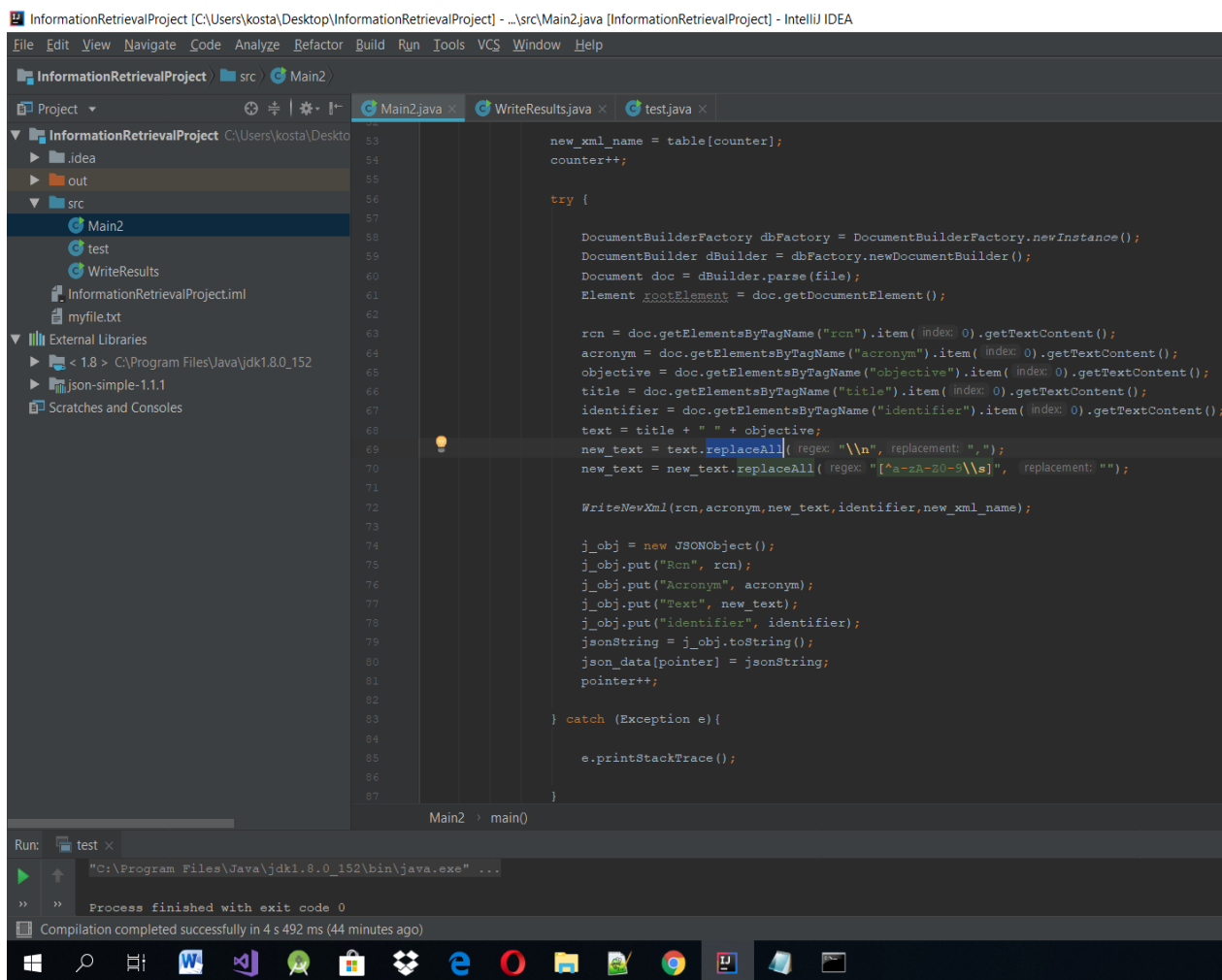


Κεχρής Κωνσταντίνος

ΑΜ: 3150071

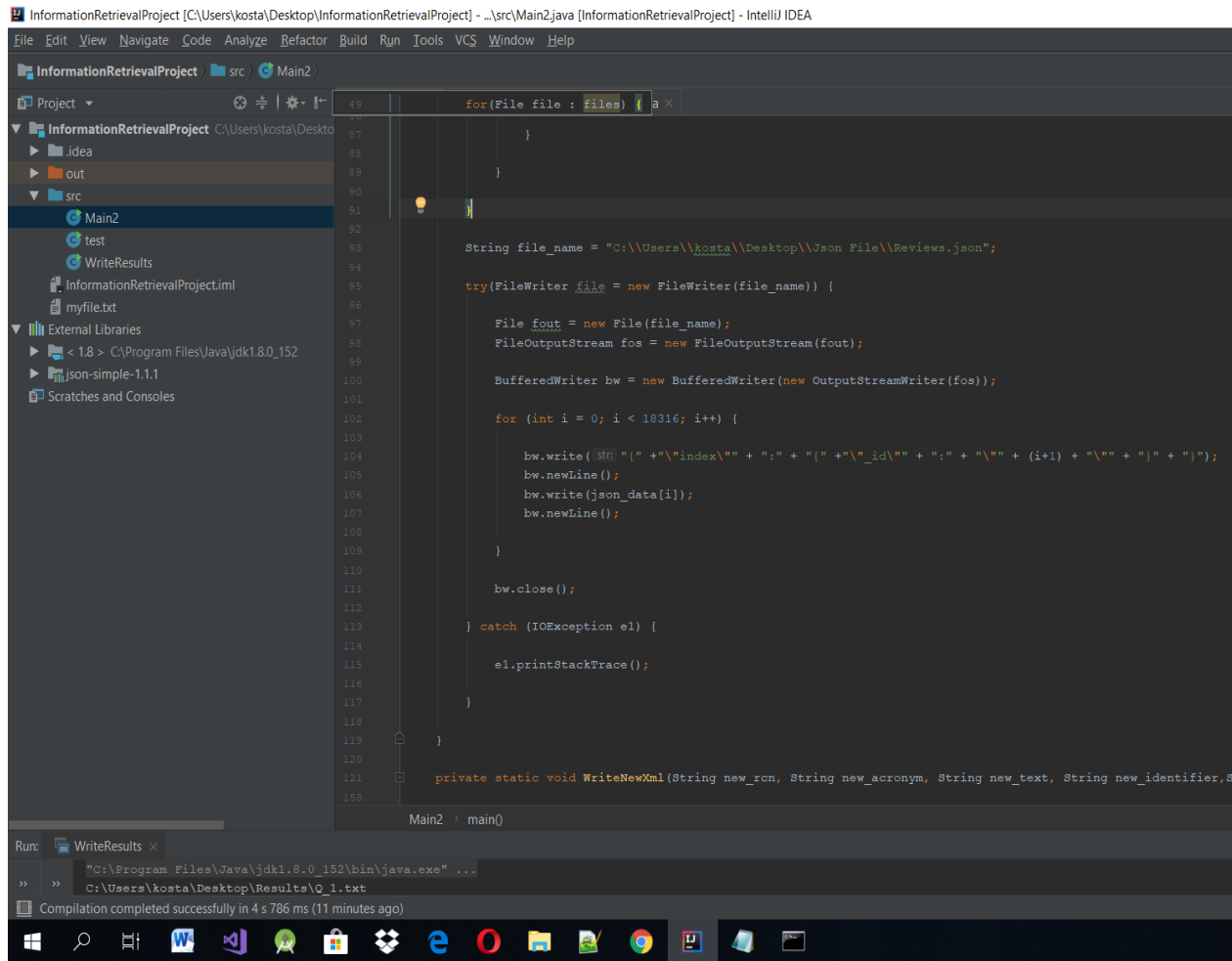
ΦΑΣΗ 1<sup>η</sup>

**1) Ανάλυση του κώδικα:** Αρχικά στην κλάση MAIN κάνω την απαραίτητη προεπεξεργασία των xml αρχείων. Διάβασμα τους με DOMC όπως είδαμε στο αντίστοιχο εργαστήριο. Κατόπιν συνένωση του πεδίου title και objective σε νέο πεδίο text. Αξίζει να αναφερθεί ότι έγινε αφαίρεση ειδικών χαρακτήρων όπως φαίνεται και μαρκαρισμένο στην εικόνα.

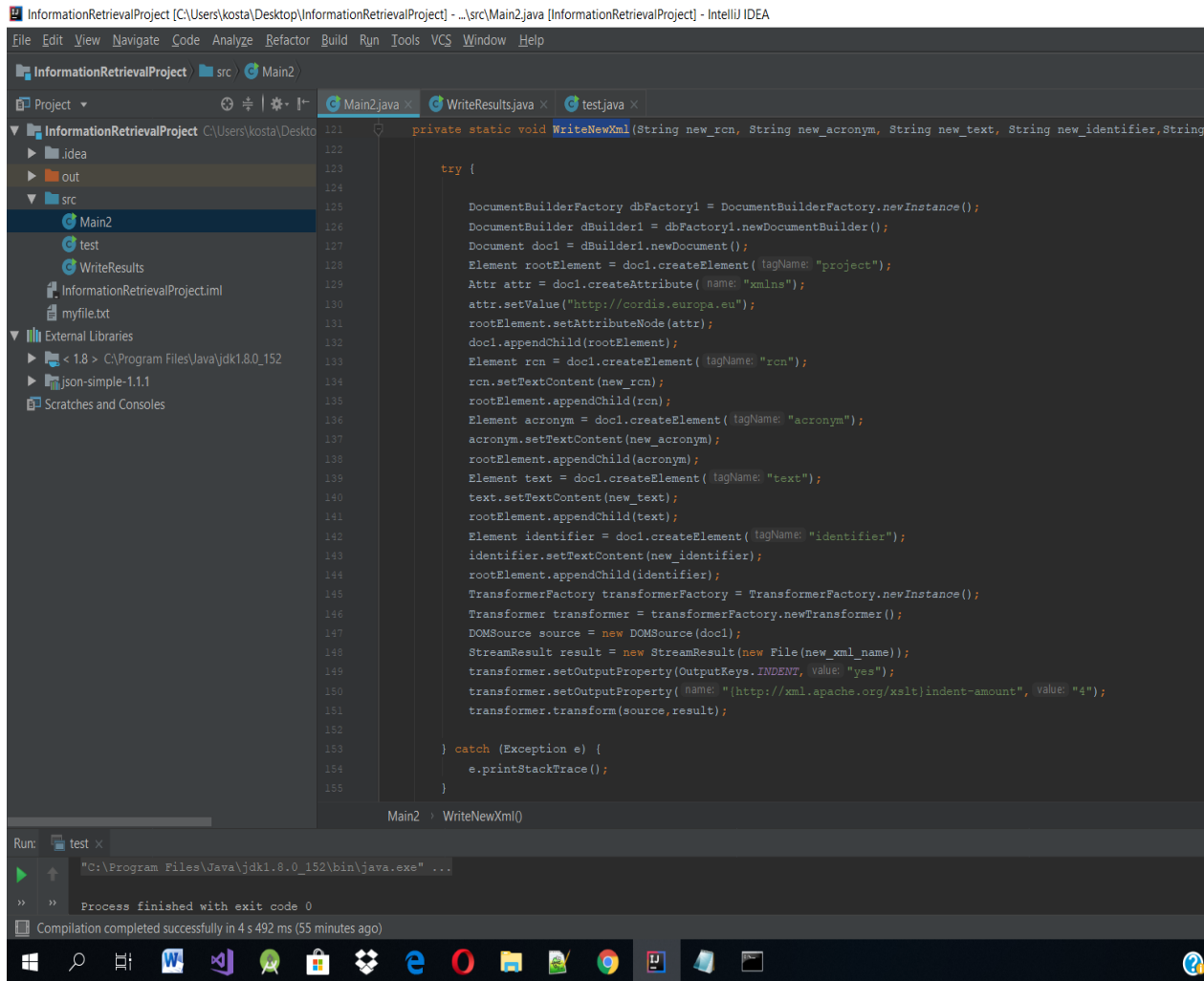


```
53 new_xml_name = table[counter];
54 counter++;
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
```

Έπειτα δημιουργώ αντικείμενο τύπου JSONObject και του περνάω τα απαραίτητα attributes (rcn,text,Acronym,identifier). Στην συνέχεια περνάω όλη την πληροφορία του εκάστοτε JSONObject σε ένα String και αποθήκευση του στον πίνακα json\_data για να το γράψουμε σε αρχείο.

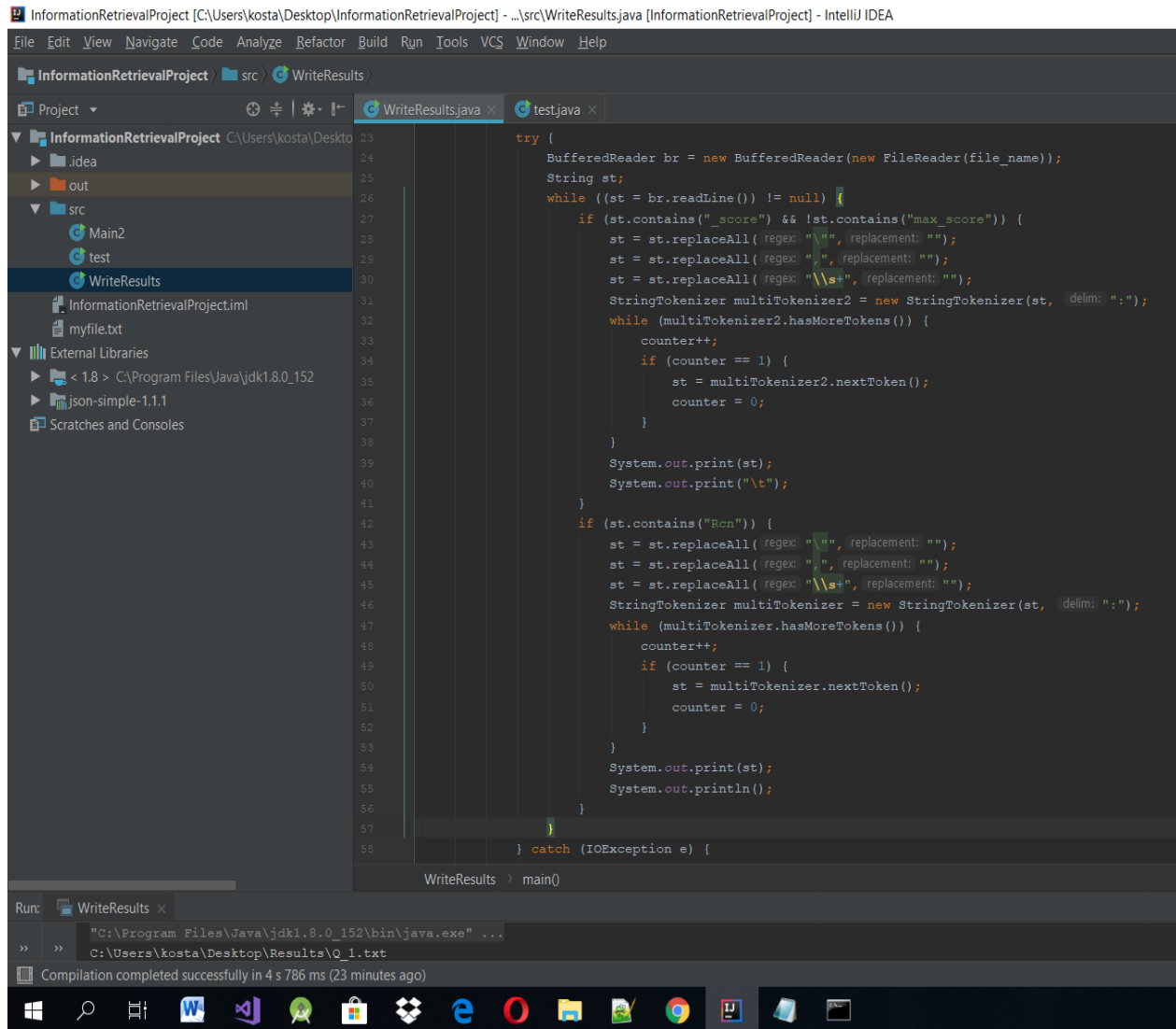


Στην συνέχεια καλώ τη μέθοδο WriteNewXml όπου γράφουμε τα νέα αρχεία xml.



```
121 private static void WriteNewXml(String new_rcn, String new_acronym, String new_text, String new_identifier, String
122
123
124
125     try {
126
127         DocumentBuilderFactory dbFactory1 = DocumentBuilderFactory.newInstance();
128         DocumentBuilder dBuilder1 = dbFactory1.newDocumentBuilder();
129         Document doc1 = dBuilder1.newDocument();
130         Element rootElement = doc1.createElement( "project");
131         Attr attr = doc1.createAttribute( "xmlns");
132         attr.setValue("http://cordis.europa.eu");
133         rootElement.setAttributeNode(attr);
134         doc1.appendChild(rootElement);
135         Element rcn = doc1.createElement( "rcn");
136         rcn.setTextContent(new_rcn);
137         rootElement.appendChild(rcn);
138         Element acronym = doc1.createElement( "acronym");
139         acronym.setTextContent(new_acronym);
140         rootElement.appendChild(acronym);
141         Element text = doc1.createElement( "text");
142         text.setTextContent(new_text);
143         rootElement.appendChild(text);
144         Element identifier = doc1.createElement( "identifier");
145         identifier.setTextContent(new_identifier);
146         rootElement.appendChild(identifier);
147         TransformerFactory transformerFactory = TransformerFactory.newInstance();
148         Transformer transformer = transformerFactory.newTransformer();
149         DOMSource source = new DOMSource(doc1);
150         StreamResult result = new StreamResult(new File(new_xml_name));
151         transformer.setOutputProperty(OutputKeys.INDENT, value: "yes");
152         transformer.setOutputProperty( "http://xml.apache.org/xslt indent-amount", value: "4");
153         transformer.transform(source, result);
154     } catch (Exception e) {
155         e.printStackTrace();
156     }
```

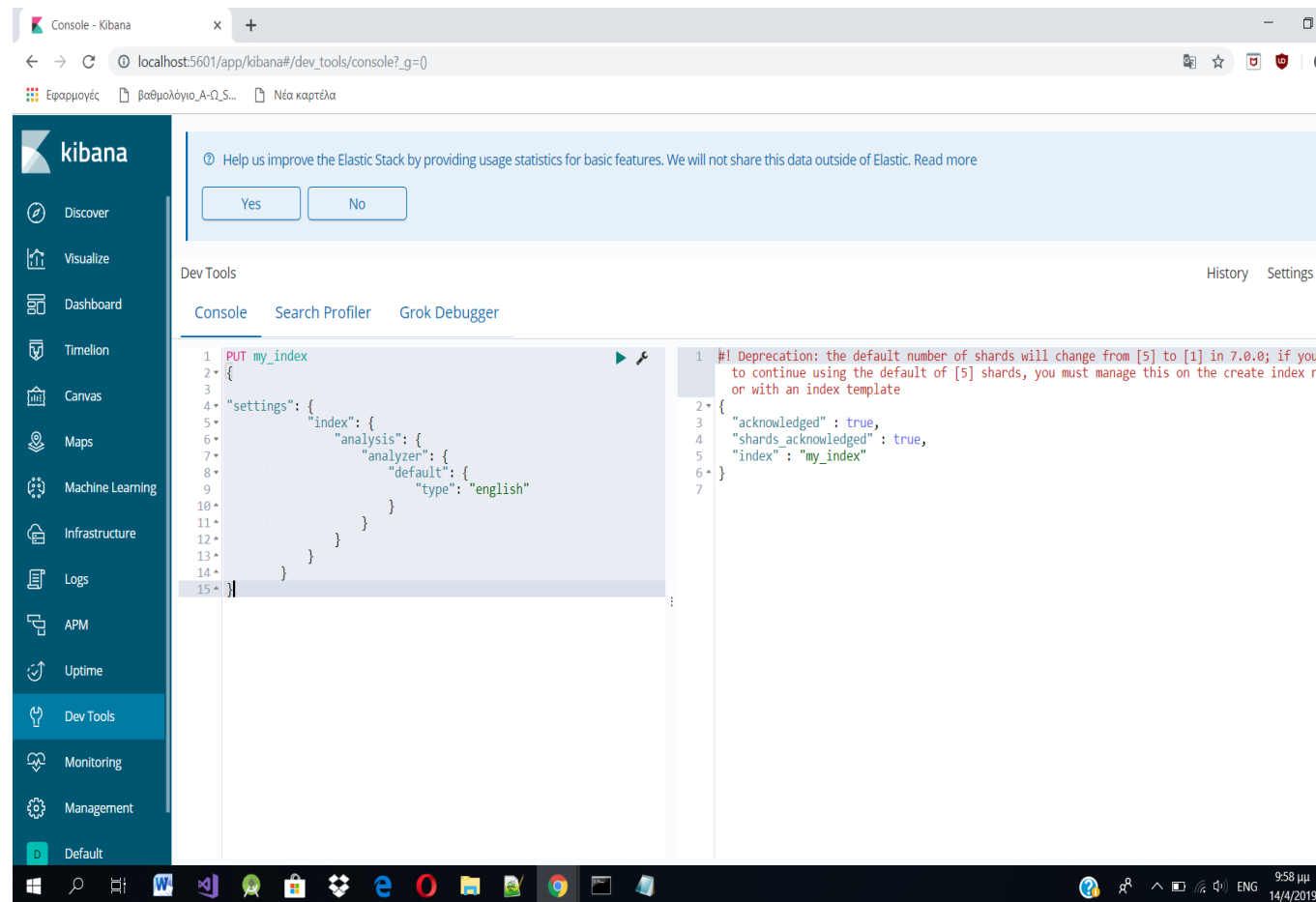
Η κλάση WriteResults χρησιμοποιείται ως εξής. Τα αποτελέσματα των queries που δώσαμε στο kibana τα πήραμε copy paste και τα βάλαμε σε 10 txt αρχεία που βρίσκονται στον φάκελο Results. Απο όλη την πληροφορία κρατάω το Rcn κάθε αρχείου και το Score του. Παίρνω με copy paste το αποτέλεσμα του WriteResults και το τοποθετώ σε νεό αρχείο το new\_res. Και στην κλάση WriteResults έγινε αφαίρεση ειδικών χαρακτήρων όπως φαίνεται και παρακάτω στην εικόνα.



Επίσης έχουμε την κλάση Test όπου χρησιμοποιείται για να δημιουργήσουμε το τελικό αρχείο RESULTS προκειμένου να τρέξουμε το trec\_eval. Τέλος χρησιμοποιώ την κλάση Test2 προκειμένου να καθαρίσω τα δέκα ερωτήματα από ειδικούς χαρακτήρες \n, \t κ.λ.π

## 2)Elasticsearch/kibana:

Αρχικά δημιουργία index με κατάλληλο analyzer. Χρησιμοποίησα τον English μιας και ανταποκρίνεται σε αυτά που ζητάει η άσκηση (stopword removal, transform to lower case,stemming).



Εισαγωγή json με `_bulk?pretty`

Console - Kibana

localhost:5601/app/kibana#/dev\_tools/console?\_g=0

Εφαρμογές βοήθολόγιο\_A-Ω... Νέα καρτέλα

kibana

Help us improve the Elastic Stack by providing usage statistics for basic features. We will not share this data outside of Elastic. Read more

Yes No

Dev Tools History Settings

Console Search Profiler Grok Debugger

```
1 GET /_cat/indices
2
3 POST /my_index/index/_bulk?pretty
4 {"index":{"_id":"1"}}
5 {"identifier":"H2020Adhoc201420","text":"Support towards
the Europe PMC initiativeContribution for 20142016 The
proposed action will provide continued support to the
European Research Council ERC in the implementation of
its Open Access strategy for projects funded in the Life
Sciences domain It follows on from the project Support
towards the Europe PMC initiativeContribution for
2013ERCEuropePMCSUP2013 which has allowed the ERC to
offer the benefits of Europe PMC to its funded
researchers for the first time in 2013 The ERC Open
Access strategy and how the present project will assist
the ERC in its implementation is explained below","rcn"
:"193157","Acronym":"ERC-EuropePMC-1-2014"}
6 {"index":{"_id":"2"}}
7 {"identifier":"H2020Adhoc201420","text":"Support to the
VicePresidents of the ERC Scientific Council 2014 The
proposed Action will provide the necessary support to the
VicePresidents of the European Research Council
Scientific Council ERC ScC to achieve key milestones and
deliverables of the ScC which are required in the first
year of implementation under the Horizon 2020 framework
programmeUnder the Horizon 2020 programme the Commission
has established the European Research Council providing a
competitive funding mechanism for investigatordriven
frontier research on European level One of the key
structural components of the ERC is the Scientific
Council consisting of 22 eminent researchers
```

1 yellow open my\_index 1sx20eQsSC22lwkI56gdYQ 5 1 18316 0 10.3mb 10.3mb  
2 green open .kibana\_task\_manager ZE1nK8Y8R8KOW\_BCSyT\_mw 1 0 2 0 12.5kb 12.5kb  
3 green open .kibana\_1 Wd4gy13JTLC7p4RHD9nmMA 1 0 3 1 16.4kb 16.4kb  
4

10:03 μμ  
14/4/2019

## Έπειτα εκτέλεση ερωτημάτων:

Console - Kibana

localhost:5601/app/kibana#/dev\_tools/console?\_g=0

Εφαρμογές βοήθολόγιο\_A-Ω... Νέα καρτέλα

kibana

Dev Tools History Settings Help

Console Search Profiler Grok Debugger

```
1 POST /my_index/_search
2 {
3   "from":1,"size":20,
4   "query":{"
5     "query_string": {
6       "query": "EUTRAVEL Optimodal European Travel Ecosystem
EuTravel aims to, Support the EU agenda towards an open
and single market for mobility services by enabling
travellers to organise a multimodal trip in accordance
with their own criteria including environmental
performance, providing multimodal travel service
providers an effective way to deliver customised
services addressing any type of specialised travel needs
and facilitating fact-based EU policy making. Promote
the creation of content, open and linked data for
travellers enriching the travelling experience. Support
travel industry players join forces towards realising an
EU shavel seamless mobility strategy and architecture
.EuTravel will research and demonstrate Inter-modal
travel optimised with respect to synchronisation between
modes, passenger experience and rights and environmental
performance (Optimodal Travel).The project objectives
will be realised by. Developing an open and readily
usable Optimodality Framework aimed at integrating
processes, data, and systems in a manner that eliminates
interoperability barriers to the marketplace emergence
of truly Optimodal travel services: from planning
through booking and the full range of related travel
support solutions. Delivering Optimodality Ecosystem
Enablers, offering an open infrastructure that allows
organisations to set up cost-effective integration of
existing systems and to create value added multimodal
travel services. Organising and developing a Living Lab
to experiment and evaluate new concepts and prototype
solutions in real life multimodal travel scenarios and
obtain data to quantify impact. Taking actions towards
sustainable development including a Stakeholder
Engagement Strategy and wide dissemination.EuTravel,
unlike other projects initiatives, will deliver an
```

```
1 {
2   "took": 135,
3   "timed_out": false,
4   "shards": {
5     "total": 5,
6     "successful": 5,
7     "skipped": 0,
8     "failed": 0
9   },
10  "hits": {
11    "total": 18310,
12    "max_score": 716.7888,
13    "hits": [
14      {
15        "index": "my_index",
16        "type": "index",
17        "_id": "205",
18        "_score": 263.2147,
19        "source": {
20          "identifier": "H2020MG2014TwoStages",
21          "text": "INFORMATION TECHNOLOGIES FOR SHIFT TO RAIL The ITRAIL Information
Technologies for Shift to rail proposal first step towards the long term IP4
IT for an Attractive Railway SHIFTRAIL Innovation Programme aims at providing
a new seamless travel experience giving access to a complete multimodal travel
offer which connects the first and last mile to long distance journeysThis is
achieved through the introduction of a ground breaking Technical Enabler based
on two concepts the traveler is placed at the heart of innovative solutions
accessing all multimodal travel services shopping ticketing and tracking
through its travelcompanion An open published framework is providing full
interoperability whilst limiting impacts on existing systems without
prerequisites for centralized standardization This Technical Enabler will be
completely settled in the context of the SHIFTRAIL IP4 and ITRAIL is proposing
a reduced approach to the scale of a specified use case without weakening any
of the key concepts of IP4 such as the usage of Semantic Web technologies meta
planning on distributed data travel companion with a protected and secured
personal wallet stored in the cloud and including the rights to travel The use
case will be defined as a specific instantiation of our open concepts and will
benefit from a completely scalable architecture fully instantiated in IP4 This
approach is addressing all the key challenges of the work program supporting a
```

10:06 μμ  
14/4/2019

### 3)Trec eval:

Τοποθέτησα τα δύο αρχεία Q\_res και το Results στο ίδιο directory που είναι το trec\_eval έτρεξα από cmd την εντολή:

trec\_eval -q Q\_res.txt RESULTS.txt τα αποτελέσματα του cmd τα πήρα copy paste και τα έβαλα σε αρχείο 3150071\_results.txt.