEN1) mutations. *BMJ Open*. 2021 Jun 30;11(6):e051343. doi: 10.1136/bmjopen-2021-051343. PMID: 34193504; PMCID: PMC8246358.

109: Lerner AJ, Arnold SE, Maxfield E, Koenig A, Toth ME, Fortin B, Mast N, Trombetta BA, Denker J, Pieper AA, Tatsuoka C, Raghupathy S, Pikuleva IA. CYP46A1 activation by low-dose efavirenz enhances brain cholesterol metabolism in subjects with early Alzheimer's disease. *Alzheimers Res Ther.* 2022 Dec 29;14(1):198. doi: 10.1186/s13195-022-01151-z. PMID: 36581878; PMCID: PMC9797897.

110: Torres-Mendoza BMG, Ortiz GG, Sánchez-Romero L, Delgado-Lara DLC, García Martínez MT, Mireles-Ramírez MA, Cruz Serrano JA, Pacheco Moisés FP. Dietary fish oil increases catalase activity in patients with probable Alzheimer's disease. *Nutr Hosp.* 2022 Dec 20;39(6):1364-1368. English. doi: 10.20960/nh.04153. PMID: 36327127.

111: Biglan K, Munsie L, Svensson KA, Ardayfio P, Pugh M, Sims J, Brys M. Safety and Efficacy of Mevidalen in Lewy Body Dementia: A Phase 2, Randomized, Placebo-Controlled Trial. *Mov Disord.* 2022 Mar;37(3):513-524. doi: 10.1002/mds.28879. Epub 2021 Dec 2. PMID: 34859493; PMCID: PMC9300146.

112: Koch M, DeKosky ST, Goodman M, Sun J, Furtado JD, Fitzpatrick AL, Mackey RH, Cai T, Lopez OL, Kuller LH, Mukamal KJ, Jensen MK. High density lipoprotein and its apolipoprotein-defined subspecies and risk of dementia. *J Lipid Res.* 2020 Mar;61(3):445-454. doi: 10.1194/jlr.P119000473. Epub 2019 Dec 31. PMID: 31892526; PMCID: PMC7053836.

113: Wunderlich G, Blahova Z, Garcia M, Jessen F. Efficacy and safety of the novel GlyT1 inhibitor BI 425809 in Alzheimer's dementia: a randomized controlled trial. *Alzheimers Res Ther.* 2023 Jan 28;15(1):24. doi: 10.1186/s13195-023-01163-3. PMID: 36709275; PMCID: PMC9883916.

114: Hu Y, Jia Y, Sun Y, Ding Y, Huang Z, Liu C, Wang Y. Efficacy and safety of simultaneous rTMS-tDCS over bilateral angular gyrus on neuropsychiatric symptoms in patients with moderate Alzheimer's disease: A prospective, randomized, sham-controlled pilot study. *Brain Stimul.* 2022 Nov-Dec;15(6):1530-1537. doi: 10.1016/j.brs.2022.11.009. Epub 2022 Nov 30. PMID: 36460293.

115: Hua X, Church K, Walker W, L'Hostis P, Viardot G, Danjou P, Hendrix S, Moebius HJ. Safety, Tolerability, Pharmacokinetics, and Pharmacodynamics of the Positive Modulator of HGF/MET, Fosgonimeton, in Healthy Volunteers and Subjects with Alzheimer's Disease: Randomized, Placebo-Controlled, Double-Blind, Phase I Clinical Trial. *J Alzheimers Dis.* 2022;86(3):1399-1413. doi: 10.3233/JAD-215511. PMID: 35180125; PMCID: PMC9108585.

116: Muñoz-Bermejo L, González-Becerra MJ, Barrios-Fernández S, Postigo-Mota S, Jerez-Barroso MDR, Martínez JAF, Suárez-Lantarón B, Marín DM, Martín-Bermúdez N, Ortés-Gómez R, Gómez-Ullate-García de León M, Martínez-Acevedo M, Rocha-Gómez L, Espejo-Antúnez S, Fraile-Bravo M, Galán MGS, Chato-Gonzalo I, Muñoz FJD, Hernández-Mocholí MÁ, Madruga-Vicente M, Prado-Solano A, Mendoza-Muñoz M, Carlos-Vivas J, Pérez-Gómez J, Pastor-Cisneros R, Fuentes-Flores P, Pereira-Payo D, De Los Ríos-Calonge J, Urbano-Mairena J, Guerra-Bustamante J, Adsuar JC. Cost-Effectiveness of the Comprehensive Interdisciplinary Program-Care in Informal Caregivers of People with Alzheimer's Disease. *Int J Environ Res Public Health.* 2022 Nov 18;19(22):15243. doi: 10.3390/ijerph192215243. PMID: 36429962; PMCID: PMC9691117.

117: Villars H, Cantet C, de Peretti E, Perrin A, Soto-Martin M, Gardette V. Impact of an educational programme on Alzheimer's disease patients' quality of

life: results of the randomized controlled trial THERAD. *Alzheimers Res Ther*. 2021 Sep 12;13(1):152. doi: 10.1186/s13195-021-00896-3. PMID: 34511121; PMCID: PMC8436545.

118: Alvarez XA, Winston CN, Barlow JW, Sarsoza FM, Alvarez I, Aleixandre M, Linares C, García-Fantini M, Kastberger B, Winter S, Rissman RA. Modulation of Amyloid- $\beta$  and Tau in Alzheimer's Disease Plasma Neuronal-Derived Extracellular Vesicles by Cerebrolysin® and Donepezil. *J Alzheimers Dis*. 2022;90(2):705-717. doi: 10.3233/JAD-220575. PMID: 36155516; PMCID: PMC9697063.

119: Levine SZ, Goldberg Y, Yoshida K, Samara M, Cipriani A, Iwatsubo T, Leucht S, Furawaka TA. Quantifying the heterogeneity of cognitive functioning in Alzheimer's disease to extend the placebo-treatment dichotomy: Latent class analysis of individual-participant data from five pivotal randomized clinical trials of donepezil. *Eur Psychiatry*. 2021 Feb 15;64(1):e16. doi: 10.1192/j.eurpsy.2021.8. PMID: 33583479; PMCID: PMC8057455.

120: Novak G, Streffer JR, Timmers M, Henley D, Brashear HR, Bogert J, Russu A, Janssens L, Tesseur I, Tritsmans L, Van Nueten L, Engelborghs S. Long-term safety and tolerability of atabecestat (JNJ-54861911), an oral BACE1 inhibitor, in early Alzheimer's disease spectrum patients: a randomized, double-blind, placebo-controlled study and a two-period extension study. *Alzheimers Res Ther*. 2020 May 14;12(1):58. doi: 10.1186/s13195-020-00614-5. PMID: 32410694; PMCID: PMC7227237.

121: Brody M, Agronin M, Herskowitz BJ, Bookheimer SY, Small GW, Hitchinson B, Ramdas K, Wishard T, McInerney KF, Vellas B, Sierra F, Jiang Z, McClain-Moss L, Perez C, Fuquay A, Rodriguez S, Hare JM, Oliva AA Jr, Baumel B. Results and insights from a phase I clinical trial of Lomecel-B for Alzheimer's disease. *Alzheimers Dement*. 2023 Jan;19(1):261-273. doi: 10.1002/alz.12651. Epub 2022 Mar 31. PMID: 35357079; PMCID: PMC10084163.

122: Takada M, Tanaka S, Tanaka K, Tsukie T, Tsukamoto-Yasui M, Suzuki K, Noguchi Y, Imaizumi A, Ishii M, Ikeuchi T. Effects of an essential amino acid mixture on behavioral and psychological symptoms of dementia and executive function in patients with Alzheimer's disease: A double-blind, randomized, placebo-controlled exploratory clinical trial. *Int J Geriatr Psychiatry*. 2022 Sep;37(9):10.1002/gps.5782. doi: 10.1002/gps.5782. PMID: 35932158; PMCID: PMC9544995.

123: Kovaleva M, Nocera JR, Hepburn K, Higgins M, Nash R, Epps F, Brewster G, Bilsborough E, Blumling AA, Griffiths PC. Attention control in a trial of an online psychoeducational intervention for caregivers. *Res Nurs Health*. 2022 Jun;45(3):337-349. doi: 10.1002/nur.22208. Epub 2022 Jan 25. PMID: 35080016; PMCID: PMC9064916.

124: Ma W, Zhu L, Tang J, Diao W, Qian L, Feng X, Zhang X. Testing the knowledge of Alzheimer's disease <i>via</i> an intervention study among community health service center staff in Jiaying, China. *Front Public Health*. 2023 Jan 27;10:969653. doi: 10.3389/fpubh.2022.969653. PMID: 36777777; PMCID: PMC9911520.

125: Baakman AC, Gavan C, van Doeselaar L, de Kam M, Broekhuizen K, Bajenaru O, Camps L, Swart EL, Kalisvaart K, Schoonenboom N, Lemstra E, Scheltens P, Cohen A, van Gerven J, Groeneveld GJ. Acute response to cholinergic challenge predicts long-term response to galantamine treatment in patients with Alzheimer's disease. *Br J Clin Pharmacol*. 2022 Jun;88(6):2814-2829. doi: 10.1111/bcp.15206.

Epub 2022 Jan 26. PMID: 34964149; PMCID: PMC9306507.

126: Pongan E, Delphin-Combe F, Krolak-Salmon P, Leveque Y, Tillmann B, Bachelet R, Getenet JC, Auguste N, Trombert B, Dorey JM, Laurent B, Rouch I. Immediate Benefit of Art on Pain and Well-Being in Community-Dwelling Patients with Mild Alzheimer's. *Am J Alzheimers Dis Other Dement.* 2020 Jan-Dec;35:1533317519859202. doi: 10.1177/1533317519859202. Epub 2019 Jul 9. PMID: 31288544.

127: Vecchio F, Quaranta D, Miraglia F, Pappalettera C, Di Iorio R, L'Abbate F, Cotelli M, Marra C, Rossini PM. Neuronavigated Magnetic Stimulation combined with cognitive training for Alzheimer's patients: an EEG graph study. *Geroscience.* 2022 Feb;44(1):159-172. doi: 10.1007/s11357-021-00508-w. Epub 2021 Dec 31. PMID: 34970718; PMCID: PMC8811083.

128: Wu X, Ji GJ, Geng Z, Wang L, Yan Y, Wu Y, Xiao G, Gao L, Wei Q, Zhou S, Wei L, Tian Y, Wang K. Accelerated intermittent theta-burst stimulation broadly ameliorates symptoms and cognition in Alzheimer's disease: A randomized controlled trial. *Brain Stimul.* 2022 Jan-Feb;15(1):35-45. doi: 10.1016/j.brs.2021.11.007. Epub 2021 Nov 6. PMID: 34752934.

129: Vidoni ED, Kamat A, Gahan WP, Ourso V, Woodard K, Kerwin DR, Binder EF, Burns JM, Cullum M, Hyman LS, Vongpatanasin W, Zhu DC, Zhang R, Keller JN. Baseline Prevalence of Polypharmacy in Older Hypertensive Study Subjects with Elevated Dementia Risk: Findings from the Risk Reduction for Alzheimer's Disease Study (rrAD). *J Alzheimers Dis.* 2020;77(1):175-182. doi: 10.3233/JAD-200122. PMID: 32716358; PMCID: PMC8108402.

130: Traini E, Carotenuto A, Fasanaro AM, Amenta F. Volume Analysis of Brain Cognitive Areas in Alzheimer's Disease: Interim 3-Year Results from the ASCOMALVA Trial. *J Alzheimers Dis.* 2020;76(1):317-329. doi: 10.3233/JAD-190623. PMID: 32508323; PMCID: PMC7369051.

131: Cuadrado F, Antolí A, Fernández-Calvo B. The effect of framing on attitudes towards Alzheimer's disease. A comparative study between younger and older adults. *PLoS One.* 2022 Jul 7;17(7):e0270959. doi: 10.1371/journal.pone.0270959. PMID: 35797268; PMCID: PMC9262181.

132: Silverman MH, Duggan S, Bardelli G, Sadler B, Key C, Medlock M, Reynolds L, Wallner B. Safety, Tolerability and Pharmacokinetics of Icapamespib, a Selective Epichaperome Inhibitor, in Healthy Adults. *J Prev Alzheimers Dis.* 2022;9(4):635-645. doi: 10.14283/jpad.2022.71. PMID: 36281667; PMCID: PMC9419134.

133: Butler T, Goldberg JD, Galvin JE, Maloney T, Ravdin L, Glodzik L, de Leon MJ, Hochman T, Bowen RL, Atwood CS. Rationale, study design and implementation of the LUCINDA Trial: Leuprolide plus Cholinesterase Inhibition to reduce Neurologic Decline in Alzheimer's. *Contemp Clin Trials.* 2021 Aug;107:106488. doi: 10.1016/j.cct.2021.106488. Epub 2021 Jun 22. PMID: 34166841; PMCID: PMC8550816.

134: Boada M, López OL, Olazarán J, Núñez L, Pfeffer M, Puente O, Piñol-Ripoll G, Gámez JE, Anaya F, Kiproff D, Alegret M, Grifols C, Barceló M, Bozzo J, Szczepiorkowski ZM, Páez A; AMBAR Trial Group. Neuropsychological, neuropsychiatric, and quality-of-life assessments in Alzheimer's disease patients treated with plasma exchange with albumin replacement from the randomized AMBAR study. *Alzheimers Dement.* 2022 Jul;18(7):1314-1324. doi:



10.1002/alz.12477. Epub 2021 Nov 2. PMID: 34726348; PMCID: PMC9540900.

135: Ptomey LT, Szabo-Reed AN, Vidoni ED, Washburn RA, Gorczyca AM, Little TD, Lee J, Helsel BC, Williams KN, Donnelly JE. A dyadic approach for a remote physical activity intervention in adults with Alzheimer's disease and their caregivers: Rationale and design for an 18-month randomized trial. *Contemp Clin Trials*. 2020 Nov;98:106158. doi: 10.1016/j.cct.2020.106158. Epub 2020 Sep 23. PMID: 32979515; PMCID: PMC7686020.

136: Tao Y, Lei B, Zhu Y, Fang X, Liao L, Chen D, Gao C. Repetitive Transcranial Magnetic Stimulation Decreases Serum Amyloid- $\beta$  and Increases Ectodomain of p75 Neurotrophin Receptor in Patients with Alzheimer's Disease. *J Integr Neurosci*. 2022 Jul 29;21(5):140. doi: 10.31083/j.jin2105140. PMID: 36137949.

137: de Heus RAA, de Jong DLK, Rijpma A, Lawlor BA, Olde Rikkert MGM, Claassen JAHR. Orthostatic Blood Pressure Recovery Is Associated With the Rate of Cognitive Decline and Mortality in Clinical Alzheimer's Disease. *J Gerontol A Biol Sci Med Sci*. 2020 Oct 15;75(11):2169-2176. doi: 10.1093/gerona/glaa129. PMID: 32449919; PMCID: PMC7566323.

138: Tariot PN, Braeckman R, Oh C. Comparison of Steady-State Pharmacokinetics of Donepezil Transdermal Delivery System with Oral Donepezil. *J Alzheimers Dis*. 2022;90(1):161-172. doi: 10.3233/JAD-220530. PMID: 36120781; PMCID: PMC9661317.

139: DiStefano MJ, Alexander GC, Polsky D, Anderson GF. Public opinion regarding U.S. Food and Drug Administration approval of aducanumab and potential policy responses: A nationally representative survey. *J Am Geriatr Soc*. 2022 Jun;70(6):1685-1694. doi: 10.1111/jgs.17692. Epub 2022 Feb 7. PMID: 35129210; PMCID: PMC9177789.

140: Stroehlein JK, Vieluf S, Zimmer P, Schenk A, Oberste M, Goelz C, van den Bongard F, Reinsberger C. Learning to play golf for elderly people with subjective memory complaints: feasibility of a single-blinded randomized pilot trial. *BMC Neurol*. 2021 May 17;21(1):200. doi: 10.1186/s12883-021-02186-9. PMID: 34001020; PMCID: PMC8127313.

141: Senderovich H, Gardner S, Berall A, Shultz R, Grant B, Santaguida V. Therapeutic Touch in the Management of Responsive Behaviors in Patients with Dementia. *Dement Geriatr Cogn Disord*. 2022;51(2):142-149. doi: 10.1159/000523752. Epub 2022 Mar 25. PMID: 35340008.

142: El-Sappagh S, Alonso JM, Islam SMR, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. *Sci Rep*. 2021 Jan 29;11(1):2660. doi: 10.1038/s41598-021-82098-3. PMID: 33514817; PMCID: PMC7846613.

143: Yu F, Salisbury D, Mathiason MA. Inter-individual differences in the responses to aerobic exercise in Alzheimer's disease: Findings from the FIT-AD trial. *J Sport Health Sci*. 2021 Jan;10(1):65-72. doi: 10.1016/j.jshs.2020.05.007. Epub 2020 Jun 4. PMID: 32505710; PMCID: PMC7858037.

144: Murray NWG, Graham PL, Sowman PF, Savage G. Theta tACS impairs episodic memory more than tDCS. *Sci Rep*. 2023 Jan 13;13(1):716. doi: 10.1038/s41598-022-27190-y. PMID: 36639676; PMCID: PMC9839727.

145: Mitchell LL, Peterson CM, Rud SR, Jutkowitz E, Sarkinen A, Trost S, Porta

CM, Finlay JM, Gaugler JE. "It's Like a Cyber-Security Blanket": The Utility of Remote Activity Monitoring in Family Dementia Care. *J Appl Gerontol*. 2020 Jan;39(1):86-98. doi: 10.1177/0733464818760238. Epub 2018 Mar 4. PMID: 29504488; PMCID: PMC6353693.

146: Lanctôt KL, Scherer RW, Li A, Vieira D, Coulibaly H, Rosenberg PB, Herrmann N, Lerner AJ, Padala PR, Brawman-Mintzer O, van Dyck CH, Porsteinsson AP, Craft S, Levey A, Burke WJ, Mintzer JE. Measuring Apathy in Alzheimer's Disease in the Apathy in Dementia Methylphenidate Trial 2 (ADMET 2): A Comparison of Instruments. *Am J Geriatr Psychiatry*. 2021 Jan;29(1):81-89. doi: 10.1016/j.jagp.2020.05.020. Epub 2020 May 27. PMID: 32565008; PMCID: PMC7704818.

147: Moga DC, Abner EL, Schmitt FA, Eckmann L, Huffmyer M, Martinez AI, Beech BF, George R, El Khouli RH, Ali D, Jicha GA. Intervention for Cognitive Reserve Enhancement in Delaying the Onset of Alzheimer's Symptomatic Expression (INCREASE) Study: Results from a Randomized Controlled Study of Medication Therapy Management Targeting a Delay in Prodromal Dementia Symptom Progression. *J Prev Alzheimers Dis*. 2022;9(4):646-654. doi: 10.14283/jpad.2022.55. PMID: 36281668; PMCID: PMC9145115.

148: Parker LJ, Marx K, Gaugler JE, Gitlin LN. Implications of the COVID-19 Pandemic on Adult Day Services and the Families They Serve. *Am J Alzheimers Dis Other Demen*. 2021 Jan-Dec;36:15333175211050152. doi: 10.1177/15333175211050152. PMID: 34647482; PMCID: PMC8745481.

149: Gaugler JE, Zmora R, Mitchell LL, Finlay J, Rosebush CE, Nkimbeng M, Baker ZG, Albers EA, Peterson CM. Remote activity monitoring for family caregivers of persons living with dementia: a mixed methods, randomized controlled evaluation. *BMC Geriatr*. 2021 Dec 18;21(1):715. doi: 10.1186/s12877-021-02634-8. PMID: 34922475; PMCID: PMC8684277.

150: Russin KJ, Nair KS, Montine TJ, Baker LD, Craft S. Diet Effects on Cerebrospinal Fluid Amino Acids Levels in Adults with Normal Cognition and Mild Cognitive Impairment. *J Alzheimers Dis*. 2021;84(2):843-853. doi: 10.3233/JAD-210471. PMID: 34602470; PMCID: PMC8673538.

151: Hickman SE, Mitchell SL, Hanson LC, Tu W, Stump TE, Unroe KT. The design and conduct of a pragmatic cluster randomized trial of an advance care planning program for nursing home residents with dementia. *Clin Trials*. 2022 Dec;19(6):623-635. doi: 10.1177/17407745221108992. Epub 2022 Jul 10. PMID: 35815777; PMCID: PMC9691516.

152: Towsley GL, Wong B, Baier RR, Neller S. An Efficacy Trial of Long-Term Care Residents With Alzheimer's Disease Using Videos to Communicate Care Preferences. *J Am Med Dir Assoc*. 2021 Jul;22(7):1559-1560. doi: 10.1016/j.jamda.2021.02.032. Epub 2021 Mar 30. PMID: 33798485; PMCID: PMC8297558.

153: Shimokawa H, Shindo T, Ishiki A, Tomita N, Ichijyo S, Watanabe T, Nakata T, Eguchi K, Kikuchi Y, Shiroto T, Takahashi J, Yasuda S, Arai H. A Pilot Study of Whole-Brain Low-Intensity Pulsed Ultrasound Therapy for Early Stage of Alzheimer's Disease (LIPUS-AD): A Randomized, Double-Blind, Placebo-Controlled Trial. *Tohoku J Exp Med*. 2022 Oct 25;258(3):167-175. doi: 10.1620/tjem.2022.J078. Epub 2022 Sep 15. PMID: 36104179.

154: Cuberas-Borrós G, Roca I, Castell-Conesa J, Núñez L, Boada M, López OL, Grifols C, Barceló M, Pareto D, Páez A. Neuroimaging analyses from a randomized,

controlled study to evaluate plasma exchange with albumin replacement in mild-to-moderate Alzheimer's disease: additional results from the AMBAR study. *Eur J Nucl Med Mol Imaging*. 2022 Nov;49(13):4589-4600. doi: 10.1007/s00259-022-05915-5. Epub 2022 Jul 22. PMID: 35867135; PMCID: PMC9606044.

155: Korthauer LE, Giugliano RP, Guo J, Sabatine MS, Sever P, Keech A, Atar D, Kurtz C, Ruff CT, Mach F, Ott BR. No association between APOE genotype and lipid lowering with cognitive function in a randomized controlled trial of evolocumab. *PLoS One*. 2022 Apr 11;17(4):e0266615. doi: 10.1371/journal.pone.0266615. PMID: 35404972; PMCID: PMC9000128.

156: Rossetto F, Isernia S, Realdon O, Borgnis F, Blasi V, Pagliari C, Cabinio M, Alberoni M, Mantovani F, Clerici M, Baglio F. A digital health home intervention for people within the Alzheimer's disease continuum: results from the Ability-TelerehABILITation pilot randomized controlled trial. *Ann Med*. 2023 Dec;55(1):1080-1091. doi: 10.1080/07853890.2023.2185672. PMID: 36929703; PMCID: PMC10030155.

157: Prins ND, Harrison JE, Chu HM, Blackburn K, Alam JJ, Scheltens P; REVERSE-SD Study Investigators. A phase 2 double-blind placebo-controlled 24-week treatment clinical study of the p38 alpha kinase inhibitor neflamapimod in mild Alzheimer's disease. *Alzheimers Res Ther*. 2021 May 27;13(1):106. doi: 10.1186/s13195-021-00843-2. PMID: 34044875; PMCID: PMC8157623.

158: Harrison TM, Ward TJ, Murphy A, Baker SL, Dominguez PA, Koeppe R, Vemuri P, Lockhart SN, Jung Y, Harvey DJ, Lovato L, Toga AW, Masdeu J, Oh H, Gitelman DR, Aggarwal N, Snyder HM, Baker LD, DeCarli C, Jagust WJ, Landau SM; U.S. POINTER Study Group. Optimizing quantification of MK6240 tau PET in unimpaired older adults. *Neuroimage*. 2023 Jan;265:119761. doi: 10.1016/j.neuroimage.2022.119761. Epub 2022 Nov 28. PMID: 36455762; PMCID: PMC9957642.

159: Hodgson NA, Petrovsky DV, Finegan K, Kallmyer BA, Pike J, Fazio S. One call makes a difference: An evaluation of the Alzheimer's Association National Helpline on dementia caregiver outcomes. *Patient Educ Couns*. 2021 Apr;104(4):896-902. doi: 10.1016/j.pec.2020.09.026. Epub 2020 Sep 23. PMID: 33004235; PMCID: PMC10001241.

160: Li B, Ji S, Peng A, Yang N, Zhao X, Feng P, Zhang Y, Chen L. Development of a Gastrointestinal-Myoelectrical-Activity-Based Nomogram Model for Predicting the Risk of Mild Cognitive Impairment. *Biomolecules*. 2022 Dec 12;12(12):1861. doi: 10.3390/biom12121861. PMID: 36551289; PMCID: PMC9775682.

161: Potashman M, Pang M, Tahir M, Shahraz S, Dichter S, Perneckzy R, Nolte S. Psychometric properties of the Alzheimer's Disease Cooperative Study - Activities of Daily Living for Mild Cognitive Impairment (ADCS-MCI-ADL) scale: a post hoc analysis of the ADCS ADC-008 trial. *BMC Geriatr*. 2023 Mar 6;23(1):124. doi: 10.1186/s12877-022-03527-0. PMID: 36879199; PMCID: PMC9990271.

162: Berens SC, Bird CM, Harrison NA. Minocycline differentially modulates human spatial memory systems. *Neuropsychopharmacology*. 2020 Dec;45(13):2162-2169. doi: 10.1038/s41386-020-00811-8. Epub 2020 Aug 24. PMID: 32839527; PMCID: PMC7784680.

163: Masuoka N, Lei C, Li H, Inamura N, Shiotani S, Yanai N, Sato K, Sakurai K, Hisatsune T. Anserine, HClO-scavenger, protected against cognitive decline in individuals with mild cognitive impairment. *Aging (Albany NY)*. 2021 Jan 20;13(2):1729-1741. doi: 10.18632/aging.202535. Epub 2021 Jan 20. PMID:

33472172; PMID: PMC7880346.

164: Yuda N, Tanaka M, Yamauchi K, Abe F, Kakiuchi I, Kiyosawa K, Miyasaka M, Sakane N, Nakamura M. Effect of the Casein-Derived Peptide Met-Lys-Pro on Cognitive Function in Community-Dwelling Adults Without Dementia: A Randomized, Double-Blind, Placebo-Controlled Trial. *Clin Interv Aging*. 2020 May 27;15:743-754. doi: 10.2147/CIA.S253116. PMID: 32546992; PMCID: PMC7266326.

165: Chamberlain S, Gabriel H, Strittmatter W, Didsbury J. An Exploratory Phase IIa Study of the PPAR delta/gamma Agonist T3D-959 Assessing Metabolic and Cognitive Function in Subjects with Mild to Moderate Alzheimer's Disease. *J Alzheimers Dis*. 2020;73(3):1085-1103. doi: 10.3233/JAD-190864. PMID: 31884472; PMCID: PMC7081093.

166: Chen C, Kowahl NR, Rainaldi E, Burq M, Munsie LM, Battiou C, Wang J, Biglan K, Marks WJ Jr, Kapur R. Wrist-worn sensor-based measurements for drug effect detection with small samples in people with Lewy Body Dementia. *Parkinsonism Relat Disord*. 2023 Apr;109:105355. doi: 10.1016/j.parkreldis.2023.105355. Epub 2023 Mar 4. PMID: 36905719.

167: Bentley TGK, Castillo D, Sadeghi N, Piber D, Carroll J, Olmstead R, Irwin MR. Costs associated with treatment of insomnia in Alzheimer's disease caregivers: a comparison of mindfulness meditation and cognitive behavioral therapy for insomnia. *BMC Health Serv Res*. 2022 Feb 19;22(1):231. doi: 10.1186/s12913-022-07619-w. PMID: 35183180; PMCID: PMC8858547.

168: Chen H, Levkoff S, Chuengsatiansup K, Sihapark S, Hinton L, Gallagher-Thompson D, Tongsiri S, Wisetpholchai B, Fritz S, Lamont A, Domlyn A, Wandersman A, Marques AH. Implementation Science in Thailand: Design and Methods of a Geriatric Mental Health Cluster-Randomized Trial. *Psychiatr Serv*. 2022 Jan 1;73(1):83-91. doi: 10.1176/appi.ps.202000028. Epub 2021 Jun 15. PMID: 34126778; PMCID: PMC8785231.

169: Lane HY, Wang SH, Lin CH. Endogenous antioxidants predicted outcome and increased after treatment: A benzoate dose-finding, randomized, double-blind, placebo-controlled trial for Alzheimer's disease. *Psychiatry Clin Neurosci*. 2023 Feb;77(2):102-109. doi: 10.1111/pcn.13504. Epub 2022 Nov 24. PMID: 36335573; PMCID: PMC10099492.

170: Cheung DSK, Ho LYW, Kwok RKH, Lai DLL, Lai CKY. The effects of involvement in training and volunteering with families of people with dementia on the knowledge and attitudes of volunteers towards dementia. *BMC Public Health*. 2022 Feb 8;22(1):258. doi: 10.1186/s12889-022-12687-y. PMID: 35135524; PMCID: PMC8822773.

171: Kim YJ, Kim HR, Jung YH, Park YH, Seo SW. Effects of Electrical Automatic Massage on Cognition and Sleep Quality in Alzheimer's Disease Spectrum Patients: A Randomized Controlled Trial. *Yonsei Med J*. 2021 Aug;62(8):717-725. doi: 10.3349/ymj.2021.62.8.717. PMID: 34296549; PMCID: PMC8298867.

172: Cummings J, Schwartz GG, Nicholls SJ, Khan A, Halliday C, Toth PP, Sweeney M, Johansson JO, Wong NCW, Kulikowski E, Kalantar-Zadeh K, Lebioda K, Ginsberg HN, Winblad B, Zetterberg H, Ray KK. Cognitive Effects of the BET Protein Inhibitor Apabetalone: A Prespecified Montreal Cognitive Assessment Analysis Nested in the BETonMACE Randomized Controlled Trial. *J Alzheimers Dis*. 2021;83(4):1703-1715. doi: 10.3233/JAD-210570. PMID: 34459400; PMCID:

PMC8609701.

173: Han SS, White K, Cisek E. A Feasibility Study of Individuals Living at Home with Alzheimer's Disease and Related Dementias: Utilization of Visual Mapping Assistive Technology to Enhance Quality of Life and Reduce Caregiver Burden. *Clin Interv Aging*. 2022 Dec 23;17:1885-1892. doi: 10.2147/CIA.S387255. PMID: 36583148; PMCID: PMC9793743.

174: Huang P, He XY, Xu M. Dengzhan shengmai capsule combined with donepezil hydrochloride in the treatment of Alzheimer's disease: preliminary findings, randomized and controlled clinical trial. *Rev Assoc Med Bras (1992)*. 2021 Feb;67(2):190-194. doi: 10.1590/1806-9282.67.02.20200378. PMID: 34287475.

175: Petrella JR, Michael AM, Qian M, Nwosu A, Sneed J, Goldberg TE, Devanand DP, Doraiswamy PM. Impact of Computerized Cognitive Training on Default Mode Network Connectivity in Subjects at Risk for Alzheimer's Disease: A 78-week Randomized Controlled Trial. *J Alzheimers Dis*. 2023;91(1):483-494. doi: 10.3233/JAD-220946. PMID: 36442202; PMCID: PMC9881022.

176: VanDusen KW, Eleswarpu S, Moretti EW, Devinney MJ, Crabtree DM, Laskowitz DT, Woldorff MG, Roberts KC, Whittle J, Browndyke JN, Cooter M, Rockhold FW, Anakwenze O, Bolognesi MP, Easley ME, Ferrandino MN, Jiranek WA, Berger M; MARBLE Study Investigators. The MARBLE Study Protocol: Modulating ApoE Signaling to Reduce Brain Inflammation, DeLirium, and Postoperative Cognitive Dysfunction. *J Alzheimers Dis*. 2020;75(4):1319-1328. doi: 10.3233/JAD-191185. PMID: 32417770; PMCID: PMC7923142.

177: Uijen IL, Aaronson JA, Karssemeijer EGA, Olde Rikkert MGM, Kessels RPC. Individual Differences in the Effects of Physical Activity on Cognitive Function in People with Mild to Moderate Dementia. *J Alzheimers Dis*. 2020;74(2):435-439. doi: 10.3233/JAD-190606. PMID: 32039840; PMCID: PMC7175929.

178: Choi WY, Lee WK, Kim TH, Ryu YK, Park A, Lee YJ, Heo SJ, Oh C, Chung YC, Kang DH. The Effects of *Spirulina maxima* Extract on Memory Improvement in Those with Mild Cognitive Impairment: A Randomized, Double-Blind, Placebo-Controlled Clinical Trial. *Nutrients*. 2022 Sep 9;14(18):3714. doi: 10.3390/nu14183714. PMID: 36145090; PMCID: PMC9505028.

179: Koh SH, Kwon HS, Choi SH, Jeong JH, Na HR, Lee CN, Yang Y, Lee AY, Lee JH, Park KW, Han HJ, Kim BC, Park JS, Lee JY, Kim S, Lee KY. Efficacy and safety of GV1001 in patients with moderate-to-severe Alzheimer's disease already receiving donepezil: a phase 2 randomized, double-blind, placebo-controlled, multicenter clinical trial. *Alzheimers Res Ther*. 2021 Mar 26;13(1):66. doi: 10.1186/s13195-021-00803-w. PMID: 33771205; PMCID: PMC7995588.

180: Macchi ZA, Ayele R, Dini M, Lamira J, Katz M, Pantilat SZ, Jones J, Kluger BM. Lessons from the COVID-19 pandemic for improving outpatient neuropalliative care: A qualitative study of patient and caregiver perspectives. *Palliat Med*. 2021 Jul;35(7):1258-1266. doi: 10.1177/02692163211017383. Epub 2021 May 18. PMID: 34006157; PMCID: PMC9447864.

181: Delrieu J, Voisin T, Saint-Aubert L, Carrie I, Cantet C, Vellas B, Payoux P, Andrieu S. The impact of a multi-domain intervention on cerebral glucose metabolism: analysis from the randomized ancillary FDG PET MAPT trial. *Alzheimers Res Ther*. 2020 Oct 19;12(1):134. doi: 10.1186/s13195-020-00683-6. PMID: 33076983; PMCID: PMC7574215.

- 182: Valis M, Dlabkova A, Hort J, Angelucci F, Pejchal J, Kuca K, Pavelek Z, Karasova JZ, Novotny M. The measured CSF/plasma donepezil concentration ratio but not individually measured CSF and plasma concentrations significantly increase over 24 h after donepezil treatment in patients with Alzheimer's disease. *Biomed Pharmacother.* 2023 Mar;159:114223. doi: 10.1016/j.biopha.2023.114223. Epub 2023 Jan 9. PMID: 36630846.
- 183: Quintana-Hernández DJ, Rojas-Hernández J, Santana-Del Pino A, Céspedes Suárez C, Pellejero Silva M, Miró-Barrachina MT, Ibáñez Fernández I, Estupiñán López JA, Borkel LF. Mindfulness Prevents Depression and Psychopathology in Elderly People with Mild to Moderate Alzheimer's Disease: A Randomized Clinical Trial. *J Alzheimers Dis.* 2023;91(1):471-481. doi: 10.3233/JAD-220889. PMID: 36442199; PMCID: PMC9881020.
- 184: Fei Y, Wang R, Lu J, Peng S, Yang S, Wang Y, Zheng K, Li R, Lin L, Li M. Probiotic intervention benefits multiple neural behaviors in older adults with mild cognitive impairment. *Geriatr Nurs.* 2023 May-Jun;51:167-175. doi: 10.1016/j.gerinurse.2023.03.006. Epub 2023 Mar 28. PMID: 36990042.
- 185: Fernández-Blázquez MA, Del Ser T, Frades-Payo B, Ávila-Villanueva M, Valentí-Soler M, Benítez-Robredo MT, Bermejo-Aguña A, Pedrero-Pérez EJ, Quilis-Sancho J, Pastor AB, Fernández-Garrido C, Morales-Alonso S, Diaz-Olalla JM, Santos NC, Maestú F, Gómez-Ramírez J. MADRID+90 study on factors associated with longevity: Study design and preliminary data. *PLoS One.* 2021 May 17;16(5):e0251796. doi: 10.1371/journal.pone.0251796. PMID: 33999936; PMCID: PMC8128242.
- 186: Lu WH, Giudici KV, Rolland Y, Guyonnet S, Li Y, Bateman RJ, de Souto Barreto P, Vellas B. Prospective Associations between Plasma Amyloid-Beta 42/40 and Frailty in Community-Dwelling Older Adults. *J Prev Alzheimers Dis.* 2021;8(1):41-47. doi: 10.14283/jpad.2020.60. PMID: 33336223; PMCID: PMC9982745.
- 187: LaBarbera KM, Sheline YI, Izzo NJ, Yuede CM, Waybright L, Yurko R, Edwards HM, Gardiner WD, Blennow K, Zetterberg H, Börjesson-Hanson A, Morgan R, Davis CS, Guttendorf RJ, Schneider LS, DeKosky S, LeVine H 3rd, Grundman M, Caggiano AO, Cirrito JR, Catalano SM, Hamby ME. A phase 1b randomized clinical trial of CT1812 to measure A $\beta$  oligomer displacement in Alzheimer's disease using an indwelling CSF catheter. *Transl Neurodegener.* 2023 May 12;12(1):24. doi: 10.1186/s40035-023-00358-w. PMID: 37173791; PMCID: PMC10176668.
- 188: Sieber F, Neufeld K, Oh ES, Gottschalk A, Wang NY. Effect of baseline cognitive impairment on association between predicted propofol effect site concentration and Bispectral index or sedation score. *BMC Anesthesiol.* 2020 May 28;20(1):129. doi: 10.1186/s12871-020-01043-5. PMID: 32466776; PMCID: PMC7254641.
- 189: Krause-Sorio B, Siddarth P, Laird KT, Ercoli L, Narr K, Barrio JR, Small G, Lavretsky H. [<sup>18</sup>F]FDDNP PET binding predicts change in executive function in a pilot clinical trial of geriatric depression. *Int Psychogeriatr.* 2021 Feb;33(2):149-156. doi: 10.1017/S1041610219002047. Epub 2020 Jan 23. PMID: 31969201; PMCID: PMC7375908.
- 190: Lane HY, Tu CH, Lin WC, Lin CH. Brain Activity of Benzoate, a D-Amino Acid Oxidase Inhibitor, in Patients With Mild Cognitive Impairment in a Randomized, Double-Blind, Placebo Controlled Clinical Trial. *Int J Neuropsychopharmacol.*

2021 May 18;24(5):392-399. doi: 10.1093/ijnp/pyab001. PMID: 33406269; PMCID: PMC8130199.

191: Kirton MJ, Burnley MT, Ramos JS, Weatherwax R, Dalleck LC. The Effects of Standardised versus Individualised Aerobic Exercise Prescription on Fitness-Fatness Index in Sedentary Adults: A Randomised Controlled Trial. *J Sports Sci Med.* 2022 Sep 1;21(3):347-355. doi: 10.52082/jssm.2022.347. PMID: 36157386; PMCID: PMC9459763.

192: Wessels AM, Tariot PN, Zimmer JA, Selzler KJ, Bragg SM, Andersen SW, Landry J, Krull JH, Downing AM, Willis BA, Shcherbinin S, Mullen J, Barker P, Schumi J, Shering C, Matthews BR, Stern RA, Vellas B, Cohen S, MacSweeney E, Boada M, Sims JR. Efficacy and Safety of Lanabecestat for Treatment of Early and Mild Alzheimer Disease: The AMARANTH and DAYBREAK-ALZ Randomized Clinical Trials. *JAMA Neurol.* 2020 Feb 1;77(2):199-209. doi: 10.1001/jamaneurol.2019.3988. Erratum in: *JAMA Neurol.* 2020 Sep 1;77(9):1179. PMID: 31764959; PMCID: PMC6902191.

193: Deckers K, Köhler S, Ngandu T, Antikainen R, Laatikainen T, Soininen H, Strandberg T, Verhey F, Kivipelto M, Solomon A. Quantifying dementia prevention potential in the FINGER randomized controlled trial using the LIBRA prevention index. *Alzheimers Dement.* 2021 Jul;17(7):1205-1212. doi: 10.1002/alz.12281. Epub 2021 Jan 6. PMID: 33403822; PMCID: PMC8359273.

194: Kile S, Au W, Parise C, Rose K, Donnel T, Hankins A, Au Y, Chan M, Ghassemi A. Five-year outcomes after IVIG for mild cognitive impairment due to alzheimer disease. *BMC Neurosci.* 2021 Aug 6;22(1):49. doi: 10.1186/s12868-021-00651-2. PMID: 34362303; PMCID: PMC8349062.

195: Cao K, Bay AA, Hajjar I, Wharton W, Goldstein F, Qiu D, Prusin T, McKay JL, Perkins MM, Hackney ME. Rationale and Design of the PARTNER Trial: Partnered Rhythmic Rehabilitation for Enhanced Motor-Cognition in Prodromal Alzheimer's Disease. *J Alzheimers Dis.* 2023;91(3):1019-1033. doi: 10.3233/JAD-220783. PMID: 36530084; PMCID: PMC10105523.

196: Hampstead BM, Stringer AY, Stilla RF, Sathian K. Mnemonic strategy training increases neocortical activation in healthy older adults and patients with mild cognitive impairment. *Int J Psychophysiol.* 2020 Aug;154:27-36. doi: 10.1016/j.ijpsycho.2019.04.011. Epub 2019 May 5. PMID: 31067489; PMCID: PMC7218756.

197: Yazdanmanesh M, Esmaili R, Nasiri M, Vasli P. Relieving care burden and promoting health-related quality of life for family caregivers of elderly people with Alzheimer's disease via an empowerment program. *Aging Clin Exp Res.* 2023 Jan;35(1):73-83. doi: 10.1007/s40520-022-02277-8. Epub 2022 Oct 25. PMID: 36282451; PMCID: PMC9593982.

198: Schurad B, Koch C, Schug B, Morte A, Vaqué A, De la Torre R, Iniesta M. Comparative Bioavailability Study of a Novel Multi-Day Patch Formulation of Rivastigmine (Twice Weekly) with Exelon® Transdermal Patch (Daily)- A Randomized Clinical Trial. *Curr Alzheimer Res.* 2022;19(7):541-553. doi: 10.2174/1567205019666220823105059. PMID: 36017827; PMCID: PMC10186381.

199: Conley AC, Albert KM, McDonald BC, Saykin AJ, Dumas JA, Newhouse PA. Estradiol treatment in young postmenopausal women with self-reported cognitive complaints: Effects on cholinergic-mediated cognitive performance. *Hum*

Psychopharmacol. 2022 Sep;37(5):e2838. doi: 10.1002/hup.2838. Epub 2022 Feb 24. PMID: 35212023; PMCID: PMC9399322.

200: Bakker C, Tasker T, Liptrot J, Hart EP, Klaassen ES, Prins S, van der Doef TF, Brown GA, Brown A, Congreve M, Weir M, Marshall FH, Cross DM, Groeneveld GJ, Nathan PJ. First-in-man study to investigate safety, pharmacokinetics and exploratory pharmacodynamics of HTL0018318, a novel M<sub>1</sub>-receptor partial agonist for the treatment of dementias. Br J Clin Pharmacol. 2021 Jul;87(7):2945-2955. doi: 10.1111/bcp.14710. Epub 2021 Feb 5. PMID: 33351971; PMCID: PMC8359307.

201: Lingler JH, Sereika SM, Butters MA, Cohen AD, Klunk WE, Knox ML, McDade E, Nadkarni NK, Roberts JS, Tamres LK, Lopez OL. A randomized controlled trial of amyloid positron emission tomography results disclosure in mild cognitive impairment. Alzheimers Dement. 2020 Sep;16(9):1330-1337. doi: 10.1002/alz.12129. Epub 2020 Jun 26. PMID: 32588971; PMCID: PMC7541680.

202: Bakker C, Prins S, Liptrot J, Hart EP, Klaassen ES, Brown GA, Brown A, Congreve M, Weir M, Marshall FH, Stevens J, Cross DM, Tasker T, Nathan PJ, Groeneveld GJ. Safety, pharmacokinetics and pharmacodynamics of HTL0009936, a selective muscarinic M<sub>1</sub>-acetylcholine receptor agonist: A randomized cross-over trial. Br J Clin Pharmacol. 2021 Nov;87(11):4439-4449. doi: 10.1111/bcp.14872. Epub 2021 May 8. PMID: 33891333; PMCID: PMC8596821.

203: Varastegan S, Kazemi R, Rostami R, Khomami S, Zandbagleh A, Hadipour AL. Remember NIBS? tACS improves memory performance in elders with subjective memory complaints. Geroscience. 2023 Apr;45(2):851-869. doi: 10.1007/s11357-022-00677-2. Epub 2022 Oct 22. PMID: 36272055; PMCID: PMC9886712.

204: Platen M, Fleßa S, Rädke A, Wucherer D, Thyrian JR, Mohr W, Scharf A, Mühlichen F, Hoffmann W, Michalowsky B. Prevalence of Low-Value Care and Its Associations with Patient-Centered Outcomes in Dementia. J Alzheimers Dis. 2021;83(4):1775-1787. doi: 10.3233/JAD-210439. PMID: 34459396; PMCID: PMC8609693.

205: Vegue Parra E, Hernández Garre JM, Echevarría Pérez P. Benefits of Dog-Assisted Therapy in Patients with Dementia Residing in Aged Care Centers in Spain. Int J Environ Res Public Health. 2021 Feb 4;18(4):1471. doi: 10.3390/ijerph18041471. PMID: 33557254; PMCID: PMC7914582.

206: Scullin MK, Jones WE, Phenis R, Beevers S, Rosen S, Dinh K, Kiselica A, Keefe FJ, Bengtson JF. Using smartphone technology to improve prospective memory functioning: A randomized controlled trial. J Am Geriatr Soc. 2022 Feb;70(2):459-469. doi: 10.1111/jgs.17551. Epub 2021 Nov 17. PMID: 34786698; PMCID: PMC8821124.

207: Taylor WD, Boyd BD, Elson D, Andrews P, Albert K, Vega J, Newhouse PA, Woodward ND, Kang H, Shokouhi S. Preliminary Evidence That Cortical Amyloid Burden Predicts Poor Response to Antidepressant Medication Treatment in Cognitively Intact Individuals With Late-Life Depression. Am J Geriatr Psychiatry. 2021 May;29(5):448-457. doi: 10.1016/j.jagp.2020.09.019. Epub 2020 Sep 28. PMID: 33032927; PMCID: PMC8004530.

208: Pitchford LM, Driver PM, Fuller JC Jr, Akers WS, Abumrad NN, Amarnath V, Milne GL, Chen SC, Ye F, Roberts LJ 2nd, Shoemaker MB, Oates JA, Rathmacher JA, Boutaud O. Safety, tolerability, and pharmacokinetics of repeated oral doses of



2-hydroxybenzylamine acetate in healthy volunteers: a double-blind, randomized, placebo-controlled clinical trial. *BMC Pharmacol Toxicol.* 2020 Jan 6;21(1):3. doi: 10.1186/s40360-020-0382-y. PMID: 31907026; PMCID: PMC6945443.

209: Hives BA, Buckler EJ, Weiss J, Schilf S, Johansen KL, Epel ES, Puterman E. The Effects of Aerobic Exercise on Psychological Functioning in Family Caregivers: Secondary Analyses of a Randomized Controlled Trial. *Ann Behav Med.* 2021 Feb 12;55(1):65-76. doi: 10.1093/abm/kaaa031. PMID: 32421163; PMCID: PMC7880222.

210: Richards K, Morrison J, Wang YY, Rangel A, Loera A, Hanlon A, Lozano A, Kovach C, Gooneratne N, Fry L, Allen R. Nighttime Agitation and Restless Legs Syndrome in Persons With Alzheimer's Disease: Study Protocol for a Double-Blind, Placebo-Controlled, Randomized Trial (NightRest). *Res Gerontol Nurs.* 2020 Nov 1;13(6):280-288. doi: 10.3928/19404921-20200918-01. Epub 2020 Sep 24. PMID: 32966585; PMCID: PMC9112187.

211: Family N, Maillet EL, Williams LTJ, Krediet E, Carhart-Harris RL, Williams TM, Nichols CD, Goble DJ, Raz S. Safety, tolerability, pharmacokinetics, and pharmacodynamics of low dose lysergic acid diethylamide (LSD) in healthy older volunteers. *Psychopharmacology (Berl).* 2020 Mar;237(3):841-853. doi: 10.1007/s00213-019-05417-7. Epub 2019 Dec 18. PMID: 31853557; PMCID: PMC7036065.

212: Kim JH, Cho MR, Shin JC, Park GC, Lee JS. Factors contributing to cognitive improvement effects of acupuncture in patients with mild cognitive impairment: a pilot randomized controlled trial. *Trials.* 2021 May 12;22(1):341. doi: 10.1186/s13063-021-05296-4. PMID: 33980288; PMCID: PMC8117619.

213: Martin-Willett R, Ellingson JE, Fries J, Helmuth T, Karoly H, Giordano G, Calhoun VD, Bryan AD. Few Structural Brain Changes Associated With Moderate-Intensity Interval Training and Low-Intensity Continuous Training in a Randomized Trial of Fitness and Older Adults. *J Aging Phys Act.* 2021 Jun 1;29(3):505-515. doi: 10.1123/japa.2019-0352. Epub 2020 Dec 2. PMID: 33271506; PMCID: PMC9092807.

214: Breier A, Brannan SK, Paul SM, Miller AC. Evidence of trospium's ability to mitigate cholinergic adverse events related to xanomeline: phase 1 study results. *Psychopharmacology (Berl).* 2023 May;240(5):1191-1198. doi: 10.1007/s00213-023-06362-2. Epub 2023 Apr 10. PMID: 37036495; PMCID: PMC10102054.

215: Park KS, Ganesh AB, Berry NT, Mobley YP, Karper WB, Labban JD, Wahlheim CN, Williams TM, Wideman L, Etnier JL. The effect of physical activity on cognition relative to APOE genotype (PAAD-2): study protocol for a phase II randomized control trial. *BMC Neurol.* 2020 Jun 6;20(1):231. doi: 10.1186/s12883-020-01732-1. PMID: 32503473; PMCID: PMC7274941.

216: Brewster GS, Epps F, Dye CE, Hepburn K, Higgins MK, Parker ML. The Effect of the "Great Village" on Psychological Outcomes, Burden, and Mastery in African American Caregivers of Persons Living With Dementia. *J Appl Gerontol.* 2020 Oct;39(10):1059-1068. doi: 10.1177/0733464819874574. Epub 2019 Sep 19. PMID: 31535922; PMCID: PMC7080571.

217: Rogalski E, Roberts A, Salley E, Saxon M, Fought A, Esparza M, Blaze E, Coventry C, Mesulam MM, Weintraub S, Mooney A, Khayum B, Rademaker A. Communication Partner Engagement: A Relevant Factor for Functional Outcomes in

Speech-Language Therapy for Aphasic Dementia. *J Gerontol B Psychol Sci Soc Sci*. 2022 Jun 1;77(6):1017-1025. doi: 10.1093/geronb/gbab165. PMID: 34528692; PMCID: PMC9159055.

218: Zhong GC, Hu TY, Yang PF, Peng Y, Wu JJ, Sun WP, Cheng L, Wang CR. Chocolate consumption and all-cause and cause-specific mortality in a US population: a post hoc analysis of the PLCO cancer screening trial. *Aging (Albany NY)*. 2021 Jul 29;13(14):18564-18585. doi: 10.18632/aging.203302. Epub 2021 Jul 29. PMID: 34329196; PMCID: PMC8351724.

219: Shadyab AH, LaCroix AZ, Feldman HH, van Dyck CH, Okonkwo OC, Tam SP, Fairchild JK, Welsh-Bohmer KA, Matthews G, Bennett D, Shadyab AA, Schafer KA, Morrison RH, Kipperman SA, Mason J, Tan D, Thomas RG, Cotman CW, Baker LD; ADCS EXERT Study Group. Recruitment of a multi-site randomized controlled trial of aerobic exercise for older adults with amnesic mild cognitive impairment: The EXERT trial. *Alzheimers Dement*. 2021 Nov;17(11):1808-1817. doi: 10.1002/alz.12401. Epub 2021 Jul 23. PMID: 34297895; PMCID: PMC9292825.

220: Sandebring-Matton A, Goikolea J, Björkhem I, Paternain L, Kemppainen N, Laatikainen T, Ngandu T, Rinne J, Soininen H, Cedazo-Minguez A, Solomon A, Kivipelto M. 27-Hydroxycholesterol, cognition, and brain imaging markers in the FINGER randomized controlled trial. *Alzheimers Res Ther*. 2021 Mar 6;13(1):56. doi: 10.1186/s13195-021-00790-y. PMID: 33676572; PMCID: PMC7937194.

221: Li W, Wen Q, Xie YH, Hu AL, Wu Q, Wang YX. Improvement of poststroke cognitive impairment by intermittent theta bursts: A double-blind randomized controlled trial. *Brain Behav*. 2022 Jun;12(6):e2569. doi: 10.1002/brb3.2569. Epub 2022 Apr 29. PMID: 35484991; PMCID: PMC9226849.

222: Yang L, Xuan C, Yu C, Jin X, Zheng P, Yan J. Effects of comprehensive intervention on life quality among the elderly with Alzheimer Disease and their caregivers based on mixed models. *Nurs Open*. 2022 Mar;9(2):1412-1422. doi: 10.1002/nop2.917. Epub 2021 May 14. PMID: 33988909; PMCID: PMC8859043.

223: Sauder C, Allen LA, Baker E, Miller AC, Paul SM, Brannan SK. Effectiveness of KarXT (xanomeline-trospium) for cognitive impairment in schizophrenia: post hoc analyses from a randomised, double-blind, placebo-controlled phase 2 study. *Transl Psychiatry*. 2022 Nov 21;12(1):491. doi: 10.1038/s41398-022-02254-9. PMID: 36414626; PMCID: PMC9681874.

224: Tawfik NM, Sabry NA, Darwish H, Mowafy M, Soliman SSA. Psychoeducational Program for the Family Member Caregivers of People with Dementia to Reduce Perceived Burden and Increase Patient's Quality of Life: A Randomized Controlled Trial. *J Prim Care Community Health*. 2021 Jan-Dec;12:21501327211014088. doi: 10.1177/21501327211014088. PMID: 33971764; PMCID: PMC8120530.

225: Courtois-Amiot P, Cloppet-Fontaine A, Poissonnet A, Benit E, Dautet M, Raynaud-Simon A, Paquet C, Lilamand M. Hypnosis for pain and anxiety management in cognitively impaired older adults undergoing scheduled lumbar punctures: a randomized controlled pilot study. *Alzheimers Res Ther*. 2022 Sep 2;14(1):120. doi: 10.1186/s13195-022-01065-w. PMID: 36056417; PMCID: PMC9438329.

226: Xu L, Fields NL, Williams IC, Gaugler JE, Kunz-Lomelin A, Cipher DJ, Feinhs G. The Senior Companion Program Plus (SCP Plus): Examining the Preliminary Effectiveness of a Lay Provider Program to Support African American Alzheimer's Disease and Related Dementias (ADRD) Caregivers. *Int J Environ Res*

Public Health. 2023 Apr 3;20(7):5380. doi: 10.3390/ijerph20075380. PMID: 37047994; PMCID: PMC10094539.

227: Teles S, Ferreira A, Paúl C. Feasibility of an online training and support program for dementia carers: results from a mixed-methods pilot randomized controlled trial. BMC Geriatr. 2022 Mar 1;22(1):173. doi: 10.1186/s12877-022-02831-z. PMID: 35232389; PMCID: PMC8887647.

228: Tomoto T, Tarumi T, Chen JN, Hynan LS, Cullum CM, Zhang R. One-year aerobic exercise altered cerebral vasomotor reactivity in mild cognitive impairment. J Appl Physiol (1985). 2021 Jul 1;131(1):119-130. doi: 10.1152/jappphysiol.00158.2021. Epub 2021 May 20. PMID: 34013755; PMCID: PMC8325610.

229: Yoshida K, Moein A, Bittner T, Ostrowitzki S, Lin H, Honigberg L, Jin JY, Quartino A. Pharmacokinetics and pharmacodynamic effect of crenezumab on plasma and cerebrospinal fluid beta-amyloid in patients with mild-to-moderate Alzheimer's disease. Alzheimers Res Ther. 2020 Jan 22;12(1):16. doi: 10.1186/s13195-020-0580-2. PMID: 31969177; PMCID: PMC6977279.

230: Sheng L, Cao W, Lin P, Chen W, Xu H, Zhong C, Yuan F, Chen H, Li H, Liu C, Yang M, Li X. Safety, Tolerability and Pharmacokinetics of Single and Multiple Ascending Doses of Benfotiamine in Healthy Subjects. Drug Des Devel Ther. 2021 Mar 9;15:1101-1110. doi: 10.2147/DDDT.S296197. PMID: 33727798; PMCID: PMC7955752.

231: Portron A, Jordan P, Draper K, Muenzer C, Dickerson D, van Iersel T, Hofmann C. A Phase I Study to Assess the Effect of Speed of Injection on Pain, Tolerability, and Pharmacokinetics After High-volume Subcutaneous Administration of Gantenerumab in Healthy Volunteers. Clin Ther. 2020 Jan;42(1):108-120.e1. doi: 10.1016/j.clinthera.2019.11.015. Epub 2019 Dec 26. PMID: 31883703.

232: Shin HY, Kim HR, Jahng GH, Jin C, Kwon S, Cho SY, Park SU, Jung WS, Moon SK, Ko CN, Park JM. Efficacy and safety of Kami-guibi-tang for mild cognitive impairment: a pilot, randomized, double-blind, placebo-controlled trial. BMC Complement Med Ther. 2021 Oct 7;21(1):251. doi: 10.1186/s12906-021-03428-6. PMID: 34620151; PMCID: PMC8495912.

233: Stephen R, Solomon A, Ngandu T, Levälähti E, Rinne JO, Kemppainen N, Parkkola R, Antikainen R, Strandberg T, Kivipelto M, Soininen H, Liu Y; FINGER study group. White Matter Changes on Diffusion Tensor Imaging in the FINGER Randomized Controlled Trial. J Alzheimers Dis. 2020;78(1):75-86. doi: 10.3233/JAD-200423. PMID: 32925045; PMCID: PMC7683078.

234: Bakker C, Tasker T, Liptrot J, Hart EP, Klaassen ES, Doll RJ, Brown GA, Brown A, Congreve M, Weir M, Marshall FH, Cross DM, Groeneveld GJ, Nathan PJ. Safety, pharmacokinetics and exploratory pro-cognitive effects of HTL0018318, a selective M<sub>1</sub> receptor agonist, in healthy younger adult and elderly subjects: a multiple ascending dose study. Alzheimers Res Ther. 2021 Apr 21;13(1):87. doi: 10.1186/s13195-021-00816-5. PMID: 33883008; PMCID: PMC8061066.

235: Men X, Marshe V, Elsheikh SS, Alexopoulos GS, Marino P, Meyers BS, Mulsant BH, Rothschild AJ, Voineskos AN, Whyte EM, Kennedy JL, Flint AJ, Müller DJ. Genomic Investigation of Remission and Relapse of Psychotic Depression Treated with Sertraline plus Olanzapine: The STOP-PD II Study. Neuropsychobiology. 2023;82(3):168-178. doi: 10.1159/000529637. Epub 2023 Apr 4. PMID: 37015192.

236: Schmitter-Edgecombe M, Brown K, Luna C, Chilton R, Sumida CA, Holder L, Cook D. Partnering a Compensatory Application with Activity-Aware Prompting to Improve Use in Individuals with Amnesic Mild Cognitive Impairment: A Randomized Controlled Pilot Clinical Trial. *J Alzheimers Dis.* 2022;85(1):73-90. doi: 10.3233/JAD-215022. PMID: 34776442; PMCID: PMC922794.

237: Chen J, Ma N, Hu G, Nousayhah A, Xue C, Qi W, Xu W, Chen S, Rao J, Liu W, Zhang F, Zhang X. rTMS modulates precuneus-hippocampal subregion circuit in patients with subjective cognitive decline. *Aging (Albany NY).* 2020 Nov 30;13(1):1314-1331. doi: 10.18632/aging.202313. Epub 2020 Nov 30. PMID: 33260151; PMCID: PMC7835048.

238: Yin W, Mamashli F, Buhl DL, Khudyakov P, Volfson D, Martenyi F, Gevorkyan H, Rosen L, Simen AA. Safety, pharmacokinetics and quantitative EEG modulation of TAK-071, a novel muscarinic M1 receptor positive allosteric modulator, in healthy subjects. *Br J Clin Pharmacol.* 2022 Feb;88(2):600-612. doi: 10.1111/bcp.14975. Epub 2021 Jul 31. PMID: 34240455; PMCID: PMC9291057.

239: Belleville S, Cloutier S, Mellah S, Willis S, Vellas B, Andrieu S, Coley N, Ngandu T; MAPT/DSA group. Is more always better? Dose effect in a multidomain intervention in older adults at risk of dementia. *Alzheimers Dement.* 2022 Nov;18(11):2140-2150. doi: 10.1002/alz.12544. Epub 2022 Jan 20. PMID: 35049127; PMCID: PMC9786573.

240: Whitfield T, Demnitz-King H, Schlosser M, Barnhofer T, Frison E, Coll-Padros N, Dautricourt S, Requier F, Delarue M, Gonneaud J, Klimecki OM, Lutz A, Paly L, Salmon E, Schild AK, Walker Z, Jessen F, Chételat G, Collette F, Wirth M, Marchant NL; Medit-Ageing Research Group. Effects of a mindfulness-based versus a health self-management intervention on objective cognitive performance in older adults with subjective cognitive decline (SCD): a secondary analysis of the SCD-Well randomized controlled trial. *Alzheimers Res Ther.* 2022 Sep 6;14(1):125. doi: 10.1186/s13195-022-01057-w. PMID: 36068621; PMCID: PMC9446839.

241: Kanatome A, Ano Y, Shinagawa K, Ide Y, Shibata M, Umeda S.  $\beta$ -Lactolin Enhances Neural Activity, Indicated by Event-Related P300 Amplitude, in Healthy Adults: A Randomized Controlled Trial. *J Alzheimers Dis.* 2021;81(2):787-796. doi: 10.3233/JAD-201413. PMID: 33814437; PMCID: PMC8203246.

242: Yang J, Zhou J, Zhou J, Wang H, Sun Z, Zhu X, He Y, Wong AHC, Liu F, Wang G. Serum amyloid P component level is associated with clinical response to escitalopram treatment in patients with major depressive disorder. *J Psychiatr Res.* 2022 Feb;146:172-178. doi: 10.1016/j.jpsychires.2021.12.051. Epub 2021 Dec 30. PMID: 34995992.

243: Molet-Benhamou L, Giudici KV, de Souto Barreto P, Cantet C, Rolland Y; MAPT/DSA group. Association between urate-lowering therapies and cognitive decline in community-dwelling older adults. *Sci Rep.* 2022 Sep 12;12(1):15299. doi: 10.1038/s41598-022-17808-6. PMID: 36097005; PMCID: PMC9467991.

244: Morgan-Trimmer S, Kudlicka A, Warmoth K, Leroi I, Oyebode JR, Pool J, Woods R, Clare L. Implementation processes in a cognitive rehabilitation intervention for people with dementia: a complexity-informed qualitative analysis. *BMJ Open.* 2021 Oct 26;11(10):e051255. doi: 10.1136/bmjopen-2021-051255. PMID: 34702730; PMCID: PMC8549661.

245: Giudici KV, de Souto Barreto P, Guyonnet S, Li Y, Bateman RJ, Vellas B; MAPT/DSA Group. Assessment of Plasma Amyloid- $\beta$ 42/40 and Cognitive Decline Among Community-Dwelling Older Adults. *JAMA Netw Open*. 2020 Dec 1;3(12):e2028634. doi: 10.1001/jamanetworkopen.2020.28634. PMID: 33331917; PMCID: PMC7747018.

246: Tonga JB, Eilertsen DE, Solem IKL, Arnevik EA, Korsnes MS, Ulstein ID. Effect of Self-Efficacy on Quality of Life in People With Mild Cognitive Impairment and Mild Dementia: The Mediating Roles of Depression and Anxiety. *Am J Alzheimers Dis Other Dement*. 2020 Jan-Dec;35:1533317519885264. doi: 10.1177/1533317519885264. PMID: 31916847.

247: Hepburn K, Nocera J, Higgins M, Epps F, Brewster GS, Lindauer A, Morhardt D, Shah R, Bonds K, Nash R, Griffiths PC. Results of a Randomized Trial Testing the Efficacy of Tele-Savvy, an Online Synchronous/Asynchronous Psychoeducation Program for Family Caregivers of Persons Living with Dementia. *Gerontologist*. 2022 Apr 20;62(4):616-628. doi: 10.1093/geront/gnab029. PMID: 33640979; PMCID: PMC7989248.

248: Zhu X, Borenstein AR, Zheng Y, Zhang W, Seidner DL, Ness R, Murff HJ, Li B, Shrubsole MJ, Yu C, Hou L, Dai Q. Ca:Mg Ratio, APOE Cytosine Modifications, and Cognitive Function: Results from a Randomized Trial. *J Alzheimers Dis*. 2020;75(1):85-98. doi: 10.3233/JAD-191223. PMID: 32280092; PMCID: PMC7737669.

249: Rafii MS, Sol O, Mobley WC, Delpretti S, Skotko BG, Burke AD, Sabbagh MN, Yuan SH, Rissman RA, Pulsifer M, Evans C, Evans AC, Beth G, Fournier N, Gray JA, Dos Santos AM, Hliva V, Vukicevic M, Kosco-Vilbois M, Streffer J, Pfeifer A, Feldman HH. Safety, Tolerability, and Immunogenicity of the ACI-24 Vaccine in Adults With Down Syndrome: A Phase 1b Randomized Clinical Trial. *JAMA Neurol*. 2022 Jun 1;79(6):565-574. doi: 10.1001/jamaneurol.2022.0983. PMID: 35532913; PMCID: PMC9086937.

250: Roberts AC, Rademaker AW, Salley EA, Mooney A, Morhardt D, Fried-Oken M, Weintraub S, Mesulam M, Rogalski E. Communication Bridge™-2 (CB2): an NIH Stage 2 randomized control trial of a speech-language intervention for communication impairments in individuals with mild to moderate primary progressive aphasia. *Trials*. 2022 Jun 13;23(1):487. doi: 10.1186/s13063-022-06162-7. PMID: 35698099; PMCID: PMC9190461.

251: Jung SJ, Jung ES, Ha KC, Baek HI, Park YK, Han SK, Chae SW, Lee SO, Chung YC. Efficacy and Safety of Sesame Oil Cake Extract on Memory Function Improvement: A 12-Week, Randomized, Double-Blind, Placebo-Controlled Pilot Study. *Nutrients*. 2021 Jul 28;13(8):2606. doi: 10.3390/nu13082606. PMID: 34444766; PMCID: PMC8399671.

252: de Sousa AVC, Grittner U, Rujescu D, Külzow N, Flöel A. Impact of 3-Day Combined Anodal Transcranial Direct Current Stimulation-Visuospatial Training on Object-Location Memory in Healthy Older Adults and Patients with Mild Cognitive Impairment. *J Alzheimers Dis*. 2020;75(1):223-244. doi: 10.3233/JAD-191234. PMID: 32280093; PMCID: PMC7306891.

253: Lee WJ, Shin YW, Kim DE, Kweon MH, Kim M. Effect of desalted *Salicornia europaea* L. ethanol extract (PM-EE) on the subjects complaining memory dysfunction without dementia: a 12 week, randomized, double-blind, placebo-controlled clinical trial. *Sci Rep*. 2020 Nov 16;10(1):19914. doi: 10.1038/s41598-020-76938-x. PMID: 33199752; PMCID: PMC7670404.

254: Sangaletti CT, Katayama KY, De Angelis K, Lemos de Moraes T, Araújo AA, Lopes HF, Camacho C, Bortolotto LA, Michelini LC, Irigoyen MC, Olofsson PS, Barnaby DP, Tracey KJ, Pavlov VA, Consolim Colombo FM. The Cholinergic Drug Galantamine Alleviates Oxidative Stress Alongside Anti-inflammatory and Cardio-Metabolic Effects in Subjects With the Metabolic Syndrome in a Randomized Trial. *Front Immunol.* 2021 Mar 11;12:613979. doi: 10.3389/fimmu.2021.613979. PMID: 33776997; PMCID: PMC7991724.

255: Coley N, Coniasse-Brioude D, Igier V, Fournier T, Poulain JP, Andrieu S; ACCEPT study group. Disparities in the participation and adherence of older adults in lifestyle-based multidomain dementia prevention and the motivational role of perceived disease risk and intervention benefits: an observational ancillary study to a randomised controlled trial. *Alzheimers Res Ther.* 2021 Sep 24;13(1):157. doi: 10.1186/s13195-021-00904-6. PMID: 34560903; PMCID: PMC8464095.

256: Smith NI, Martinez AI, Huffmyer M, Eckmann L, George R, Abner EL, Jicha GA, Moga DC. Acceptability of patient-centered, multi-disciplinary medication therapy management recommendations: results from the INCREASE randomized study. *BMC Geriatr.* 2023 Mar 10;23(1):137. doi: 10.1186/s12877-023-03876-4. PMID: 36894900; PMCID: PMC9999619.

257: Rosenbloom M, Barclay T, Johnsen J, Erickson L, Svitak A, Pyle M, Frey W, Hanson LR. Double-Blind Placebo-Controlled Pilot Investigation of the Safety of a Single Dose of Rapid-Acting Intranasal Insulin in Down Syndrome. *Drugs R D.* 2020 Mar;20(1):11-15. doi: 10.1007/s40268-020-00296-2. PMID: 32077057; PMCID: PMC7067723.

258: Bristol AA, Convery KA, Sotelo V, Schneider CE, Lin SY, Fletcher J, Rupper R, Galvin JE, Brody AA. Protocol for an embedded pragmatic clinical trial to test the effectiveness of Aliviado Dementia Care in improving quality of life for persons living with dementia and their informal caregivers. *Contemp Clin Trials.* 2020 Jun;93:106005. doi: 10.1016/j.cct.2020.106005. Epub 2020 Apr 19. PMID: 32320844; PMCID: PMC7269690.

259: André C, Rehel S, Kuhn E, Landeau B, Moulinet I, Touron E, Ourry V, Le Du G, Mézenge F, Tomadesso C, de Flores R, Bejanin A, Sherif S, Delcroix N, Manrique A, Abbas A, Marchant NL, Lutz A, Klimecki OM, Collette F, Arenaza-Urquijo EM, Poisnel G, Vivien D, Bertran F, de la Sayette V, Chételat G, Rauchs G; Medit-Ageing Research Group. Association of Sleep-Disordered Breathing With Alzheimer Disease Biomarkers in Community-Dwelling Older Adults: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Neurol.* 2020 Jun 1;77(6):716-724. doi: 10.1001/jamaneurol.2020.0311. PMID: 32202593; PMCID: PMC7091393.

260: Clare L, Kudlicka A, Collins R, Evans S, Pool J, Henderson C, Knapp M, Litherland R, Oyebode J, Woods R. Implementing a home-based personalised cognitive rehabilitation intervention for people with mild-to-moderate dementia: GREAT into Practice. *BMC Geriatr.* 2023 Feb 13;23(1):93. doi: 10.1186/s12877-022-03705-0. PMID: 36782122; PMCID: PMC9925212.

261: Williams KN, Shaw CA, Perkhounkova Y, Hein M, Coleman CK. Satisfaction, utilization, and feasibility of a telehealth intervention for in-home dementia care support: A mixed methods study. *Dementia (London).* 2021 Jul;20(5):1565-1585. doi: 10.1177/1471301220957905. Epub 2020 Sep 9. PMID: 32902313; PMCID: PMC7940465.

262: Krause-Sorio B, Siddarth P, Kilpatrick L, Laird KT, Milillo MM, Ercoli L, Narr KL, Lavretsky H. Combined treatment with escitalopram and memantine increases gray matter volume and cortical thickness compared to escitalopram and placebo in a pilot study of geriatric depression. *J Affect Disord.* 2020 Sep 1;274:464-470. doi: 10.1016/j.jad.2020.05.092. Epub 2020 May 24. PMID: 32663977; PMCID: PMC7368564.

263: Wang Y, Hernandez G, Mack WJ, Schneider LS, Yin F, Brinton RD. Retrospective analysis of phytoSERM for management of menopause-associated vasomotor symptoms and cognitive decline: a pilot study on pharmacogenomic effects of mitochondrial haplogroup and APOE genotype on therapeutic efficacy. *Menopause.* 2020 Jan;27(1):57-65. doi: 10.1097/GME.0000000000001418. PMID: 31567873; PMCID: PMC7100617.

264: Lee YC, Sousa TV, Stretton-Smith PA, Gold C, Geretsegger M, Baker FA. Demographic and clinical profile of residents living with dementia and depressive symptoms in Australian private residential aged care: Data from the Music Interventions for Dementia and Depression in ELderly care (MIDDEL) cluster-randomised controlled trial. *Australas J Ageing.* 2022 Dec;41(4):e387-e396. doi: 10.1111/ajag.13104. Epub 2022 Jul 8. PMID: 35801957; PMCID: PMC10084363.

265: Ogawa T, Sawane K, Ookoshi K, Kawashima R. Supplementation with Flaxseed Oil Rich in Alpha-Linolenic Acid Improves Verbal Fluency in Healthy Older Adults. *Nutrients.* 2023 Mar 21;15(6):1499. doi: 10.3390/nu15061499. PMID: 36986229; PMCID: PMC10056498.

266: Thela L, Decloedt E, Zetterberg H, Gisslén M, Lesosky M, Gleich M, Koutsilieris E, Scheller C, Hye A, Joska J. Blood and cerebrospinal fluid biomarker changes in patients with HIV-associated neurocognitive impairment treated with lithium: analysis from a randomised placebo-controlled trial. *J Neurovirol.* 2023 Apr;29(2):156-166. doi: 10.1007/s13365-023-01116-4. Epub 2023 Feb 15. PMID: 36790601; PMCID: PMC10185609.

267: Xu L, Yu H, Sun H, Hu B, Geng Y. Dietary Melatonin Therapy Alleviates the Lamina Cribrosa Damages in Patients with Mild Cognitive Impairments: A Double-Blinded, Randomized Controlled Study. *Med Sci Monit.* 2020 May 7;26:e923232. doi: 10.12659/MSM.923232. PMID: 32376818; PMCID: PMC7233010.

268: Bradley K, Smith R, Hughson JA, Atkinson D, Bessarab D, Flicker L, Radford K, Smith K, Strivens E, Thompson S, Blackberry I, LoGiudice D. Let's CHAT (community health approaches to) dementia in Aboriginal and Torres Strait Islander communities: protocol for a stepped wedge cluster randomised controlled trial. *BMC Health Serv Res.* 2020 Mar 12;20(1):208. doi: 10.1186/s12913-020-4985-1. PMID: 32164678; PMCID: PMC7069169.

269: Habes M, Jacobson AM, Braffett BH, Rashid T, Ryan CM, Shou H, Cui Y, Davatzikos C, Luchsinger JA, Biessels GJ, Bebu I, Gubitosi-Klug RA, Bryan RN, Nasrallah IM; DCCT/EDIC Research Group. Patterns of Regional Brain Atrophy and Brain Aging in Middle- and Older-Aged Adults With Type 1 Diabetes. *JAMA Netw Open.* 2023 Jun 1;6(6):e2316182. doi: 10.1001/jamanetworkopen.2023.16182. PMID: 37261829; PMCID: PMC10236234.

270: Ashton NJ, Moseby-Knappe M, Benedet AL, Grötschel L, Lantero-Rodriguez J, Karikari TK, Hassager C, Wise MP, Stammet P, Kjaergaard J, Friberg H, Nielsen N, Cronberg T, Zetterberg H, Blennow K. Alzheimer Disease Blood Biomarkers in

Patients With Out-of-Hospital Cardiac Arrest. JAMA Neurol. 2023 Apr 1;80(4):388-396. doi: 10.1001/jamaneurol.2023.0050. PMID: 36877496; PMCID: PMC9989959.

271: Surr CA, Holloway I, Walwyn REA, Griffiths AW, Meads D, Martin A, Kelley R, Ballard C, Fossey J, Burnley N, Chenoweth L, Creese B, Downs M, Garrod L, Graham EH, Lilley-Kelly A, McDermid J, McLellan V, Millard H, Perfect D, Robinson L, Robinson O, Shoosmith E, Siddiqi N, Stokes G, Wallace D, Farrin AJ. Effectiveness of Dementia Care Mapping™ to reduce agitation in care home residents with dementia: an open-cohort cluster randomised controlled trial. Aging Ment Health. 2021 Aug;25(8):1410-1423. doi: 10.1080/13607863.2020.1745144. Epub 2020 Apr 13. PMID: 32279541.

272: Stephen R, Ngandu T, Liu Y, Peltonen M, Antikainen R, Kemppainen N, Laatikainen T, Lötjönen J, Rinne J, Strandberg T, Tuomilehto J, Vanninen R, Soininen H, Kivipelto M, Solomon A; FINGER Study Group. Change in CAIDE Dementia Risk Score and Neuroimaging Biomarkers During a 2-Year Multidomain Lifestyle Randomized Controlled Trial: Results of a Post-Hoc Subgroup Analysis. J Gerontol A Biol Sci Med Sci. 2021 Jul 13;76(8):1407-1414. doi: 10.1093/gerona/glab130. PMID: 33970268; PMCID: PMC8277089.

273: Damiano RF, Loureiro JC, Pais MV, Pereira RF, Corradi MM, Di Santi T, Bezerra GAM, Radanovic M, Talib LL, Forlenza OV. Revisiting global cognitive and functional state 13 years after a clinical trial of lithium for mild cognitive impairment. Braz J Psychiatry. 2023 Mar 11;45(1):46-49. doi: 10.47626/1516-4446-2022-2767. PMID: 36049127; PMCID: PMC9976922.

274: Dauphinot V, Boublay N, Moutet C, Achi S, Bathsavanis A, Krolak-Salmon P. Comparison of Instrumental Activities of Daily Living assessment by face-to-face or telephone interviews: a randomized, crossover study. Alzheimers Res Ther. 2020 Mar 13;12(1):24. doi: 10.1186/s13195-020-00590-w. PMID: 32169093; PMCID: PMC7068883.

275: Saif N, Niotis K, Dominguez M, Hodes JF, Woodbury M, Amini Y, Sadek G, Scheyer O, Caesar E, Hristov H, Knowlton N, Lee P, McInnis M, Isaacson RS. Education Research: Online Alzheimer education for high school and college students: A randomized controlled trial. Neurology. 2020 Oct 20;95(16):e2305-e2313. doi: 10.1212/WNL.0000000000009859. Epub 2020 Jul 14. PMID: 32665410; PMCID: PMC7713789.

276: Terracciano A, Artese A, Yeh J, Edgerton L, Granville L, Aschwanden D, Luchetti M, Glueckauf RL, Stephan Y, Sutin AR, Katz P. Effectiveness of Powerful Tools for Caregivers on Caregiver Burden and on Care Recipient Behavioral and Psychological Symptoms of Dementia: A Randomized Controlled Trial. J Am Med Dir Assoc. 2020 Aug;21(8):1121-1127.e1. doi: 10.1016/j.jamda.2019.11.011. Epub 2019 Dec 19. PMID: 31866419; PMCID: PMC7302999.

277: Bakker C, van der Aart J, Labots G, Liptrot J, Cross DM, Klaassen ES, Dickinson S, Tasker T, Groeneveld GJ. Safety and Pharmacokinetics of HTL0018318, a Novel M<sub>1</sub> Receptor Agonist, Given in Combination with Donepezil at Steady State: A Randomized Trial in Healthy Elderly Subjects. Drugs R D. 2021 Sep;21(3):295-304. doi: 10.1007/s40268-021-00352-5. Epub 2021 Jun 23. PMID: 34164794; PMCID: PMC8363684.

278: Abe S, Ezaki O, Suzuki M. Medium-Chain Triglycerides (8:0 and 10:0) Increase Mini-Mental State Examination (MMSE) Score in Frail Elderly Adults in a



Randomized Controlled Trial. J Nutr. 2020 Sep 1;150(9):2383-2390. doi: 10.1093/jn/nxaa186. PMID: 32652024.

279: Tohda C, Matsui M, Inada Y, Yang X, Kuboyama T, Kimbara Y, Watari H. Combined Treatment with Two Water Extracts of *Eleutherococcus senticosus* Leaf and Rhizome of *Drynaria fortunei* Enhances Cognitive Function: A Placebo-Controlled, Randomized, Double-Blind Study in Healthy Adults. Nutrients. 2020 Jan 23;12(2):303. doi: 10.3390/nu12020303. PMID: 31979283; PMCID: PMC7070741.

280: Fowler NR, Perkins AJ, Gao S, Sachs GA, Boustani MA. Risks and Benefits of Screening for Dementia in Primary Care: The Indiana University Cognitive Health Outcomes Investigation of the Comparative Effectiveness of Dementia Screening (IU CHOICE) Trial. J Am Geriatr Soc. 2020 Mar;68(3):535-543. doi: 10.1111/jgs.16247. Epub 2019 Dec 2. PMID: 31792940; PMCID: PMC7187902.

281: Kim DR, Song S, Kim GM, Chang JH, Tak YJ, Huh U, Cho JS, Liao Y, Han KS, Ko MH, Park JH. Effects of ICT-Based Multicomponent Program on Body Composition and Cognitive Function in Older Adults: A Randomized Controlled Clinical Study. Clin Interv Aging. 2021 Jun 21;16:1161-1171. doi: 10.2147/CIA.S306894. PMID: 34188459; PMCID: PMC8232840.

282: Mraz H, Tong CY, Liu KPY. Semantic-based memory-encoding strategy and cognitive stimulation in enhancing cognitive function and daily task performance for older adults with mild cognitive impairment: A pilot non-randomised cluster controlled trial. PLoS One. 2023 Mar 27;18(3):e0283449. doi: 10.1371/journal.pone.0283449. PMID: 36972279; PMCID: PMC10042350.

283: Koch G, Motta C, Bonni S, Pellicciari MC, Picazio S, Casula EP, Maiella M, Di Lorenzo F, Ponzio V, Ferrari C, Scaricamazza E, Caltagirone C, Martorana A. Effect of Rotigotine vs Placebo on Cognitive Functions Among Patients With Mild to Moderate Alzheimer Disease: A Randomized Clinical Trial. JAMA Netw Open. 2020 Jul 1;3(7):e2010372. doi: 10.1001/jamanetworkopen.2020.10372. Erratum in: JAMA Netw Open. 2020 Aug 3;3(8):e2019190. PMID: 32667654; PMCID: PMC7364345.

284: Surr CA, Holloway I, Walwyn RE, Griffiths AW, Meads D, Kelley R, Martin A, McLellan V, Ballard C, Fossey J, Burnley N, Chenoweth L, Creese B, Downs M, Garrod L, Graham EH, Lilley-Kelley A, McDermid J, Millard H, Perfect D, Robinson L, Robinson O, Shoosmith E, Siddiqi N, Stokes G, Wallace D, Farrin AJ. Dementia Care Mapping™ to reduce agitation in care home residents with dementia: the EPIC cluster RCT. Health Technol Assess. 2020 Mar;24(16):1-172. doi: 10.3310/hta24160. PMID: 32216870; PMCID: PMC7132533.

285: Ljubenkov PA, Edwards L, Iaccarino L, La Joie R, Rojas JC, Koestler M, Harris B, Boeve BF, Borroni B, van Swieten JC, Grossman M, Pasquier F, Frisoni GB, Mummery CJ, Vandenberghe R, Le Ber I, Hannequin D, McGinnis SM, Auriacombe S, Onofri M, Goodman IJ, Riordan HJ, Wisniewski G, Hesterman J, Marek K, Haynes BA, Patzke H, Koenig G, Hilt D, Moebius H, Boxer AL. Effect of the Histone Deacetylase Inhibitor FRM-0334 on Progranulin Levels in Patients With Progranulin Gene Haploinsufficiency: A Randomized Clinical Trial. JAMA Netw Open. 2021 Sep 1;4(9):e2125584. doi: 10.1001/jamanetworkopen.2021.25584. PMID: 34559230; PMCID: PMC8463943.

286: Altomare D, Barkhof F, Caprioglio C, Collij LE, Scheltens P, Lopes Alves I, Bouwman F, Berkhof J, van Maurik IS, Garibotto V, Moro C, Delrieu J, Payoux P, Saint-Aubert L, Hitzel A, Molinuevo JL, Grau-Rivera O, Gispert JD, Drzezga A,

Jessen F, Zeyen P, Nordberg A, Savitcheva I, Jelic V, Walker Z, Edison P, Demonet JF, Gismondi R, Farrar G, Stephens AW, Frisoni GB; Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) Consortium. Clinical Effect of Early vs Late Amyloid Positron Emission Tomography in Memory Clinic Patients: The AMYPAD-DPMS Randomized Clinical Trial. *JAMA Neurol.* 2023 Jun 1;80(6):548-557. doi: 10.1001/jamaneurol.2023.0997. PMID: 37155177; PMCID: PMC10167601.

287: Lin CH, Chen PK, Wang SH, Lane HY. Effect of Sodium Benzoate on Cognitive Function Among Patients With Behavioral and Psychological Symptoms of Dementia: Secondary Analysis of a Randomized Clinical Trial. *JAMA Netw Open.* 2021 Apr 1;4(4):e216156. doi: 10.1001/jamanetworkopen.2021.6156. PMID: 33881530; PMCID: PMC8060832.

288: Bonner GJ, Freels S, Ferrans C, Steffen A, Suarez ML, Dancy BL, Watkins YJ, Collinge W, Hart AS, Aggarwal NT, Wilkie DJ. Advance Care Planning for African American Caregivers of Relatives With Dementias: Cluster Randomized Controlled Trial. *Am J Hosp Palliat Care.* 2021 Jun;38(6):547-556. doi: 10.1177/1049909120916127. Epub 2020 Apr 20. PMID: 32308012; PMCID: PMC8443116.

289: Benedict C, Blennow K, Zetterberg H, Cedernaes J. Effects of acute sleep loss on diurnal plasma dynamics of CNS health biomarkers in young men. *Neurology.* 2020 Mar 17;94(11):e1181-e1189. doi: 10.1212/WNL.0000000000008866. Epub 2020 Jan 8. PMID: 31915189; PMCID: PMC7220231.

290: Conley AC, Albert KM, Boyd BD, Kim SG, Shokouhi S, McDonald BC, Saykin AJ, Dumas JA, Newhouse PA. Cognitive complaints are associated with smaller right medial temporal gray-matter volume in younger postmenopausal women. *Menopause.* 2020 Nov;27(11):1220-1227. doi: 10.1097/GME.0000000000001613. PMID: 33110037; PMCID: PMC9153070.

291: Vijverberg EGB, Axelsen TM, Bihlet AR, Henriksen K, Weber F, Fuchs K, Harrison JE, Kühn-Wache K, Alexandersen P, Prins ND, Scheltens P. Rationale and study design of a randomized, placebo-controlled, double-blind phase 2b trial to evaluate efficacy, safety, and tolerability of an oral glutaminy cyclase inhibitor varoglutamstat (PQ912) in study participants with MCI and mild AD-VIVIAD. *Alzheimers Res Ther.* 2021 Aug 23;13(1):142. doi: 10.1186/s13195-021-00882-9. PMID: 34425883; PMCID: PMC8381483.

292: Cheung T, Li TMH, Ho YS, Kranz G, Fong KNK, Leung SF, Lam SC, Yeung WF, Lam JYT, Fong KH, Beisteiner R, Xiang YT, Cheng CPW. Effects of Transcranial Pulse Stimulation (TPS) on Adults with Symptoms of Depression-A Pilot Randomized Controlled Trial. *Int J Environ Res Public Health.* 2023 Jan 28;20(3):2333. doi: 10.3390/ijerph20032333. PMID: 36767702; PMCID: PMC9915638.

293: Sperling R, Henley D, Aisen PS, Raman R, Donohue MC, Ernststrom K, Rafii MS, Streffer J, Shi Y, Karcher K, Raghavan N, Tymofyeyev Y, Bogert J, Brashear HR, Novak G, Thippawong J, Saad ZS, Kolb H, Rofael H, Sanga P, Romano G. Findings of Efficacy, Safety, and Biomarker Outcomes of Atabecestat in Preclinical Alzheimer Disease: A Truncated Randomized Phase 2b/3 Clinical Trial. *JAMA Neurol.* 2021 Mar 1;78(3):293-301. doi: 10.1001/jamaneurol.2020.4857. PMID: 33464300; PMCID: PMC7816119.

294: Morató X, Marquí M, Tartari JP, Lafuente A, Abdelnour C, Alegret M, Jofresa S, Buendía M, Pancho A, Aguilera N, Ibarria M, Diego S, Cuevas R, Cañada L, Calvet A, Antonio EE, Pérez-Cordón A, Sanabria Á, de Rojas I, Nuñez-Llaves R, Cano A, Orellana A, Montreal L, Cañabate P, Rosende-Roca M, Vargas L, Bojaryn

U, Ricciardi M, Ariton DM, Espinosa A, Ortega G, Muñoz N, Leonart N, Alarcón-Martín E, Moreno M, Preckler S, Tantinya N, Ramis M, Nogales AB, Seguer S, Martín E, Pytel V, Valero S, Gurruchaga M, Tárraga L, Ruiz A, Boada M. A randomized, open-label clinical trial in mild cognitive impairment with EGb 761 examining blood markers of inflammation and oxidative stress. *Sci Rep.* 2023 Apr 3;13(1):5406. doi: 10.1038/s41598-023-32515-6. PMID: 37012306; PMCID: PMC10070452.

295: Grill JD, Raman R, Ernstrom K, Sultzer DL, Burns JM, Donohue MC, Johnson KA, Aisen PS, Sperling RA, Karlawish J; A4 Study Team. Short-term Psychological Outcomes of Disclosing Amyloid Imaging Results to Research Participants Who Do Not Have Cognitive Impairment. *JAMA Neurol.* 2020 Dec 1;77(12):1504-1513. doi: 10.1001/jamaneurol.2020.2734. PMID: 32777010; PMCID: PMC7418046.

296: Jeong JH, Lee JY, Kim JY, Seo YK, Kang WC, Kang HW, Park SJ, Jang HK, Park YC, Jung IC. Safety and efficacy of Jujadokseo-hwan for memory deficit (amnesia) in mild neurocognitive disorder: A protocol for randomized, double-blind, placebo-controlled, parallel-group, multicenter clinical trial. *Medicine (Baltimore).* 2020 Feb;99(8):e19231. doi: 10.1097/MD.00000000000019231. PMID: 32080123; PMCID: PMC7034695.

297: Mitchell SL, D'Agata EMC, Hanson LC, Loizeau AJ, Habtemariam DA, Tsai T, Anderson RA, Shaffer ML. The Trial to Reduce Antimicrobial Use in Nursing Home Residents With Alzheimer Disease and Other Dementias (TRAIN-AD): A Cluster Randomized Clinical Trial. *JAMA Intern Med.* 2021 Sep 1;181(9):1174-1182. doi: 10.1001/jamainternmed.2021.3098. PMID: 34251396; PMCID: PMC8276127.

298: Shaw CA, Williams KN, Perkhounkova Y, Hein M, Coleman CK. Effects of a Video-based Intervention on Caregiver Confidence for Managing Dementia Care Challenges: Findings from the FamTechCare Clinical Trial. *Clin Gerontol.* 2020 Oct-Dec;43(5):508-517. doi: 10.1080/07317115.2020.1729917. Epub 2020 Feb 19. PMID: 32072866; PMCID: PMC7434650.

299: Jeong JH, Jung C, Kim J, Kim JY, Kim HS, Park YC, Lee JH, Jung IC. Investigation of combined treatment of acupuncture and neurofeedback for improving cognitive function in mild neurocognitive disorder: A randomized, assessor-blind, pilot study. *Medicine (Baltimore).* 2021 Sep 17;100(37):e27218. doi: 10.1097/MD.00000000000027218. PMID: 34664858; PMCID: PMC8448021.

300: Anstey KJ, Cherbuin N, Kim S, McMaster M, D'Este C, Lautenschlager N, Rebok G, McRae I, Torres SJ, Cox KL, Pond CD. An Internet-Based Intervention Augmented With a Diet and Physical Activity Consultation to Decrease the Risk of Dementia in At-Risk Adults in a Primary Care Setting: Pragmatic Randomized Controlled Trial. *J Med Internet Res.* 2020 Sep 24;22(9):e19431. doi: 10.2196/19431. PMID: 32969833; PMCID: PMC7545332.

301: Sheline YI, Snider BJ, Beer JC, Seok D, Fagan AM, Suckow RF, Lee JM, Waligorska T, Korecka M, Aselcioglu I, Morris JC, Shaw LM, Cirrito JR. Effect of escitalopram dose and treatment duration on CSF A $\beta$  levels in healthy older adults: A controlled clinical trial. *Neurology.* 2020 Nov 10;95(19):e2658-e2665. doi: 10.1212/WNL.00000000000010725. Epub 2020 Sep 10. PMID: 32913021; PMCID: PMC7713735.

302: Robert P, Manera V, Derreumaux A, Ferrandez Y Montesino M, Leone E, Fabre R, Bourgeois J. Efficacy of a Web App for Cognitive Training (MeMo) Regarding Cognitive and Behavioral Performance in People With Neurocognitive Disorders:

Randomized Controlled Trial. J Med Internet Res. 2020 Mar 11;22(3):e17167. doi: 10.2196/17167. PMID: 32159519; PMCID: PMC7097721.

303: El Alili M, Smaling HJA, Joling KJ, Achterberg WP, Francke AL, Bosmans JE, van der Steen JT. Cost-effectiveness of the Namaste care family program for nursing home residents with advanced dementia in comparison with usual care: a cluster-randomized controlled trial. BMC Health Serv Res. 2020 Sep 4;20(1):831. doi: 10.1186/s12913-020-05570-2. PMID: 32887591; PMCID: PMC7473814.

304: Levakov G, Kaplan A, Yaskolka Meir A, Rinott E, Tsaban G, Zelicha H, Blüher M, Ceglarek U, Stumvoll M, Shelef I, Avidan G, Shai I. The effect of weight loss following 18 months of lifestyle intervention on brain age assessed with resting-state functional connectivity. Elife. 2023 Apr 6;12:e83604. doi: 10.7554/eLife.83604. PMID: 37022140; PMCID: PMC10174688.

305: Berry DA, Dhadda S, Kanekiyo M, Li D, Swanson CJ, Irizarry M, Kramer LD, Berry SM. Lecanemab for Patients With Early Alzheimer Disease: Bayesian Analysis of a Phase 2b Dose-Finding Randomized Clinical Trial. JAMA Netw Open. 2023 Apr 3;6(4):e237230. doi: 10.1001/jamanetworkopen.2023.7230. PMID: 37040116; PMCID: PMC10091161.

306: Howard R, Zubko O, Bradley R, Harper E, Pank L, O'Brien J, Fox C, Tabet N, Livingston G, Bentham P, McShane R, Burns A, Ritchie C, Reeves S, Lovestone S, Ballard C, Noble W, Nilforooshan R, Wilcock G, Gray R; Minocycline in Alzheimer Disease Efficacy (MADE) Trialist Group. Minocycline at 2 Different Dosages vs Placebo for Patients With Mild Alzheimer Disease: A Randomized Clinical Trial. JAMA Neurol. 2020 Feb 1;77(2):164-174. doi: 10.1001/jamaneurol.2019.3762. PMID: 31738372; PMCID: PMC6865324.

307: Tsai RM, Miller Z, Koestler M, Rojas JC, Ljubenkov PA, Rosen HJ, Rabinovici GD, Fagan AM, Cobigo Y, Brown JA, Jung JI, Hare E, Geldmacher DS, Natelson-Love M, McKinley EC, Luong PN, Chuu EL, Powers R, Mumford P, Wolf A, Wang P, Shamloo M, Miller BL, Roberson ED, Boxer AL. Reactions to Multiple Ascending Doses of the Microtubule Stabilizer TPI-287 in Patients With Alzheimer Disease, Progressive Supranuclear Palsy, and Corticobasal Syndrome: A Randomized Clinical Trial. JAMA Neurol. 2020 Feb 1;77(2):215-224. doi: 10.1001/jamaneurol.2019.3812. PMID: 31710340; PMCID: PMC6865783.