

Η μηχανή αναζήτησης Google

Υλοποίηση

Βιβλιοθήκες εφαρμογής

Η υλοποίηση της άσκησης έγινε με τη χρήση της βιβλιοθήκης **Lucene** (version 6.2.1):

/lib/lucene-core-6.2.1.jar

/lib/lucene-analyzers-common-6.2.1.jar

/lib/lucene-demo-6.2.1.jar

Του html parser **jsoup** (version 1.10.1):

/lib/jsoup-1.10.1.jar

Και του **porter stemmer**, του οποίου ο κώδικας έχει χρησιμοποιηθεί στη κλάση Stemmer του προγράμματος

Γραφικό περιβάλλον εφαρμογής

Για τη δημιουργία του γραφικού περιβάλλοντος της εφαρμογής (καθώς και τη συγγραφή του κώδικα) χρησιμοποιήθηκε το NetBeans IDE 8.0.2 σε συνδυασμό με γραφικά στοιχεία της βιβλιοθήκης swing της java.

Περιβάλλον ανάπτυξης εφαρμογής

Λειτουργικό σύστημα:

Windows edition

Windows 10 Education

© 2016 Microsoft Corporation. All rights reserved.



Έκδοση java:

Java 8 Update 91 (64-bit)	Oracle Corporation	12-May-16	204 MB	8.0.910.14
Java SE Development Kit 8 Update 60 (64-bit)	Oracle Corporation	03-Sep-15	517 MB	8.0.600.27

```
java version "1.8.0_91"  
Java(TM) SE Runtime Environment (build 1.8.0_91-b14)  
Java HotSpot(TM) 64-Bit Server VM (build 25.91-b14, mixed mode)
```

Επεξήγηση βασικών στοιχείων

The screenshot shows the Doogler search interface with the following numbered annotations:

- 1. Search Query input field
- 2. Search in: checkboxes for <PLACES>, <PEOPLE>, <TITLE>, and <BODY>
- 3. Number of top results displayed: 10
- 4. Input Files list area
- 5. Add Files... button
- 6. Delete button
- 7. Search button
- 8. Close button
- 9. Results table area

#	Title	Path	Score
---	-------	------	-------

- Κελί εισαγωγής όρων αναζήτησης
 - Διηματοτητα επεξεργασίας ερωτημάτων vector space πολλών λέξεων
 - Διηματοτητα επεξεργασίας ερωτημάτων boolean (AND, OR, NOT)
 - Δυνατότητα επεξεργασίας ερωτημάτων φράσης (“ερώτημα”)
- Επιλογή των πεδίων στα οποία θα πραγματοποιηθεί η αναζήτηση (<PLACES>, <PEOPLE>, <TITLE>, <BODY>)
- Αριθμός εμφανιζόμενων κορυφαίων αποτελεσμάτων
- Λίστα καταγραφής αρχείων εισόδου
- Κουμπί προσθήκης αρχείου/ων στη λίστα (αυτόματη ενημέρωση των ευρετηρίων)
- Κουμπί διαγραφής επιλεγμένου αρχείου από τη λίστα αρχείων εισόδου (αυτόματη ενημέρωση των ευρετηρίων)
- Κουμπί έναρξης αναζήτησης του ερωτήματος (1.) στα κείμενα των αρχείων της λίστας εισόδου (4.)
- Κουμπί τερματισμού εφαρμογής

9. Πίνακας εμφάνισης αποτελεσμάτων αναζήτησης (στοιχισμένων κατά φθίνουσα σειρά βαθμολογίας) με δυνατότητα εξερεύνησης ολόκληρου του κειμένου με διπλό κλικ σε κάποιο αντικείμενο

Οδηγίες εκτέλεσης

Αρχικά ο χρήστης καλείται να προσθέσει τα αρχεία κειμένων reuters (με κατάληξη .sgm ή .txt) στη μηχανή αναζήτησης, είτε μέσω της επιλογής του μενού File->Add Files..., είτε μέσω του κουμπιού Add Files...

Αφού ο χρήστης επιλέξει τα αρχεία στα οποία θέλει να πραγματοποιηθεί η αναζήτηση δημιουργούνται τα αναγκαία ευρετήρια του Lucene στο φάκελο /index καθώς και γίνεται ο διαχωρισμός των ξεχωριστών κειμένων στο φάκελο /split (διαδικασία διάρκειας 5 δευτερολέπτων για όλα τα 21 αρχεία της συλλογής).

Ο χρήστης μπορεί ανα πάσα στιγμή να διαγράψει ένα αρχείο από τα ευρετήρια της αναζήτησης, επιλέγοντας το στη λίστα και πατώντας το κουμπί Delete.

Μετά την ολοκλήρωση επιλογής λίστας αρχείων προς αναζήτηση ο χρήστης καλείται να συμπληρώσει το ερώτημα του στο κελί πάνω δεξιά, να επιλέξει τα πεδία (tags) στα οποία θα γίνει η αναζήτηση και τον αριθμό των επιθυμητών αποτελεσμάτων πάνω δεξιά.

Παραδείγματα υποστηριζόμενων τύπων ερωτημάτων:

- word1 word2 word3
- word1 AND word2 OR word3
- word1 NOT word2
- “word1 word2 word3”
- Κάθε συνδυασμός των παραπάνω

Έπειτα ο χρήστης μπορεί να κάνει κλικ στο κουμπί Search για την έναρξη της αναζήτησης. Τα αποτελέσματα της αναζήτησης (Τίτλος κειμένου, τοποθεσία αρχείου κειμένου, βαθμολογία) εμφανίζονται στον πίνακα στα δεξιά, στοιχισμένα κατά φθίνουσα σειρά βαθμολογίας. Εκεί ο χρήστης με διπλό κλικ σε κάποιο στοιχείο του πίνακα μπορεί να ανοίξει το αρχείο .txt που περιέχει το κείμενο.

Παραδείγματα εκτέλεσης προγράμματος

Doogle - Βιτούλοδότης Κωνσταντίνος

File Help

Search Query:
minister AND prime

Search in:
☐ <PLACES> ☐ <PEOPLE> ☒ <TITLE> ☒ <BODY>

Input Files:

C:\Users\KONIN\Desktop\data\yeut2-000.sgm
C:\Users\KONIN\Desktop\data\yeut2-001.sgm
C:\Users\KONIN\Desktop\data\yeut2-002.sgm
C:\Users\KONIN\Desktop\data\yeut2-006.sgm
C:\Users\KONIN\Desktop\data\yeut2-011.sgm
C:\Users\KONIN\Desktop\data\yeut2-014.sgm
C:\Users\KONIN\Desktop\data\yeut2-018.sgm

Results:

#	Title	Path	Score
1	ITALIAN PRIME MINISTER CRAXI SAYS HE WILL RESIGN	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-001.sgm\1129.txt	21.43541
2	ITALIAN PRESIDENT FACES PROBLEMS CHOOSING LEADER	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-001.sgm\1900.txt	10.396866
3	ITALIAN GOVERNMENT RESIGNS	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-001.sgm\1183.txt	10.18975
4	CHRISTIAN DEMOCRAT MINISTERS RESIGN	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-014.sgm\14976.txt	9.585439
5	NATIONAL OPPOSITION AHEAD IN N.Z. OPINION POLL	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-002.sgm\2556.txt	9.171602
6	FRENCH GOVERNMENT WINS CONFIDENCE VOTE	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-014.sgm\14825.txt	9.159647
7	FRENCH GOVERNMENT WINS CONFIDENCE VOTE	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-014.sgm\14905.txt	9.159647
8	ITALIAN COALITION MEETS AS GOVERNMENT CRISIS LOOMS	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-000.sgm\219.txt	9.09595
9	OECD SAYS JAPAN RESOLVED TO STIMULATE DEMAND	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-000.sgm\351.txt	8.811401
10	HAWKE DECIDES AGAINST EARLY AUSTRALIAN ELECTION	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-011.sgm\11779.txt	8.782905
11	1988 FRENCH MINISTERIAL BUDGETS FROZEN - SOURCES	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-018.sgm\18552.txt	7.9204082
12	LEBANESE PRESIDENT AGREES NEW PEACE MOVES	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-000.sgm\318.txt	7.7812977
13	SWEDEN TO GO AHEAD WITH S. AFRICAN TRADE SANCTIONS	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-000.sgm\338.txt	7.7812977
14	BOYCOTT FORCES NAKASONE TO ADOPT STOP-GAP BUDGET	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-006.sgm\6403.txt	7.7812977
15	NAKASONE SAYS DOLLAR FALL ONLY TEMPORARY	D:\Files\Documents\NetBeansProjects\LuceneFirstApplication\split\yeut2-018.sgm\18370.txt	7.7812977

Add Files...Delete

SearchClose

Doogle - Βιτουλαδότης Κωνσταντίνος

File

Help

Search Query:

"usa"

Search in:

☒ <PLACES>

☐ <PEOPLE>

☐ <TITLE>

☐ <BODY>

Number of top results displayed: 100

Input Files:

C:\Users\KONIN\Desktop\data\reut2-000.sgm

C:\Users\KONIN\Desktop\data\reut2-001.sgm

C:\Users\KONIN\Desktop\data\reut2-002.sgm

C:\Users\KONIN\Desktop\data\reut2-003.sgm

C:\Users\KONIN\Desktop\data\reut2-004.sgm

C:\Users\KONIN\Desktop\data\reut2-005.sgm

C:\Users\KONIN\Desktop\data\reut2-006.sgm

C:\Users\KONIN\Desktop\data\reut2-007.sgm

C:\Users\KONIN\Desktop\data\reut2-008.sgm

C:\Users\KONIN\Desktop\data\reut2-009.sgm

C:\Users\KONIN\Desktop\data\reut2-010.sgm

C:\Users\KONIN\Desktop\data\reut2-011.sgm

C:\Users\KONIN\Desktop\data\reut2-012.sgm

C:\Users\KONIN\Desktop\data\reut2-013.sgm

C:\Users\KONIN\Desktop\data\reut2-014.sgm

C:\Users\KONIN\Desktop\data\reut2-015.sgm

C:\Users\KONIN\Desktop\data\reut2-016.sgm

C:\Users\KONIN\Desktop\data\reut2-017.sgm

C:\Users\KONIN\Desktop\data\reut2-018.sgm

C:\Users\KONIN\Desktop\data\reut2-019.sgm

C:\Users\KONIN\Desktop\data\reut2-020.sgm

Add Files...

Delete

Results:

#	Title	Path	Score
13	U.S. M-1 MONEY SUPPLY ROSE...	D:\Files\Documents\NetBeansP...	0.4446299
14	GENERAL BINDING <GBND> I...	D:\Files\Documents\NetBeansP...	0.4446299
15	LIBERTY ALL-STAR <USA> SE...	D:\Files\Documents\NetBeansP...	0.4446299
16	COCA COLA <KO> UNIT AND ...	D:\Files\Documents\NetBeansP...	0.4446299
17	FORD MOTOR CREDIT <F> T...	D:\Files\Documents\NetBeansP...	0.4446299
18	STERLING SOFTWARE <SSW>...	D:\Files\Documents\NetBeansP...	0.4446299
19	<SCHULT HOMES CORP> MAK...	D:\Files\Documents\NetBeansP...	0.4446299
20	FLUOR <FLR> UNIT GETS CO...	D:\Files\Documents\NetBeansP...	0.4446299
21	SUFFIELD FINANCIAL CORP <...	D:\Files\Documents\NetBeansP...	0.4446299
22	OHIO MATTRESS <OMT> MAY...	D:\Files\Documents\NetBeansP...	0.4446299
23	<HIGH POINT FINANCIAL COR...	D:\Files\Documents\NetBeansP...	0.4446299
24	LANDMARK BANCSHARES <LB...	D:\Files\Documents\NetBeansP...	0.4446299
25	U.S. GRAIN CARLOADINGS FA...	D:\Files\Documents\NetBeansP...	0.4446299
26	HONG KONG FIRM UPS WRAT...	D:\Files\Documents\NetBeansP...	0.4446299
27	COLECO INDUSTRIES INC <CL...	D:\Files\Documents\NetBeansP...	0.4446299
28	DIAMOND SHAMROCK (DIA) C...	D:\Files\Documents\NetBeansP...	0.4446299
29	LIEBERT CORP <LIEB> APPRO...	D:\Files\Documents\NetBeansP...	0.4446299
30	AM INTERNATIONAL INC <AM...	D:\Files\Documents\NetBeansP...	0.4446299
31	COLECO INDUSTRIES <CLC> ...	D:\Files\Documents\NetBeansP...	0.4446299
32	OLIN CORP <OLM> TO ELECT ...	D:\Files\Documents\NetBeansP...	0.4446299

Search

Close

Doogle - Βιτουλαδότης Κωνσταντίνος

File Help

Search Query:
usa AND guns

Search in:
☒ <PLACES> ☐ <PEOPLE> ☒ <TITLE> ☒ <BODY>

Number of top results displayed: 100

Input Files:
C:\Users\KONIN\Desktop\data\reut2-000.sgm
C:\Users\KONIN\Desktop\data\reut2-001.sgm
C:\Users\KONIN\Desktop\data\reut2-002.sgm
C:\Users\KONIN\Desktop\data\reut2-003.sgm
C:\Users\KONIN\Desktop\data\reut2-004.sgm
C:\Users\KONIN\Desktop\data\reut2-005.sgm
C:\Users\KONIN\Desktop\data\reut2-006.sgm
C:\Users\KONIN\Desktop\data\reut2-007.sgm
C:\Users\KONIN\Desktop\data\reut2-008.sgm
C:\Users\KONIN\Desktop\data\reut2-009.sgm
C:\Users\KONIN\Desktop\data\reut2-010.sgm
C:\Users\KONIN\Desktop\data\reut2-011.sgm
C:\Users\KONIN\Desktop\data\reut2-012.sgm
C:\Users\KONIN\Desktop\data\reut2-013.sgm
C:\Users\KONIN\Desktop\data\reut2-014.sgm
C:\Users\KONIN\Desktop\data\reut2-015.sgm
C:\Users\KONIN\Desktop\data\reut2-016.sgm
C:\Users\KONIN\Desktop\data\reut2-017.sgm
C:\Users\KONIN\Desktop\data\reut2-018.sgm
C:\Users\KONIN\Desktop\data\reut2-019.sgm
C:\Users\KONIN\Desktop\data\reut2-020.sgm


Results:

#	Title	Path	Score
---	-------	------	-------

Add Files...Delete

SearchClose

Nothing Found

 The search returned no results!

OK

