



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΑΣ, ΔΙΟΙΚΗΣΗΣ ΚΑΙ
ΠΛΗΡΟΦΟΡΙΚΗΣ

Τμήμα Πληροφορικής και
Τηλεπικοινωνιών

Η μηχανή αναζήτησης TReSA

Συγγραφείς:

Λάμπρος Γραμματικόπουλος, 2022201800038, dit18038@uop.gr

Κωσταντίνος Νικόλαος Κολοτούρος, 2022201800090, dit18090@uop.gr

Διδάσκων:

Χρήστος Τρυφωνόπουλος




Δεκέμβριος 2020

Πίνακας περιεχομένων

Οδηγίες εκτέλεσης του κώδικα μέσα από τον IDE (configurations με IntelliJ IDEA)	2
Οδηγίες εκτέλεσης του εκτελέσιμου αρχείου (TReSA.jar)	4
Λεπτομέρειες υλοποίησης	5
Παραδείγματα επίδειξης του προγράμματος (με τα 100 πρώτα άρθρα (0-99))	6
Επεξήγηση των βασικών στοιχείων του προγράμματός (user manual).....	13
Πηγές:.....	22

Οδηγίες εκτέλεσης του κώδικα μέσα από τον IDE (configurations με IntelliJ IDEA)

1. Εγκαθιστούμε το πιο πρόσφατο JDK της Java από το ακόλουθο link:
<https://www.oracle.com/java/technologies/downloads/>
2. Εγκαθιστούμε το πιο πρόσφατο JDK της JavaFX από το ακόλουθο link: <https://gluonhq.com/products/javafx/>
3. Εγκαθιστούμε το IDE της επιλογής μας, εμείς χρησιμοποιήσαμε το IntelliJ IDEA, από το ακόλουθο Link:
<https://www.jetbrains.com/idea/download/#section=windows>
4. Ανοίγουμε το φάκελο του project με το IDE και κάνουμε τα απαραίτητα configurations. Στην περίπτωση μας τα βήματα για να μπορέσει να εκτελεστεί το project είναι τα εξής με το IntelliJ IDEA:
 - i. Πατάμε πάνω αριστερά το κουμπί «file» επιλέγουμε «Project Structure» -> «Project Settings» -> «Project», στο πεδίο «SDK» επιλέγουμε το SDK της java που έχουμε εγκαταστήσει και στο πεδίο «Language» επιλέγουμε «15 – Text Blocks». Πατάμε το κουμπί «Apply» και μετά το κουμπί «OK».

- ii. Πατάμε στο κουμπί πάνω δεξιά «Add Configuration». Πατάμε στο κουμπί σε σχήμα σταυρού πάνω αριστερά  και επιλέγουμε «Application». Στο πεδίο «Main class» συμπληρώνουμε: `sample.Main`. Πατάμε στο κουμπί «Modify options», επιλέγουμε «add VM options» στο πεδίο «VM options» συμπληρώνουμε: `--module-path "εδώ τοποθετούμε το path για το lib του JavaFX SDK (.\\javafx-sdk-xx.x.x\\lib)" --add-modules javafx.controls,javafx.fxml`. Στο πρώτο πάνω αριστερά drop down menu, επιλέγουμε το path για το JDK που εγκαταστήσαμε. Πατάμε το κουμπί «Apply» και μετά το κουμπί «OK».
 - iii. Τα libraries του project έχουν γίνει linked από εμάς με τα περιεχόμενα του φακέλου lib του project. Σε περίπτωση όμως που κάτι δεν έχει πάει καλά, μπορείτε να τα συνδέσετε εσείς με τον ακόλουθο τρόπο: Πατάμε πάνω αριστερά το κουμπί «file» επιλέγουμε «Project Structure» -> «Libraries», πατάμε στο κουμπί σε σχήμα σταυρού πάνω αριστερά  και τοποθετούμε όλο το φάκελο lib της JavaFX και τις εξής βιβλιοθήκες του Lucene (lucene-analyzers-common-8.11.0, lucene-core-8.11.0, lucene-queryparser-8.11.0). Ο φάκελος αυτός και οι βιβλιοθήκες υπάρχουν στον φάκελο lib του project και τις επιλέγουμε από εκεί για συντομία. Πατάμε το κουμπί «Apply» και μετά το κουμπί «OK».
5. Τέλος για να τρέξουμε τον κώδικα πατάμε τον συνδυασμό πλήκτρων Ctrl + F5 ή πατάμε το πάνω δεξιά κουμπί σε σχήμα πράσινου τριγώνου .

Οδηγίες εκτέλεσης του εκτελέσιμου αρχείου (TReSA.jar)

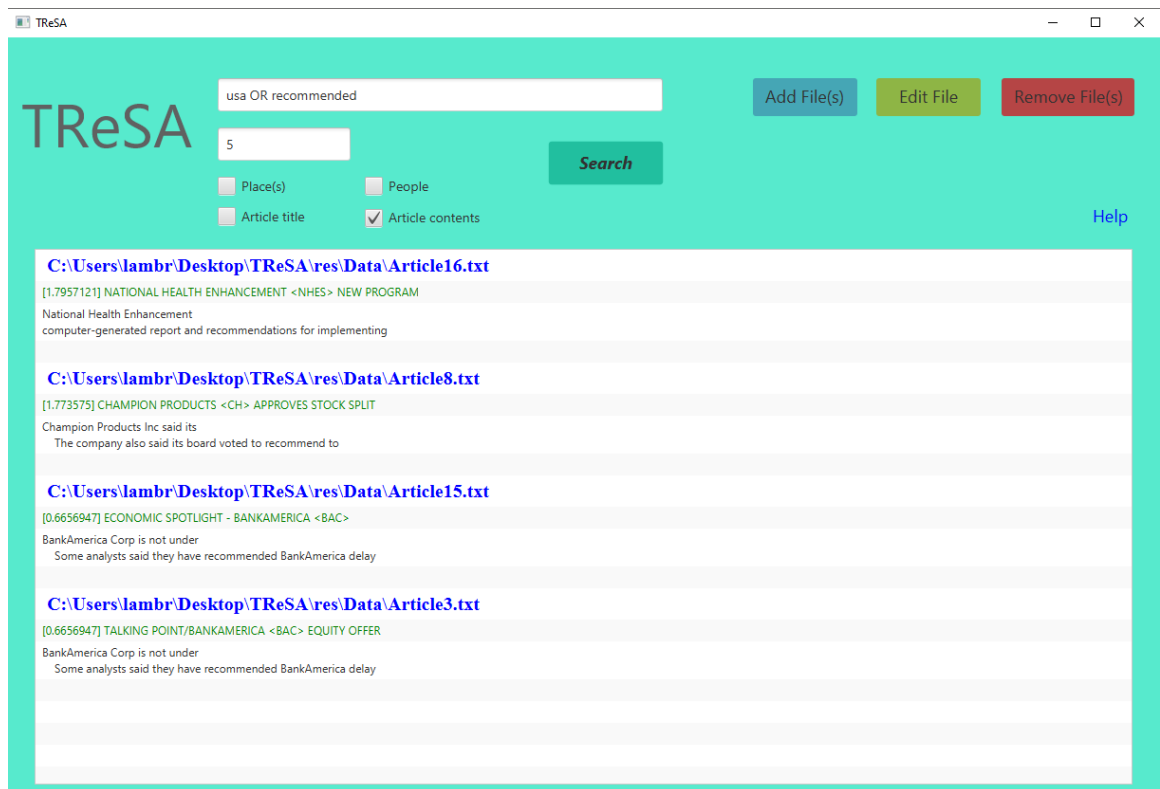
1. Τοποθετούμε (αν δεν υπάρχει ήδη) τον φάκελο res στον ίδιο φάκελο με το αρχείο TReSA.jar.
2. Ανοίγουμε ένα τερματικό (terminal)
3. Πλοηγούμαστε στον φάκελο που βρίσκεται το TReSA.jar με την εντολή: `cd εδώ τοποθετούμε το path για το φάκελο στον οποίο βρίσκεται το TReSA.jar αρχείο`.
4. Εκτελούμε την ακόλουθη εντολή: `java --module-path "εδώ τοποθετούμε το path για το lib του JavaFX SDK \(.\\javafx-sdk-xx.x.x\\lib\)" --add-modules javafx.controls,javafx.fxml -jar TReSA.jar`
Main

Λεπτομέρειες υλοποίησης

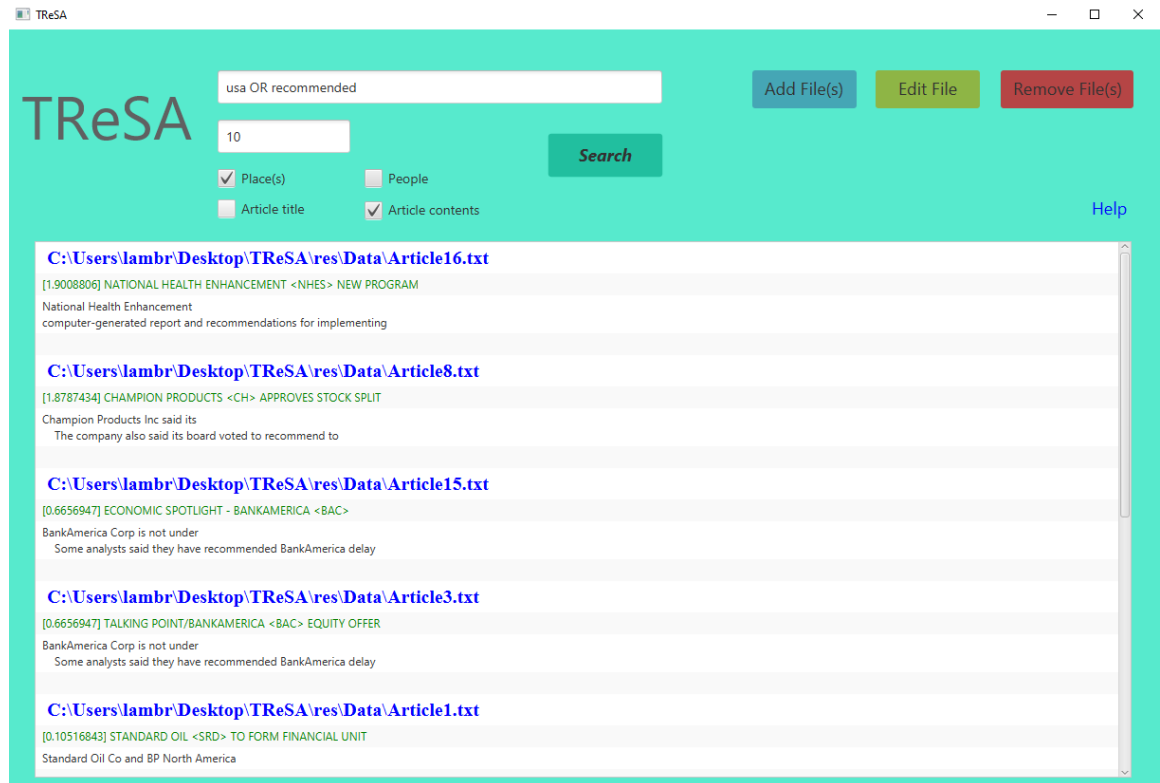
- Η μηχανή αναζήτησης υποστηρίζει όλους τους ζητούμενους τύπους ερωτήσεων «queries».
- Τα άρθρα εμφανίζονται από την μεγαλύτερη (πάνω) προς την μικρότερη (κάτω) σειρά σχετικότητας με βάση το score τους.
- Ο χρήστης μπορεί να ανοίξει για να διαβάσει ένα άρθρο πατώντας πάνω στην διαδρομή (path) του άρθρου. Στη συνέχεια στο εμφανιζόμενο παράθυρο μπορεί να διαβάσει το άρθρο, καθώς και να ψάξει τα πιο σχετικά άρθρα με βάση τον Τίτλο, την Περιοχή και τους Ανθρώπους του άρθρου που επέλεξε να διαβάσει.
- Ο χρήστης μπορεί να θέσει ένα όριο στο πλήθος των άρθρων που του επιστρέφει η μηχανή αναζήτησης (Top-k) κατά την αναζήτηση άρθρων αλλά και κατά την αναζήτηση των σχετικών άρθρων ενός άρθρου.
- Για να προσθέσει, επεξεργαστεί και διαγράψει ένα ή περισσότερα αρχεία ο χρήστης, τα επιλέγει από τα τοπικά αρχεία του φακέλου του project στον υπολογιστή όπου είναι αποθηκευμένα.
- Ο χρήστης μπορεί να ανοίξει πολλαπλά άρθρα ταυτόχρονα από τα αποτελέσματα της ερώτησης που έκανε στη μηχανή αναζήτησης.
- Το ευρετήριο θα δημιουργηθεί μόνο την πρώτη φορά που ο χρήστης θα εκτελέσει το πρόγραμμα και θα ενημερώνεται όταν ο χρήστης προσθέτει, επεξεργάζεται και διαγράφει αρχεία.

Παραδείγματα επίδειξης του προγράμματος (με τα 100 πρώτα άρθρα (0-99))

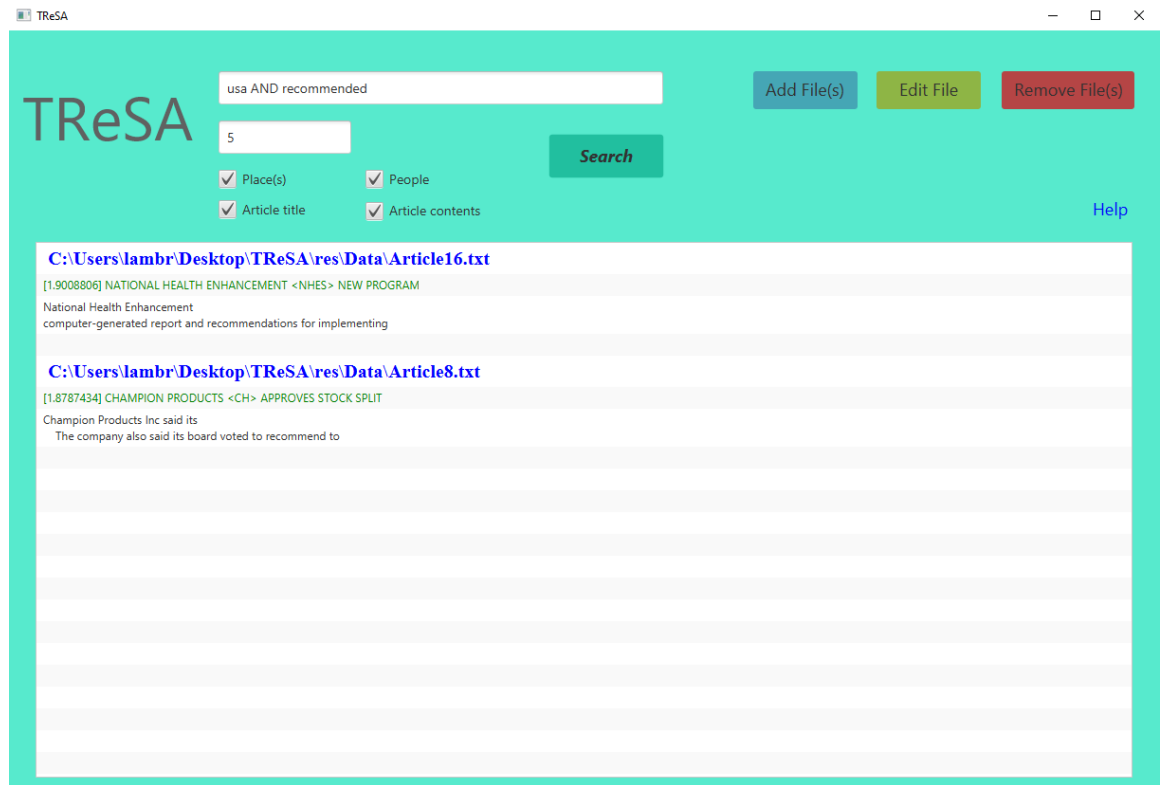
- i. Στην παρακάτω εικόνα αναζητούμε τα 10 πρώτα άρθρα που περιέχουν την λέξη «usa» ή την λέξη «recommended» στο πεδίο «BODY» των άρθρων και έχουμε 4 αποτελέσματα:



- ii. Στην παρακάτω εικόνα αναζητούμε τα 10 πρώτα άρθρα που περιέχουν την λέξη «usa» ή την λέξη «recommended» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES, «BODY») και έχουμε 10 αποτελέσματα:



- iii. Στην παρακάτω εικόνα αναζητούμε τα 5 πρώτα άρθρα που περιέχουν την λέξη «usa» και ταυτόχρονα την λέξη «recommended» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES», «PEOPLE», «TITLE», «BODY») και έχουμε 2 αποτελέσματα:



- iv. Στην παρακάτω εικόνα αναζητούμε τα 10 πρώτα άρθρα που περιέχουν την λέξη «usa» αλλά όχι τις λέξεις «recommended» και «connection» και ταυτόχρονα περιέχουν την λέξη «available» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES», «PEOPLE», «TITLE», «BODY») και έχουμε 3 αποτελέσματα:

The screenshot shows the TReSA search interface. The search bar contains the query: `((usa NOT recommended) NOT connection) AND available`. Below the search bar, there is a text input field with the number '5'. To the right of the search bar are three buttons: 'Add File(s)', 'Edit File', and 'Remove File(s)'. Below the search bar, there are four checkboxes: 'Place(s)', 'People', 'Article title', and 'Article contents', all of which are checked. A 'Search' button is located to the right of these checkboxes. A 'Help' link is located in the top right corner. The search results are displayed in a list format, showing the file path, the article title, and the article content. The results are as follows:

File Path	Article Title	Article Content
C:\Users\lambr\Desktop\TReSA\res\Data\Article16.txt	[1.7850523] NATIONAL HEALTH ENHANCEMENT <NHES> NEW PROGRAM	National Health Enhancement be available in 60 days. computer-generated report and recommendations for implementing
C:\Users\lambr\Desktop\TReSA\res\Data\Article10.txt	[1.7051699] COBANCO INC <CBCO> YEAR NET	Shr 34 cts vs 1.19 dlrs Note: 4th qtr not available. Year includes 1985
C:\Users\lambr\Desktop\TReSA\res\Data\Article87.txt	[1.7051699] SORG <SRG> STOCKHOLDERS FORM GROUP	Sorg Inc said a group composed of The company was not immediately available to provide
C:\Users\lambr\Desktop\TReSA\res\Data\Article4.txt	[1.18953] NATIONAL AVERAGE PRICES FOR FARMER-OWNED RESERVE	The U.S. Agriculture Department n.a.-not available.

- v. Στην παρακάτω εικόνα αναζητούμε τα 4 πρώτα άρθρα που περιέχουν την λέξη «billion», τον αριθμό «68,004,000» και την ημερομηνία «7/1/82» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES», «PEOPLE», «TITILE», «BODY») και έχουμε 4 αποτελέσματα:

The screenshot shows the TReSA search interface. At the top, there is a search bar containing the query "billion 68,004,000 7/1/82". Below the search bar, there are four checkboxes: "Place(s)", "People", "Article title", and "Article contents", all of which are checked. To the right of the search bar, there are three buttons: "Add File(s)", "Edit File", and "Remove File(s)". Below the search bar, there is a "Search" button. The search results are displayed in a list of four articles, each with a file path and a brief description of the article's content.

TReSA

billion 68,004,000 7/1/82

4

Search

☒ Place(s) ☒ People ☒ Article title ☒ Article contents

Help

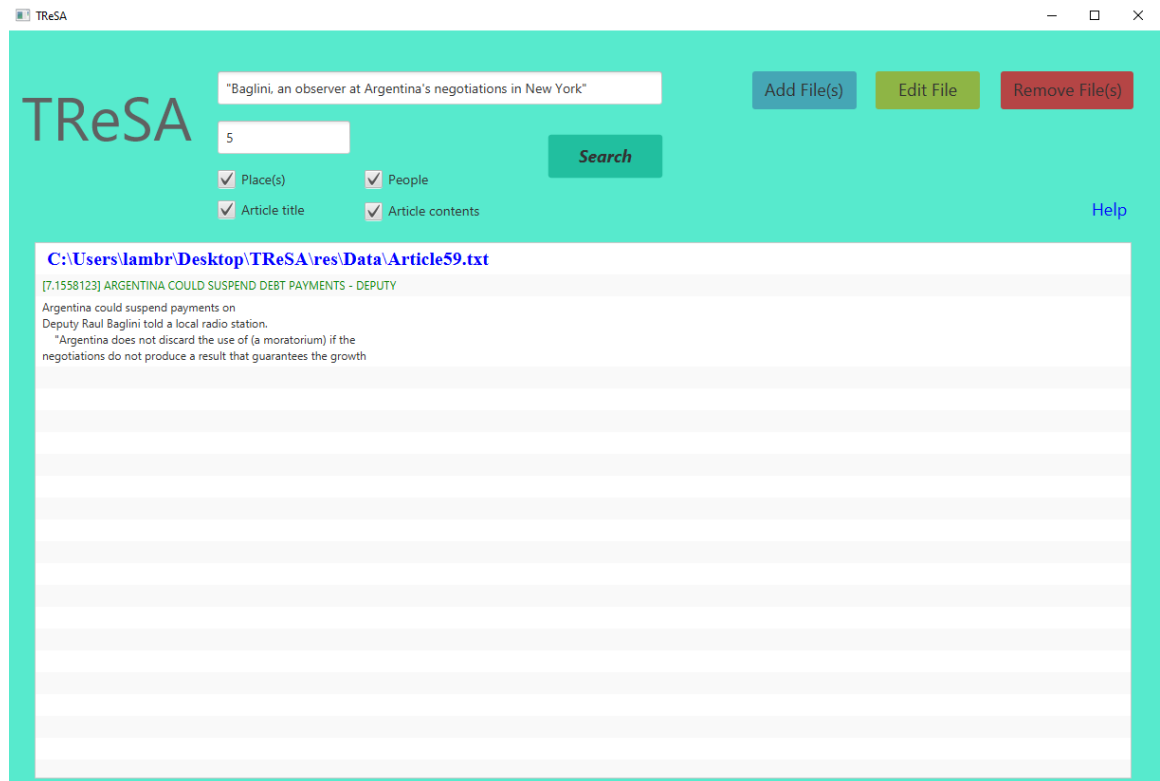
C:\Users\lambr\Desktop\TReSA\res\Data\Article64.txt
[3.3676414] FARMERS GROUP INC <FGRP> 4TH QTR NET
Shr 80 cts vs 72 cts
Revs 1.12 billion vs 992.9 mln
Avg shrs 69,127,000 vs 68,004,000

C:\Users\lambr\Desktop\TReSA\res\Data\Article4.txt
[1.5724702] NATIONAL AVERAGE PRICES FOR FARMER-OWNED RESERVE
The U.S. Agriculture Department
corn/sorghum after 7/1/82. Level VI covers wheat entered after

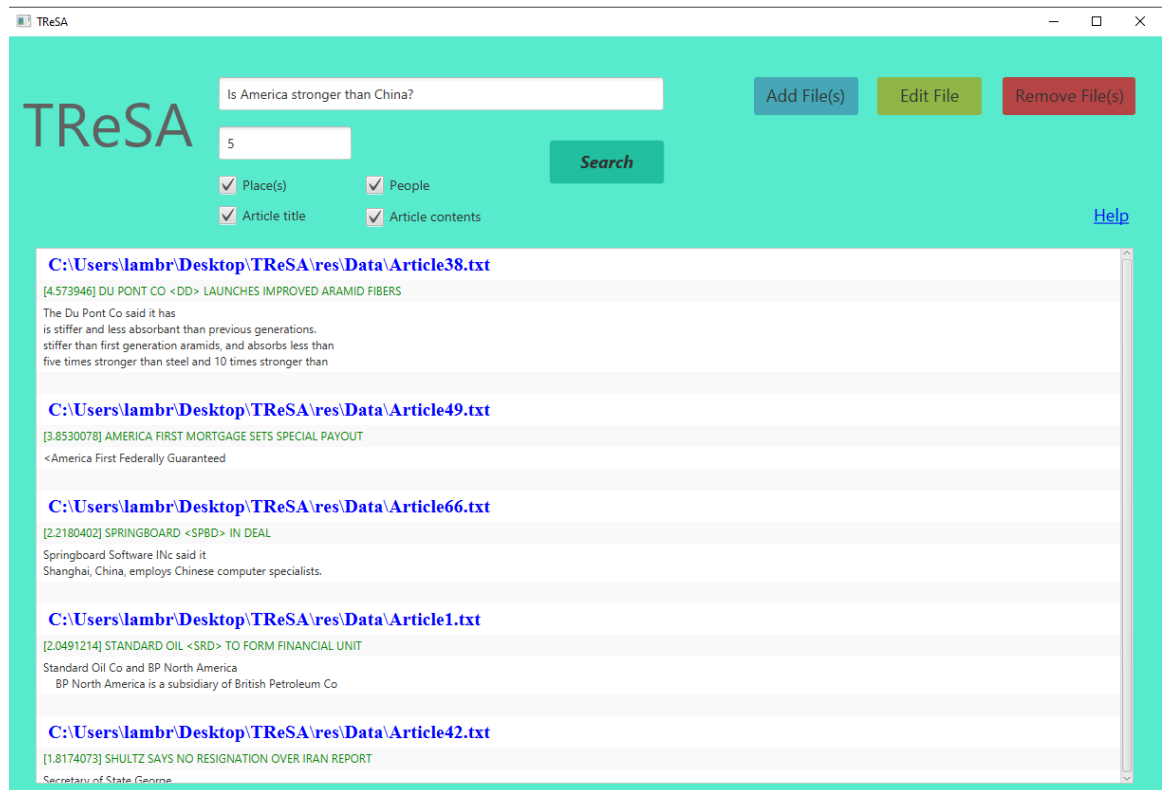
C:\Users\lambr\Desktop\TReSA\res\Data\Article61.txt
[1.378097] TREASURY BALANCES AT FED ROSE ON FEB 25
Treasury balances at the Federal
Reserve rose on Feb 25 to 4.151 billion dlrs from 2.727 billion
billion dlrs from 25.780 billion dlrs on the same respective
billion dlrs on Feb 25 compared with 28.507 billion dlrs on Feb

C:\Users\lambr\Desktop\TReSA\res\Data\Article47.txt
[1.3128628] N.Z. OFFICIAL FOREIGN RESERVES FALL IN JANUARY
New Zealand's official foreign
reserves fell to 7.15 billion N.Z. Dlrs in January from 7.20
billion dlrs in December and compared with 3.03 billion a year

- vi. Στην παρακάτω εικόνα αναζητούμε τα 5 πρώτα άρθρα που περιέχουν την ακριβώς την φράση «Baglini, an observer at Argentina's negotiations in New York» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES», «PEOPLE», «TITLE», «BODY») και έχουμε 1 αποτέλεσμα:

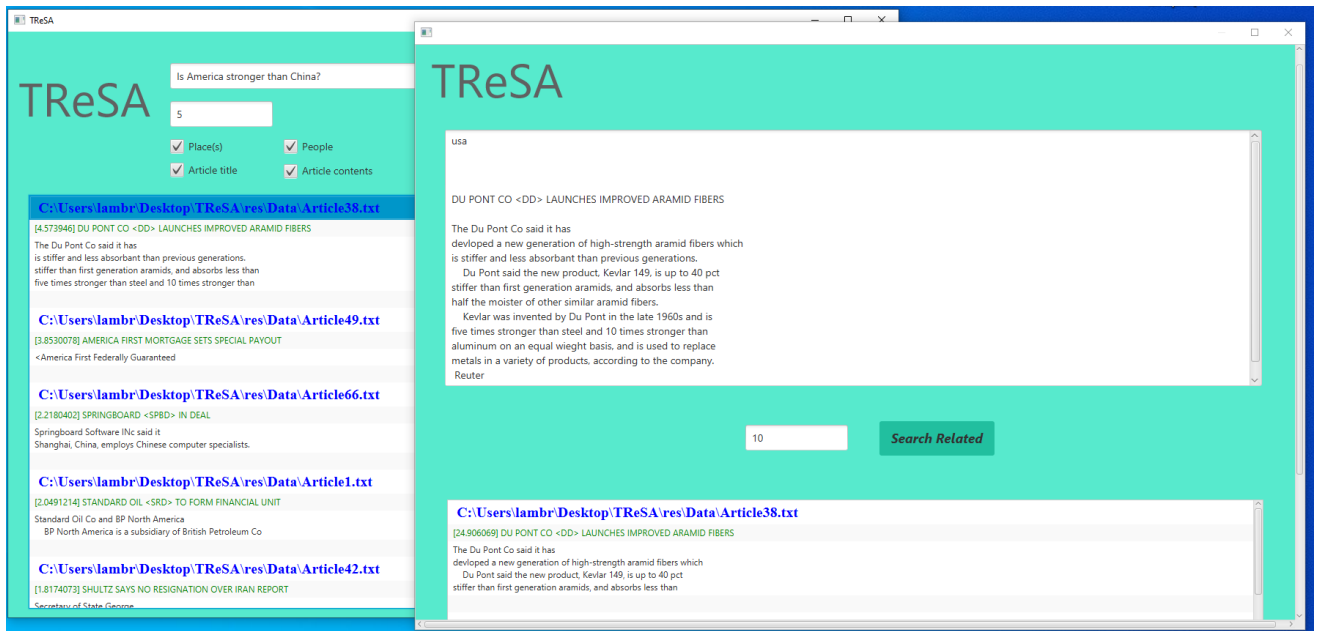


- vii. Στην παρακάτω εικόνα αναζητούμε τα 5 πρώτα άρθρα που περιέχουν τις λέξεις που περιέχονται σε μια τυχαία αναζήτηση «Is America stronger than China?» σε οποιοδήποτε από τα πεδία των άρθρων («PLACES», «PEOPLE», «TITLE», «BODY») και έχουμε 5 αποτελέσματα:

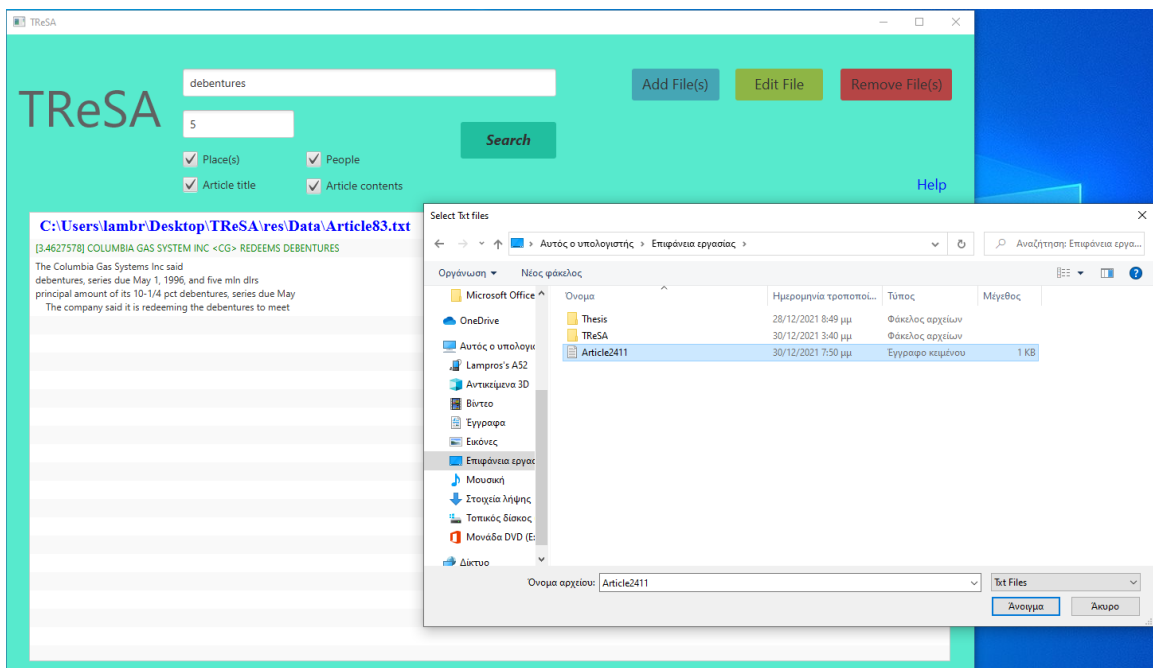
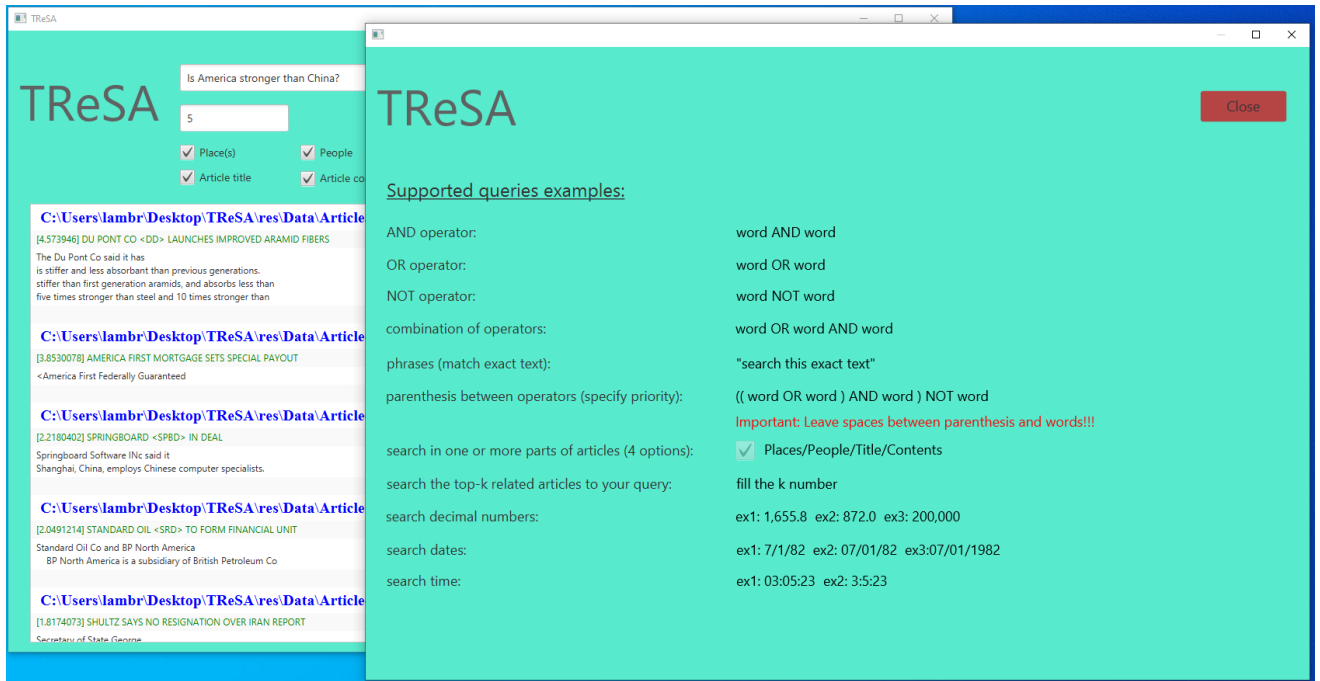


Επεξήγηση των βασικών στοιχείων του προγράμματός (user manual)

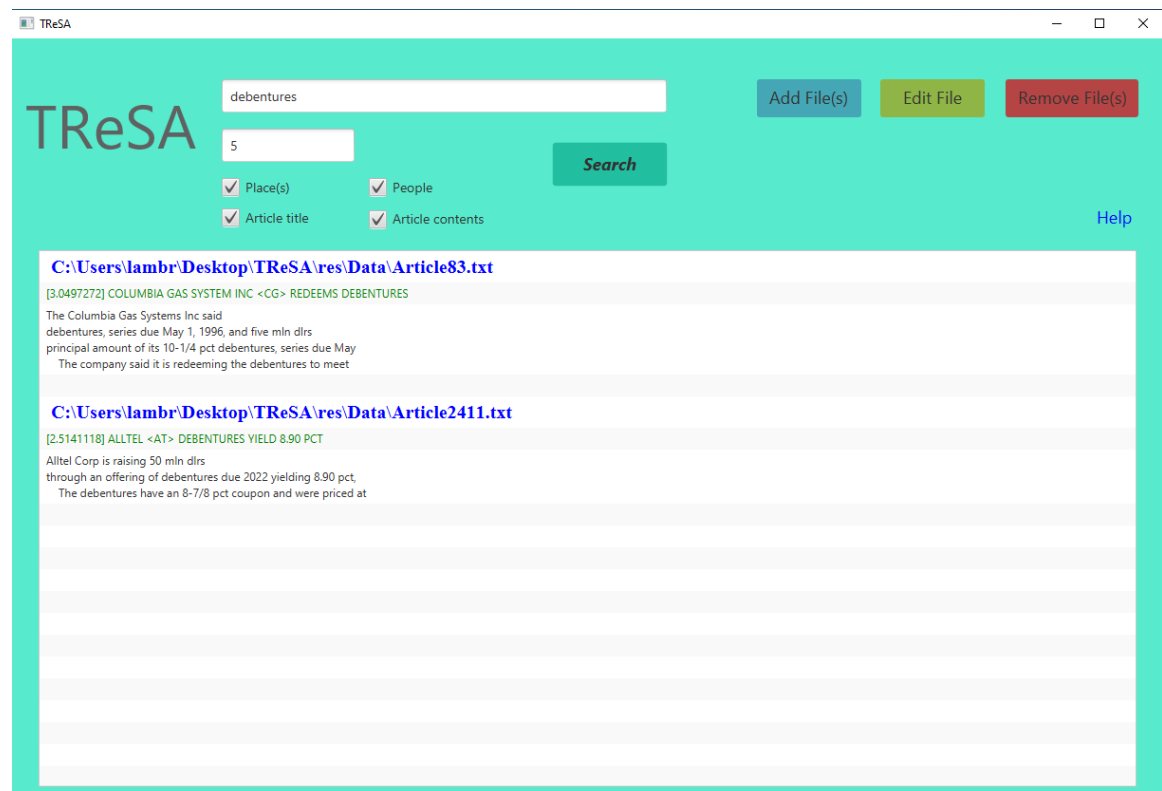
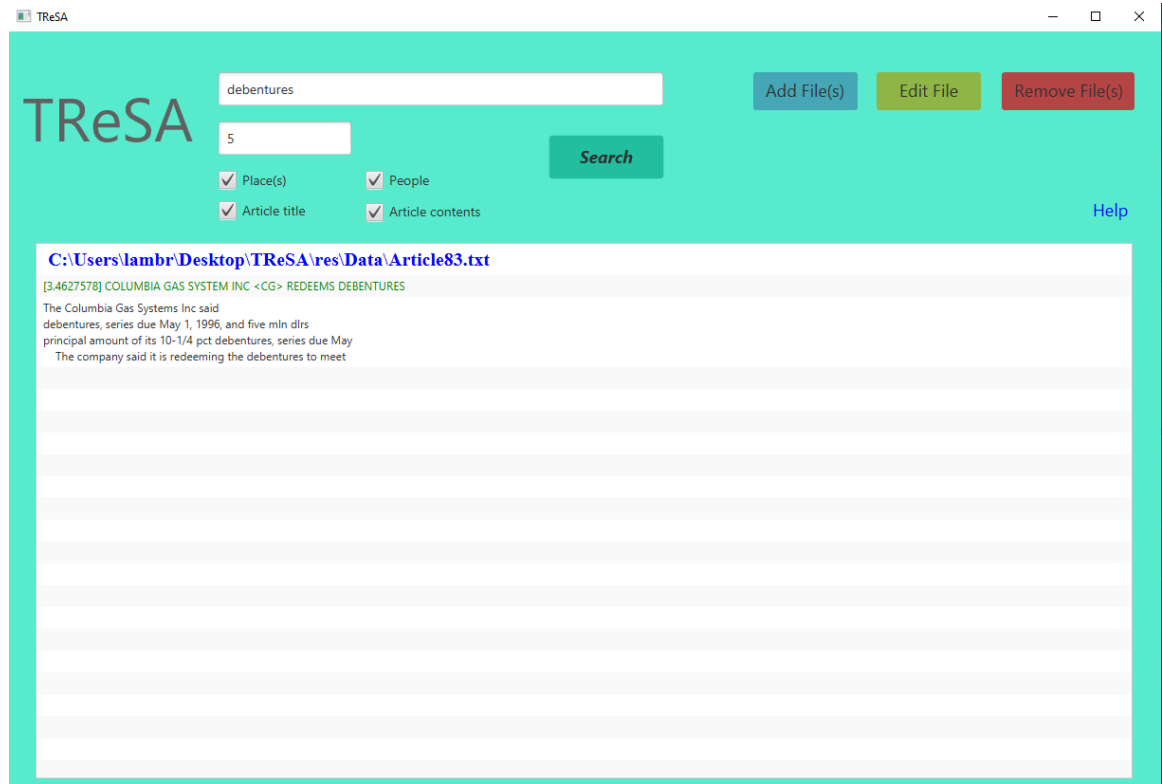
- ι. Στην παρακάτω εικόνα φαίνεται πως κανείς μπορεί να δει ένα κείμενο της επιλογής του και άλλα σχετικά κείμενα με αυτό:



- ii. Στην παρακάτω εικόνα φαίνεται πως κανείς μπορεί να δει τα παραδείγματα των υποστηριζόμενων ερωτημάτων από τον σύνδεσμο «Help»:



- iii. Στις παρακάτω εικόνες φαίνεται πως κανείς μπορεί να προσθέσει ένα ή περισσότερα αρχεία στην μηχανή αναζήτησης από το κουμπί «Add File(s)»:



TReSA

debtentures

5

Search

☒ Place(s)

☒ Article title

☒ People

☒ Article contents

Add File(s)

Edit File

Remove File(s)

[Help](#)

C:\Users\lambr\Desktop\TReSA\res\Data\Article83.txt

[3.0497272] COLUMBIA GAS SYSTEM INC <CG> REDEEMS DEBENTURES

The Columbia Gas Systems Inc said debtentures, series due May 1, 1996, and five mln dlrs principal amount of its 10-1/4 pct debtentures, series due May

The company said it is redeeming the debtentures to meet

C:\Users\lambr\Desktop\TReSA\res\Data\Article2411.txt

[2.5141118] ALLTEL <AT> DEBENTURES YIELD 8.90 PCT

Alltel Corp is raising 50 mln dlrs through an offering of debtentures due 2022 yielding 8.90 pct,

The debtentures have an 8-7/8 pct coupon and were priced at

Άνοιγμα

← → ↑

Αυτός ο υπολογιστής > Επιφάνεια εργασίας > TReSA > res > Data

Αναζήτηση Data

Οργάνωση

Νέος φάκελος

Microsoft Office

OneDrive

Αυτός ο υπολογιστής

Lamprios's AS2

Αντικείμενα 3D

Βίντεο

Εγγραφο

Εικόνες

Επιφάνεια εργασίας

Μουσική

Στοιχεία λήψης

Τοπικός δίσκος

Μονάδα DVD (E

Όνομα

Ημερομηνία τροποποι...

Τύπος

Μέγεθος

Article86

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article87

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article88

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article89

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article90

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article91

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article92

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article93

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article94

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article95

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article96

30/12/2021 12:20 μμ

Εγγραφο κειμένου

3 KB

Article97

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article98

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article99

30/12/2021 12:20 μμ

Εγγραφο κειμένου

1 KB

Article2411

30/12/2021 7:50 μμ

Εγγραφο κειμένου

1 KB

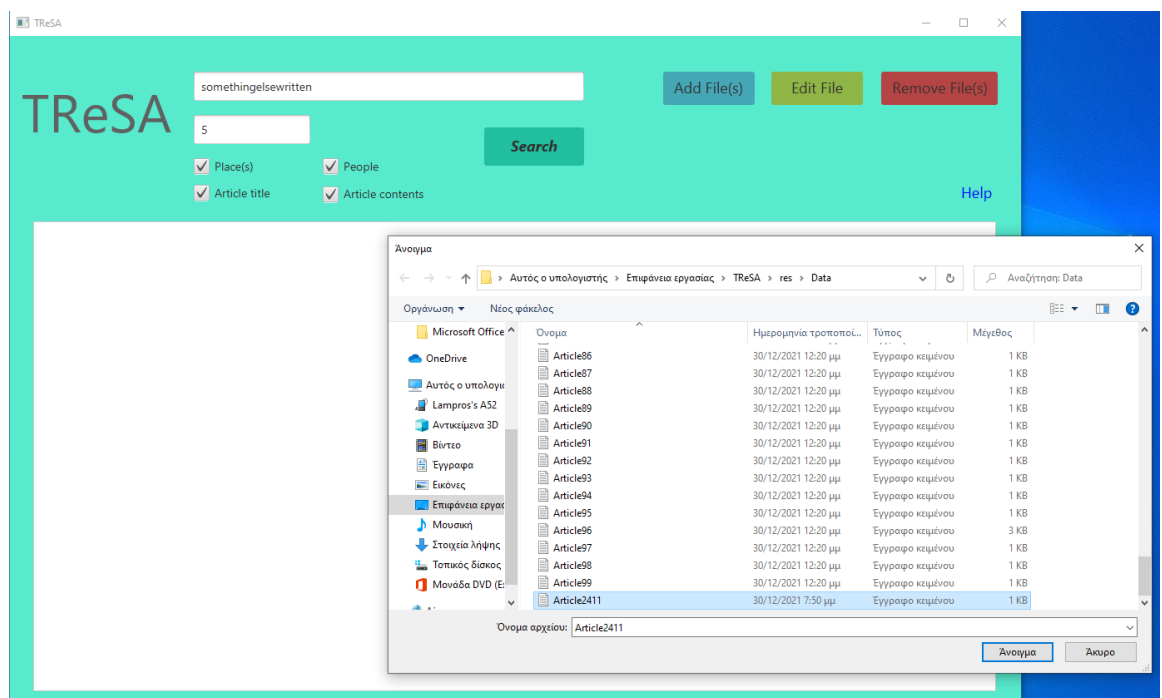
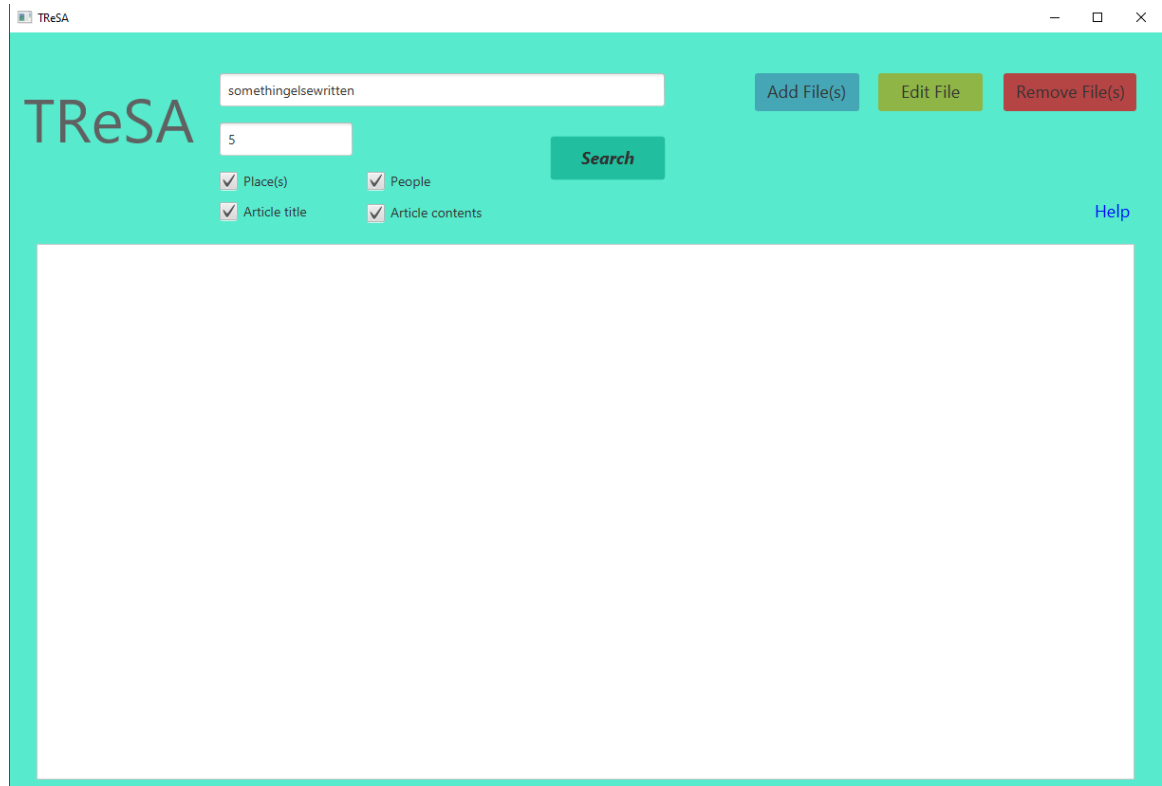
Όνομα αρχείου

Article2411

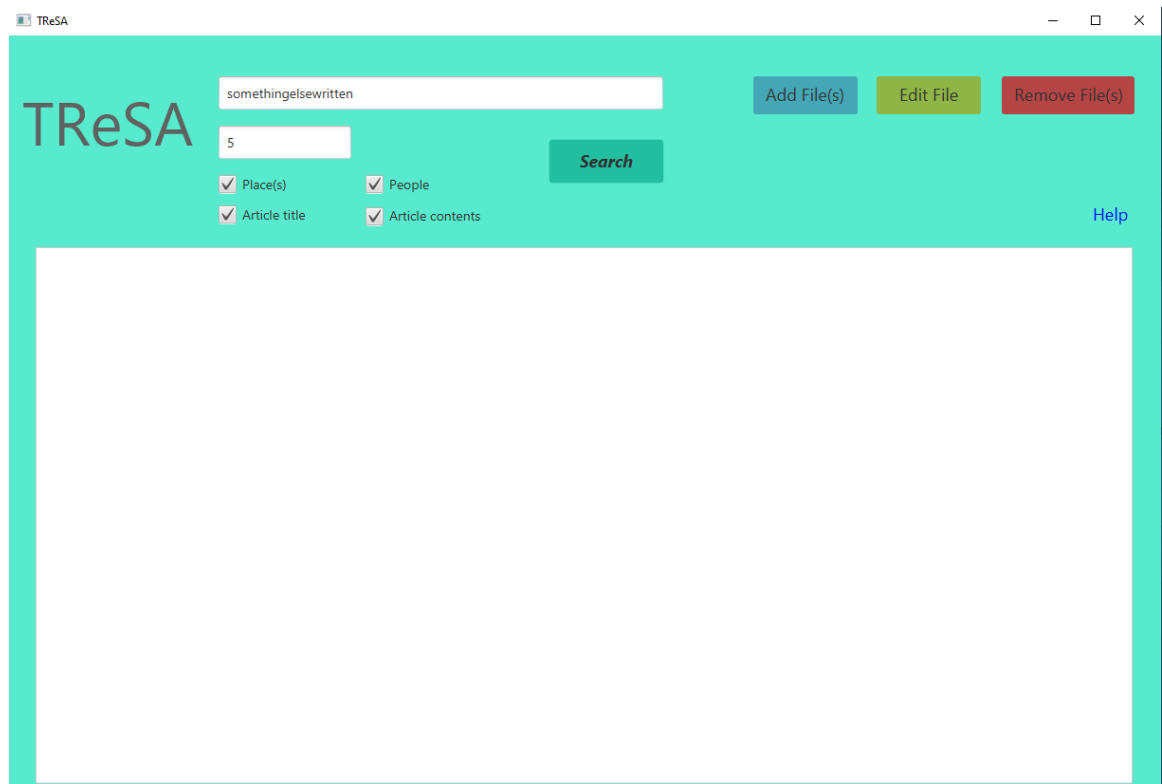
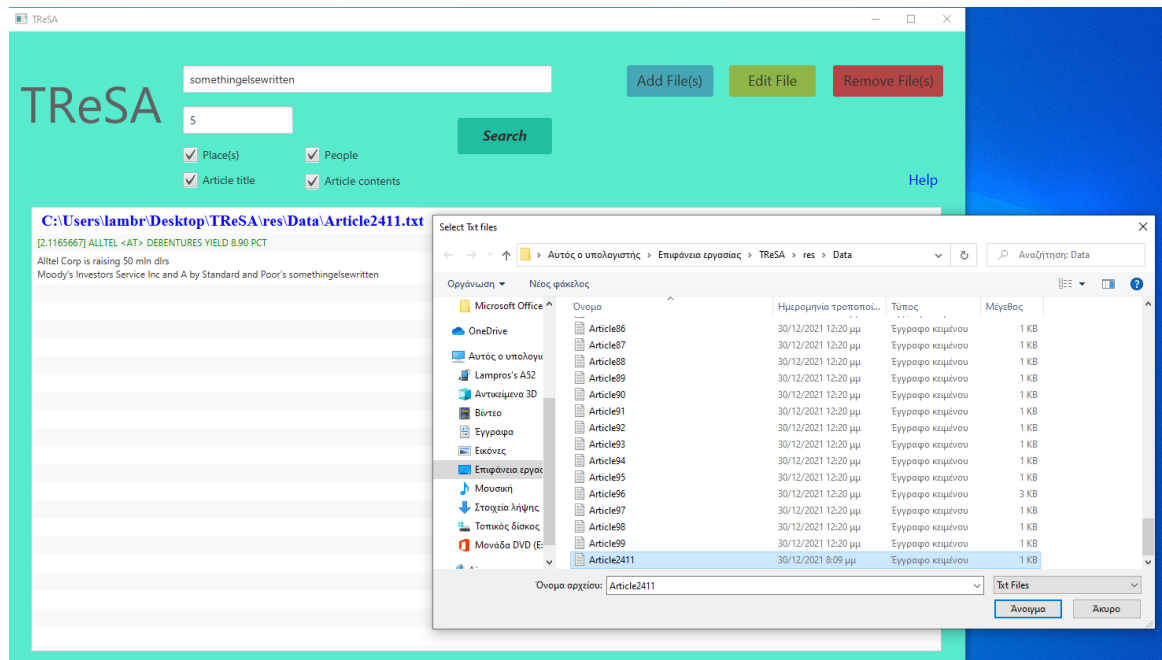
Άνοιγμα

Ακύρο

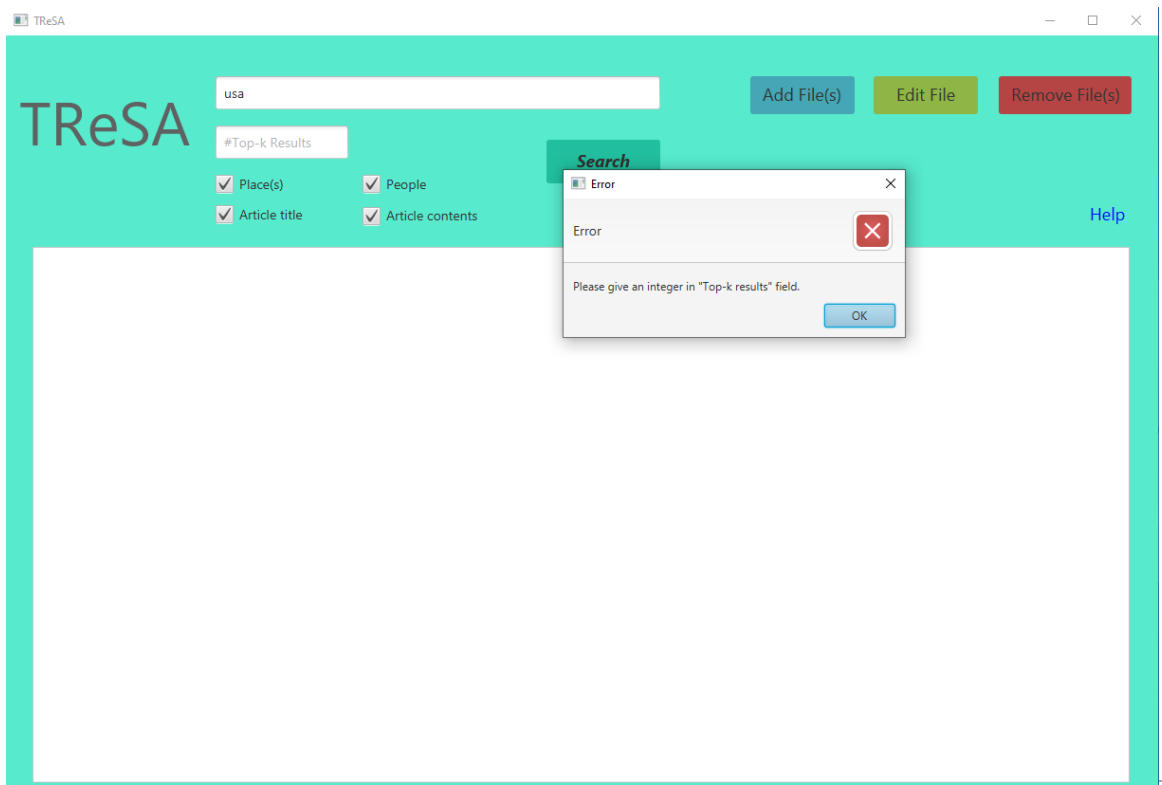
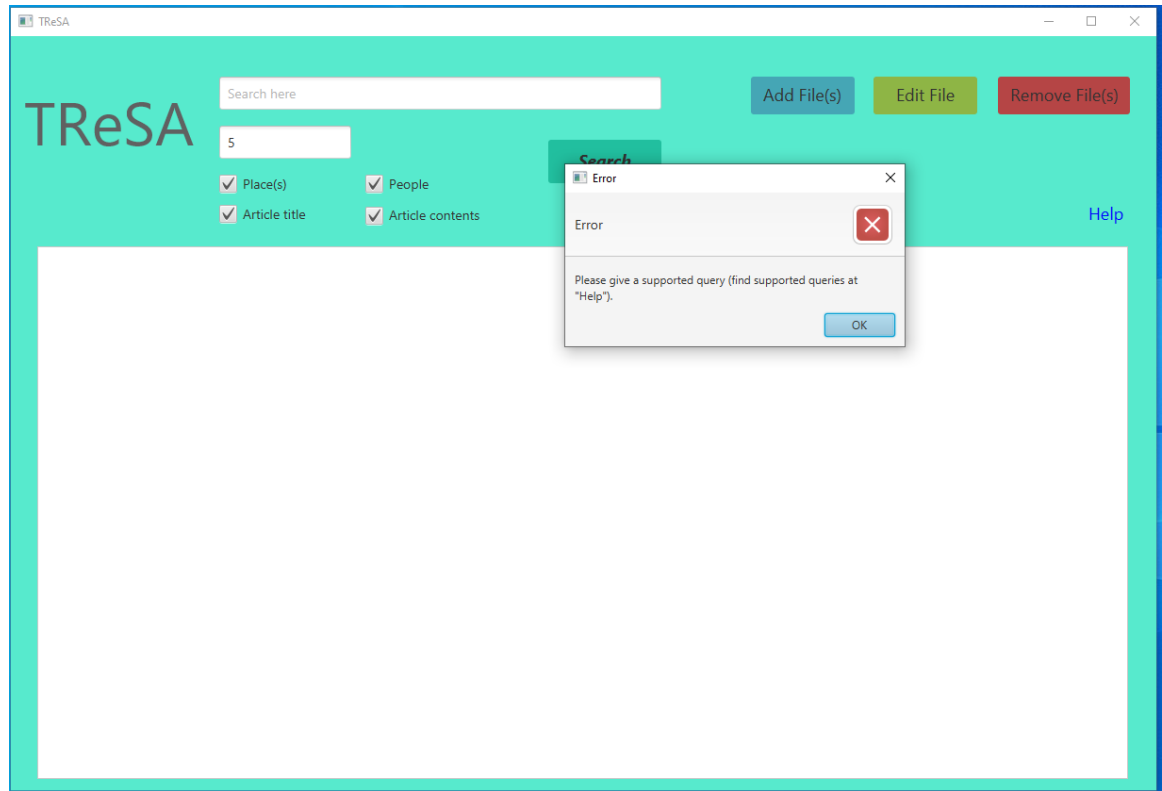
- iv. Στις παρακάτω εικόνες φαίνεται πως κανείς μπορεί να αλλάξει ένα αρχείο της μηχανής αναζήτησης από το κουμπί «Edit File»:

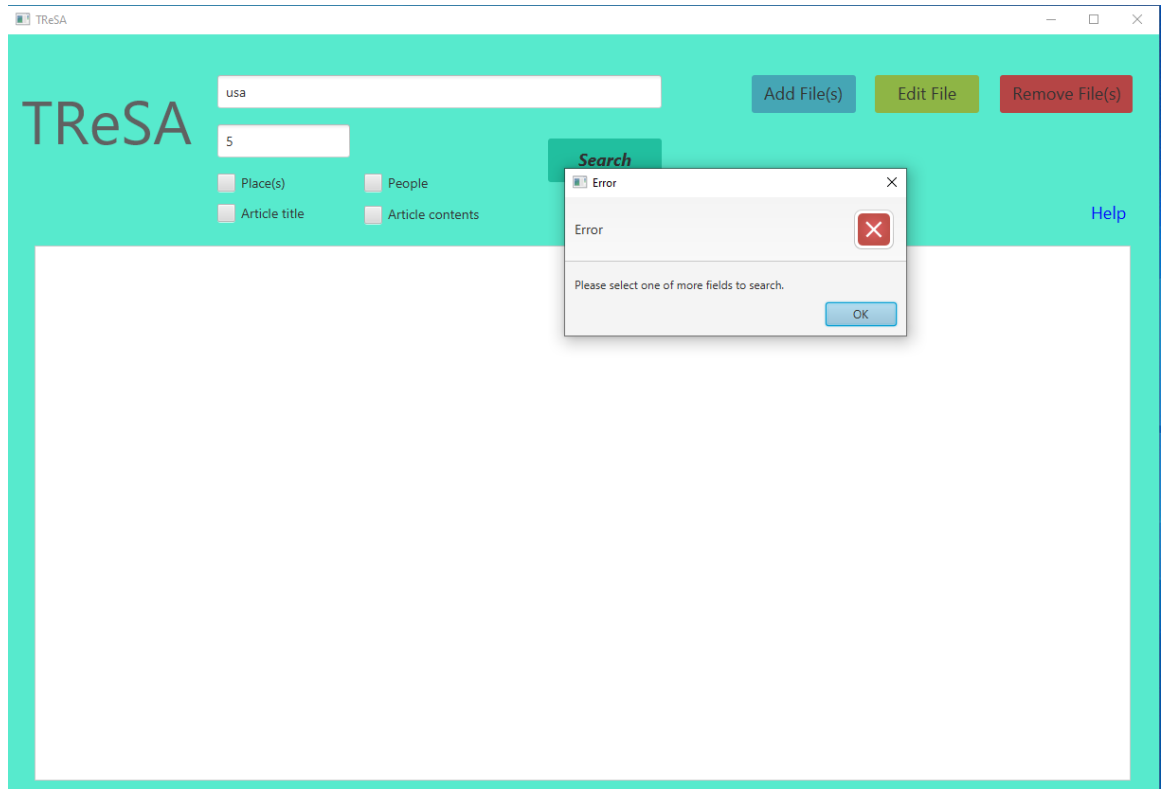


- ν. Στις παρακάτω εικόνες φαίνεται πως κανείς μπορεί να διαγράψει ένα ή περισσότερα αρχεία στην μηχανή αναζήτησης από το κουμπί «Remove File(s)»:



- vi. Στις παρακάτω εικόνες φαίνονται κάποια από τα μηνύματα λάθους «Alerts» της μηχανής αναζήτησης:





Πηγές:

https://lucene.apache.org/core/8_11_0/index.html

<https://www.baeldung.com/lucene-analyzers>

https://www.tutorialspoint.com/lucene/lucene_updatedocument.htm

<https://www.programcreek.com/java-api-examples/org.apache.lucene.search.BooleanQuery>

<https://stackoverflow.com/questions/15728569/multiple-field-query-handling-in-lucene>

<https://stackoverflow.com/questions/45067838/javafx-list-view-coloring-the-list-cell>

<https://stackoverflow.com/questions/48515198/phrasedquery-is-not-working-in-apache-lucene-7-2-1>

<https://mkyong.com/java/how-to-check-if-directory-is-empty-in-java/>