UNIVERSITY OF WATERLOO

Faculty of Engineering

Department of Management Sciences

MSCI 541/720 –HW2

Name: Rui Zhang

Student ID: 20599896

Date: Feburary 4th, 2018

Problem 1:

Screenshots of the code is at the back.
Proof that the code works is integrated with problem 2's figures.
Problem 2:

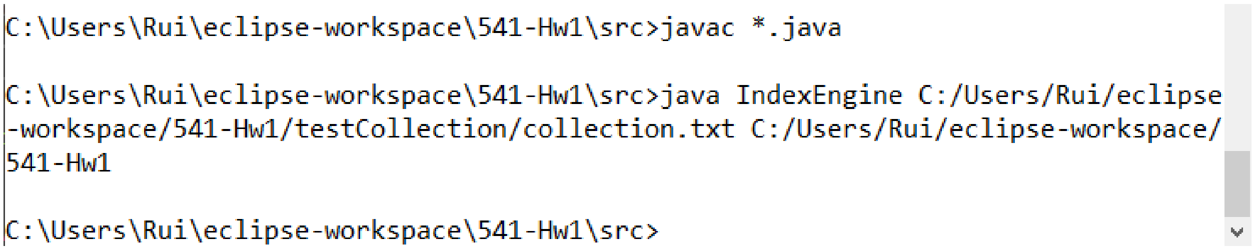
Part A) (I have included as many test cases that I can think of. Please contact me if you need additional items)
Assumptions:
In problem 2, the instructions lists that each document should have its own rank. I assumed that this is meant that the ranking system is based upon per query and not for the entire queries lists. It does not make sense to have the first document for a particular query be ranked based on the previously retrieved document from another query. Therefore after each query, the rank is reset to 1 to the first document retrieved.
In addition, since the score is the number of items retrieved minus the rank, if the query only has one document retrieved, then that would mean that the score is 0. I did not find this acceptable because it doesn't make sense to have a score of 0. Therefore, if the document only retrieved one document, the score would still be 1 but for other cases, it would be number of items retrieved minus one.

Figure	Description
1	Proof that the Index Engine capable of running from command prompt and accepts two arguments: the directory to my test collection file, directory to where the index engine is
2	Proof that the BooleanAND capable of running from command prompt and accepts three arguments: the directory to my test collection results such as the inverted index and lexicon etc, name of the test queries, name of the results file
3	Collection of my test documents separated with DOC tags
4	The output of my Index Engine separating the documents into their own designated folder and files while naming them by the internal id
5	The output results of my BooleanAnd retrieval program
6	The view of what each of the files looks like including metadata, lexicon, inverted index etc. However, the inverted index and the lexicons are saved as bytes through Java serialization
7	A view of the 4 collection documents, queries/topics, and results file. The first topic is "Lorem Ipsum" and since it is on every document, the results file lists the topic ID (501) and displays all 4 documents meaning those 4 documents includes those two words.
8	Only one of the documents have the topic "dummy text" so therefore the program only outputted one document that has the two texts and associated its topicID (502).
9	This figure demonstrates that if none of the documents have the query words then nothing will be inputted into the results file.
10	This figure demonstrates that the program will only output a document to results if the document contains every single word that is in the query. In this example, 3 documents have the word "it" but only one document has the entire query word and therefore only that document is written into the results file.
11	This figure is a confirmation that if the documents contain every single word in the query, then the document would be outputted into the results file. Since 3 of the 4 documents have the word "text", only the three documents are put into the results file. For the results file, one may see that the rank unique per query and the score is just the number of items retrieved minus the rank.

Note: I made my command prompt with black test and white background to save ink.

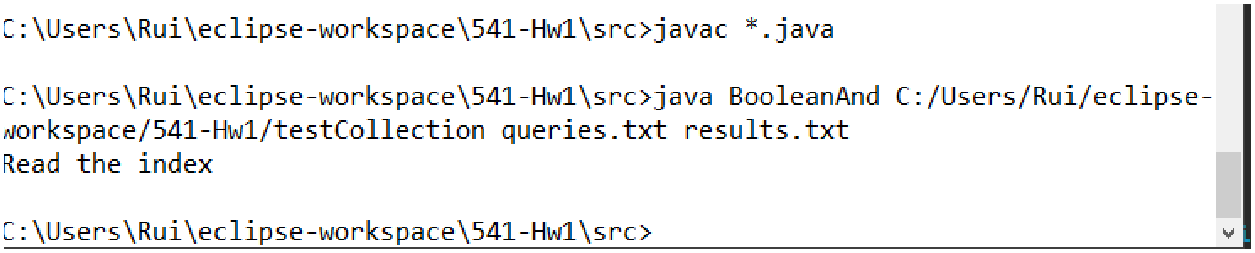
In the results file, you may see the rank and score being counter proportional. As the rank goes up, score goes down. If only one document is retrieved, then the rank and score are both 1 (as discussed in my assumptions above).

Figure 1)



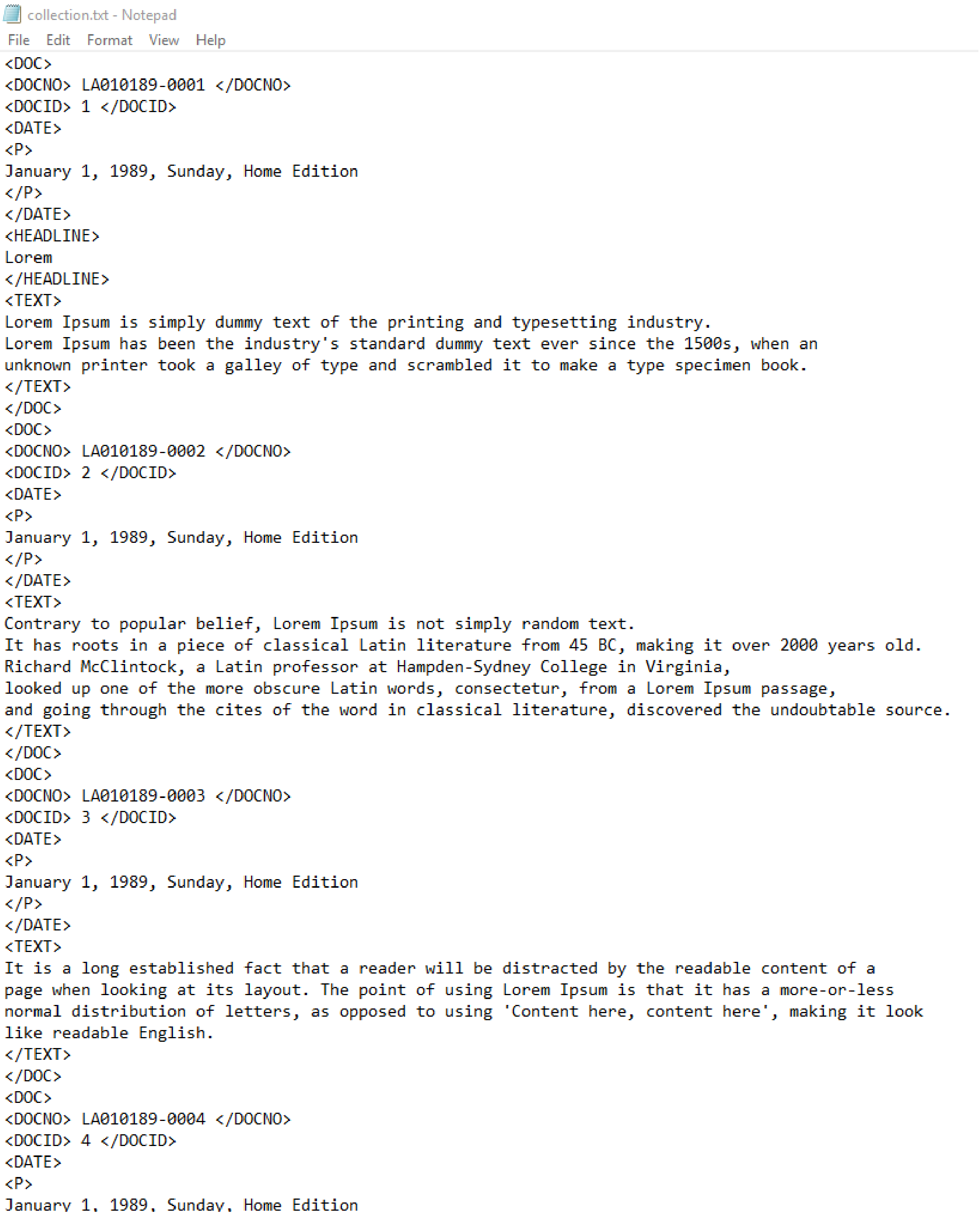
```
C:\Users\Rui\eclipse-workspace\541-Hw1\src>javac *.java

C:\Users\Rui\eclipse-workspace\541-Hw1\src>java IndexEngine C:/Users/Rui/eclipse
-workspace/541-Hw1/testCollection/collection.txt C:/Users/Rui/eclipse-workspace/
541-Hw1

C:\Users\Rui\eclipse-workspace\541-Hw1\src>
```

Figure 2)



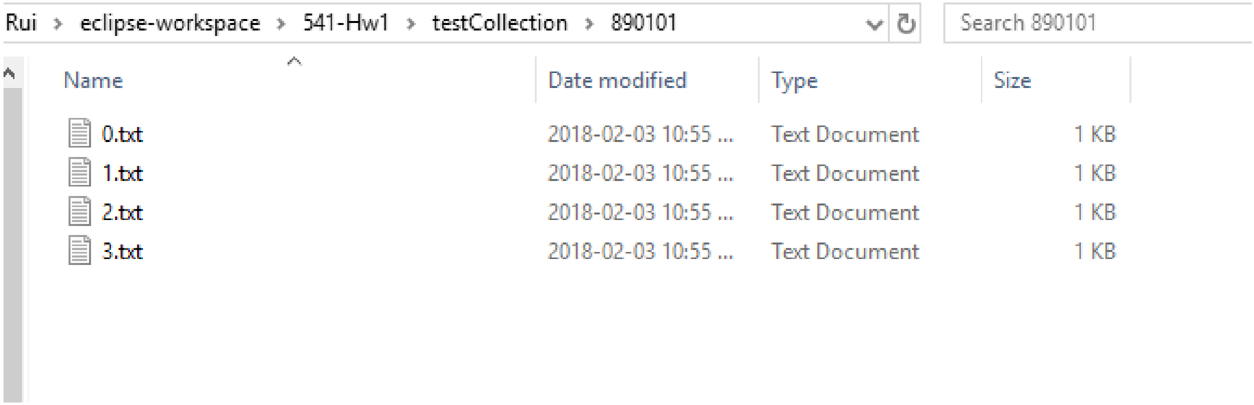
```
C:\Users\Rui\eclipse-workspace\541-Hw1\src>javac *.java

C:\Users\Rui\eclipse-workspace\541-Hw1\src>java BooleanAnd C:/Users/Rui/eclipse-
workspace/541-Hw1/testCollection queries.txt results.txt
Read the index

C:\Users\Rui\eclipse-workspace\541-Hw1\src>
```

Figure 3)



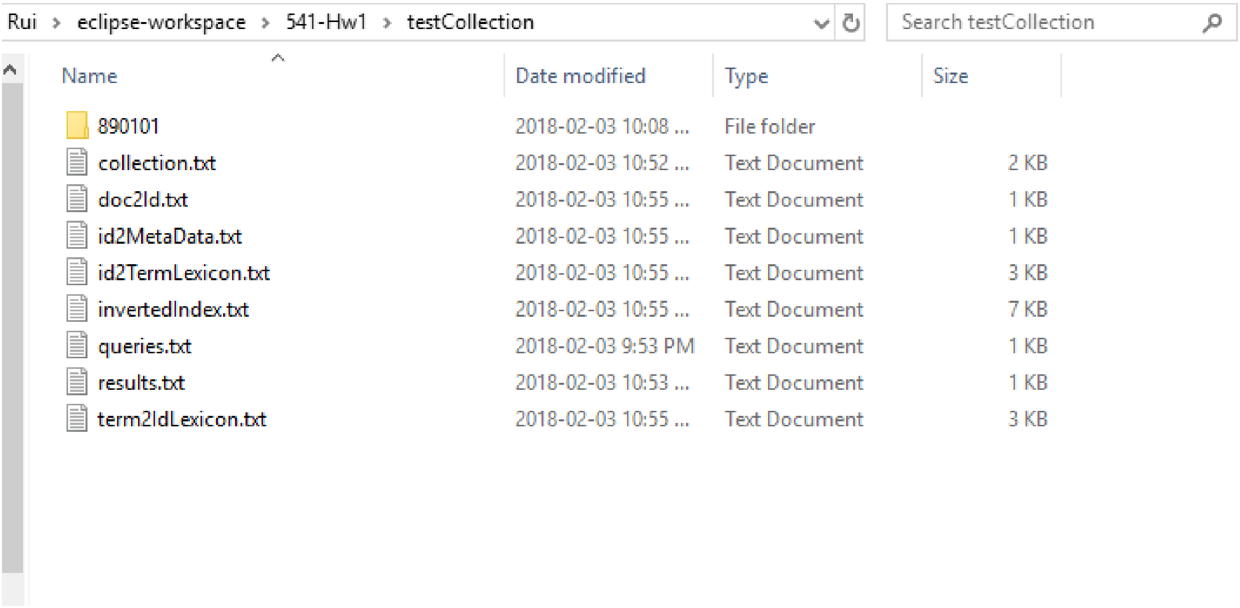
```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is simply dummy text of the printing and typesetting industry.
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an
unknown printer took a galley of type and scrambled it to make a type specimen book.
</TEXT>
</DOC>
<DOC>
<DOCNO> LA010189-0002 </DOCNO>
<DOCID> 2 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<TEXT>
Contrary to popular belief, Lorem Ipsum is not simply random text.
It has roots in a piece of classical Latin literature from 45 BC, making it over 2000 years old.
Richard McClintock, a Latin professor at Hampden-Sydney College in Virginia,
looked up one of the more obscure Latin words, consectetur, from a Lorem Ipsum passage,
and going through the cites of the word in classical literature, discovered the undoubtable source.
</TEXT>
</DOC>
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<TEXT>
It is a long established fact that a reader will be distracted by the readable content of a
page when looking at its layout. The point of using Lorem Ipsum is that it has a more-or-less
normal distribution of letters, as opposed to using 'Content here, content here', making it look
like readable English.
</TEXT>
</DOC>
<DOC>
<DOCNO> LA010189-0004 </DOCNO>
<DOCID> 4 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
```

Figure 4)



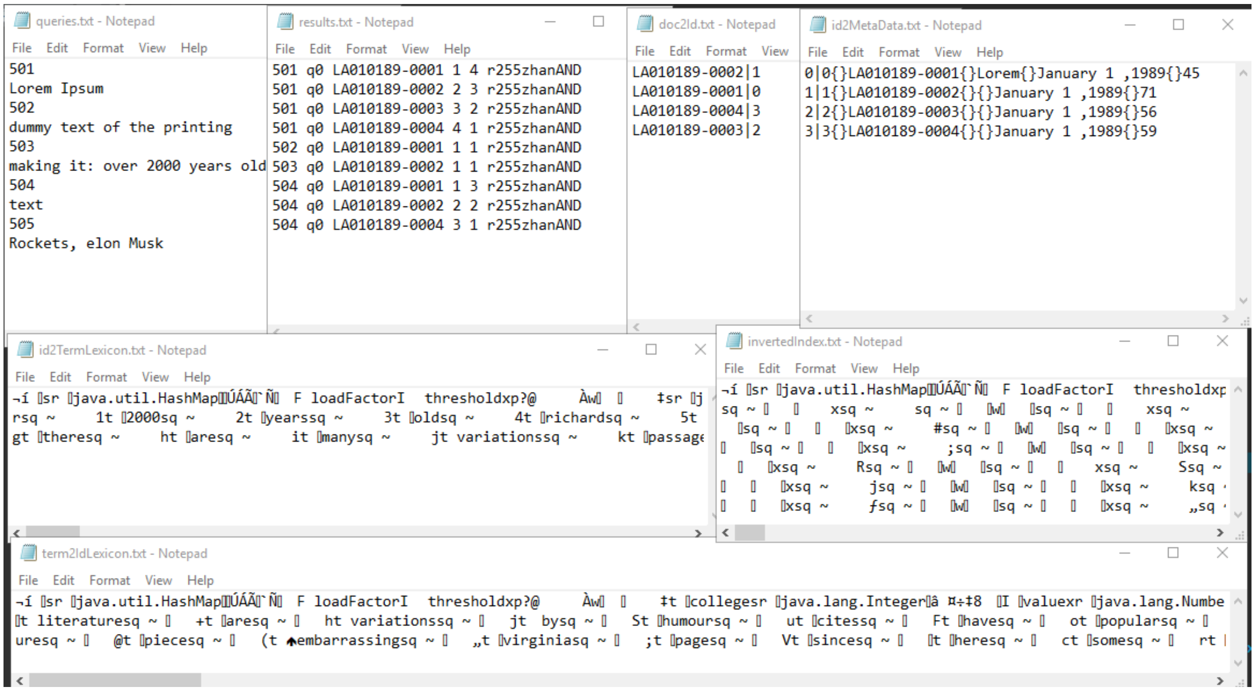
Name	Date modified	Type	Size
0.txt	2018-02-03 10:55 ...	Text Document	1 KB
1.txt	2018-02-03 10:55 ...	Text Document	1 KB
2.txt	2018-02-03 10:55 ...	Text Document	1 KB
3.txt	2018-02-03 10:55 ...	Text Document	1 KB

Figure 5)



Name	Date modified	Type	Size
890101	2018-02-03 10:08 ...	File folder	
collection.txt	2018-02-03 10:52 ...	Text Document	2 KB
doc2Id.txt	2018-02-03 10:55 ...	Text Document	1 KB
id2MetaData.txt	2018-02-03 10:55 ...	Text Document	1 KB
id2TermLexicon.txt	2018-02-03 10:55 ...	Text Document	3 KB
invertedIndex.txt	2018-02-03 10:55 ...	Text Document	7 KB
queries.txt	2018-02-03 9:53 PM	Text Document	1 KB
results.txt	2018-02-03 10:53 ...	Text Document	1 KB
term2IdLexicon.txt	2018-02-03 10:55 ...	Text Document	3 KB

Figure 6)



```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old
504
text
505
Rockets, elon Musk
```

```
501 q0 LA010189-0001 1 4 r255zhanAND
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

```
LA010189-0002|1
LA010189-0001|0
LA010189-0004|3
LA010189-0003|2
```

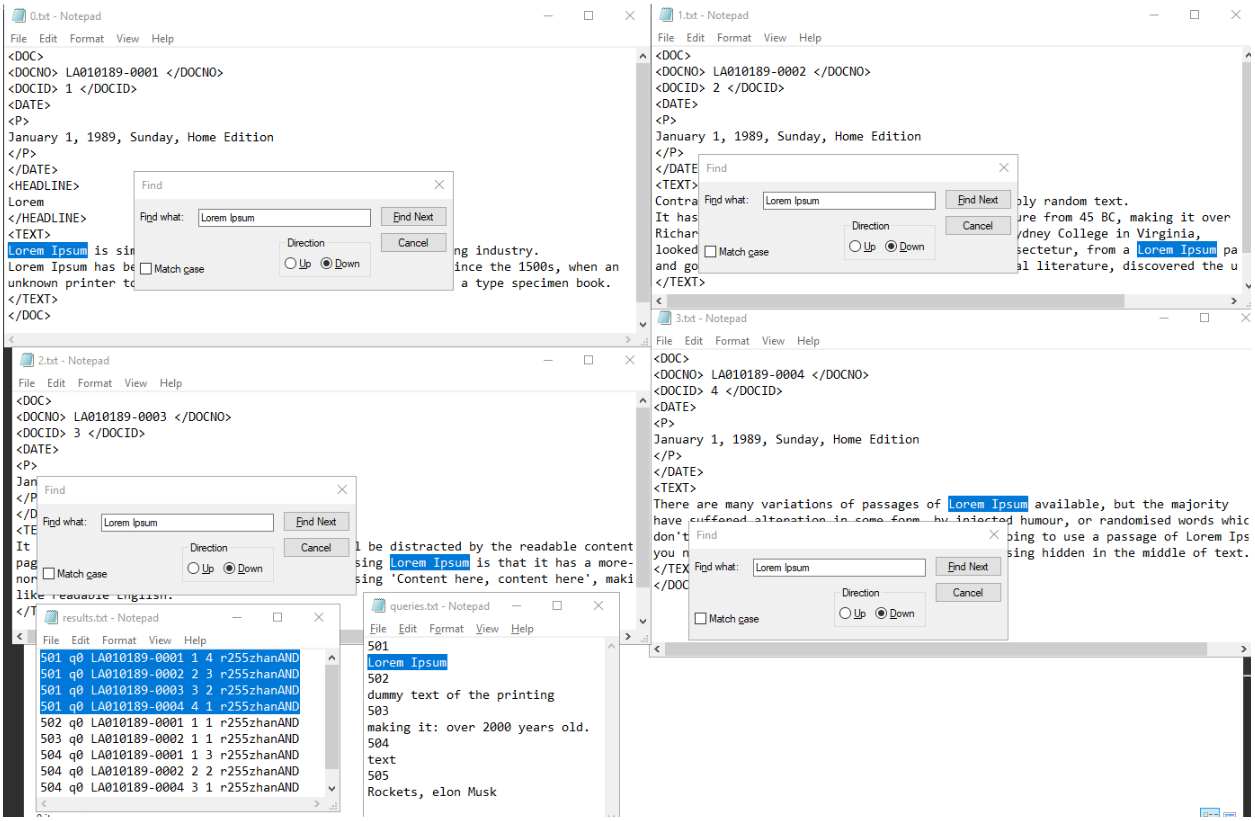
```
0|0{}LA010189-0001{}Lorem{}January 1 ,1989{}45
1|1{}LA010189-0002{}{}January 1 ,1989{}71
2|2{}LA010189-0003{}{}January 1 ,1989{}56
3|3{}LA010189-0004{}{}January 1 ,1989{}59
```

```
¬í ▯sr ▯java.util.HashMap▯▯ÚÁÃ▯`Ñ▯  F loadFactorI   thresholdxp?@      Àw▯  ▯     ‡sr ▯j
rsq ~     1t ▯2000sq ~     2t ▯yearssq ~     3t ▯oldsq ~     4t ▯richardsq ~     5t
gt ▯theresq ~     ht ▯aresq ~     it ▯manysq ~     jt variationssq ~     kt ▯passage
```

```
¬í ▯sr ▯java.util.HashMap▯▯ÚÁÃ▯`Ñ▯  F loadFactorI   thresholdxp
sq ~ ▯   ▯    xsq ~      sq ~ ▯   ▯w▯   ▯sq ~ ▯    ▯    xsq ~
  ▯sq ~ ▯   ▯   ▯xsq ~      #sq ~ ▯   ▯w▯   ▯sq ~ ▯   ▯   ▯xsq ~
▯   ▯sq ~ ▯   ▯   ▯xsq ~      ;sq ~ ▯   ▯w▯   ▯sq ~ ▯   ▯   ▯xsq ~
  ▯   ▯xsq ~      Rsq ~ ▯   ▯w▯   ▯sq ~ ▯   ▯    xsq ~     Ssq ~
▯   ▯   ▯xsq ~      jsq ~ ▯   ▯w▯   ▯sq ~ ▯   ▯   ▯xsq ~     ksq ‹
▯   ▯   ▯xsq ~      ƒsq ~ ▯   ▯w▯   ▯sq ~ ▯   ▯   ▯xsq ~     „sq ‹
```

```
¬í ▯sr ▯java.util.HashMap▯▯ÚÁÃ▯`Ñ▯  F loadFactorI   thresholdxp?@      Àw▯  ▯     ‡t ▯collegesr ▯java.lang.Integer▯â ¤÷‡8  ▯I ▯valuexr ▯java.lang.Numbe
▯t literaturesq ~ ▯   +t ▯aresq ~ ▯   ht variationssq ~ ▯   jt  bysq ~ ▯   St ▯humoursq ~ ▯   ut ▯citessq ~ ▯   Ft ▯havesq ~ ▯   ot ▯popularsq ~ ▯
uresq ~ ▯   @t ▯piecesq ~ ▯   (t ▲embarrassingsq ~ ▯   „t ▯virginiasq ~ ▯   ;t ▯pagesq ~ ▯   Vt ▯sincesq ~ ▯   ▯t ▯heresq ~ ▯   ct ▯somesq ~ ▯   rt |
```

Figure 7)



```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is sin                                    ng industry.
Lorem Ipsum has be                                    ince the 1500s, when an
unknown printer to                                     a type specimen book.
</TEXT>
</DOC>
```

Find what: Lorem Ipsum

Direction ○ Up ● Down

☐ Match case

```
<DOC>
<DOCNO> LA010189-0002 </DOCNO>
<DOCID> 2 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE
<TEXT>
Contra                                  bly random text.
It has                                  ure from 45 BC, making it over
Richar                                  ydney College in Virginia,
looked                                  sectetur, from a Lorem Ipsum pa
and go                                  al literature, discovered the u
</TEXT>
```

Find what: Lorem Ipsum

Direction ○ Up ● Down

☐ Match case

```
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
Jan
</P
</D
<TE
It                                   l be distracted by the readable content
pag                                  sing Lorem Ipsum is that it has a more-
nor                                  sing 'Content here, content here', maki
like readable English.
</T
```

Find what: Lorem Ipsum

Direction ○ Up ● Down

☐ Match case

```
<DOC>
<DOCNO> LA010189-0004 </DOCNO>
<DOCID> 4 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<TEXT>
There are many variations of passages of Lorem Ipsum available, but the majority
have suffered alteration in some form, by injected humour, or randomised words whic
don't t                                 oing to use a passage of Lorem Ips
you n                                   sing hidden in the middle of text.
</TEX
</DOC
```

Find what: Lorem Ipsum

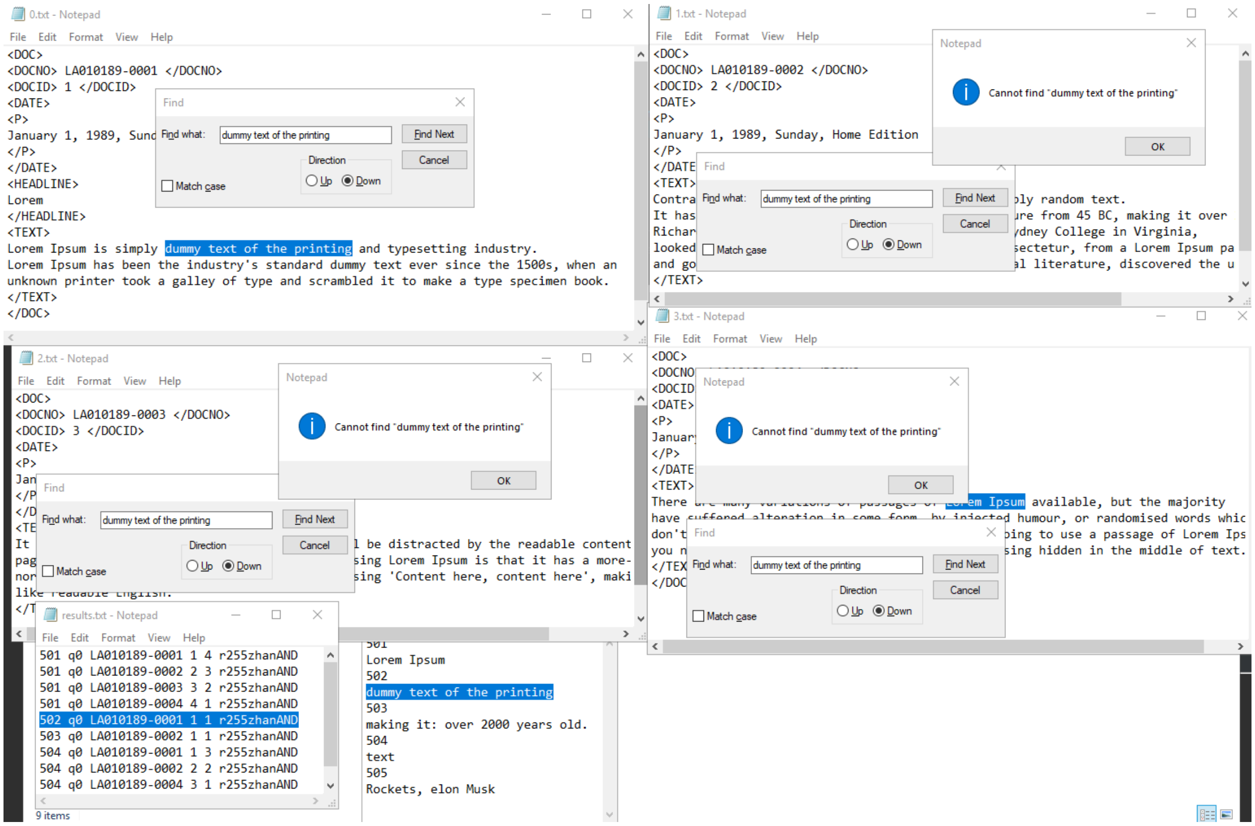
Direction ○ Up ● Down

☐ Match case

```
501 q0 LA010189-0001 1 4 r255zhanAND
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old.
504
text
505
Rockets, elon Musk
```

Figure 8)



0.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sund
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is simply dummy text of the printing and typesetting industry.
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an
unknown printer took a galley of type and scrambled it to make a type specimen book.
</TEXT>
</DOC>
```

Find — Find what: dummy text of the printing — Direction ○ Up ● Down — □ Match case

1.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0002 </DOCNO>
<DOCID> 2 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE
<TEXT>
Contra                                ply random text.
It has                                ure from 45 BC, making it over
Richar                                ydney College in Virginia,
looked                                sectetur, from a Lorem Ipsum pa
and go                                al literature, discovered the u
</TEXT>
```

Notepad: Cannot find "dummy text of the printing"

Find — Find what: dummy text of the printing — Direction ○ Up ● Down — □ Match case

2.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
Jan
</P
</D
<TE
It                                    l be distracted by the readable content
pag                                   sing Lorem Ipsum is that it has a more-
nor                                   sing 'Content here, content here', maki
like readable English.
</T
```

Notepad: Cannot find "dummy text of the printing"

Find — Find what: dummy text of the printing — Direction ○ Up ● Down — □ Match case

3.txt - Notepad

```
<DOC>
<DOCNO
<DOCID
<DATE>
<P>
Januar
</P>
</DATE
<TEXT>
There are many variations of passages of Lorem Ipsum available, but the majority
have suffered alteration in some form, by injected humour, or randomised words whic
don't                                  ping to use a passage of Lorem Ips
you n                                  sing hidden in the middle of text.
</TEX
</DOC
```

Notepad: Cannot find "dummy text of the printing"

Find — Find what: dummy text of the printing — Direction ○ Up ● Down — □ Match case

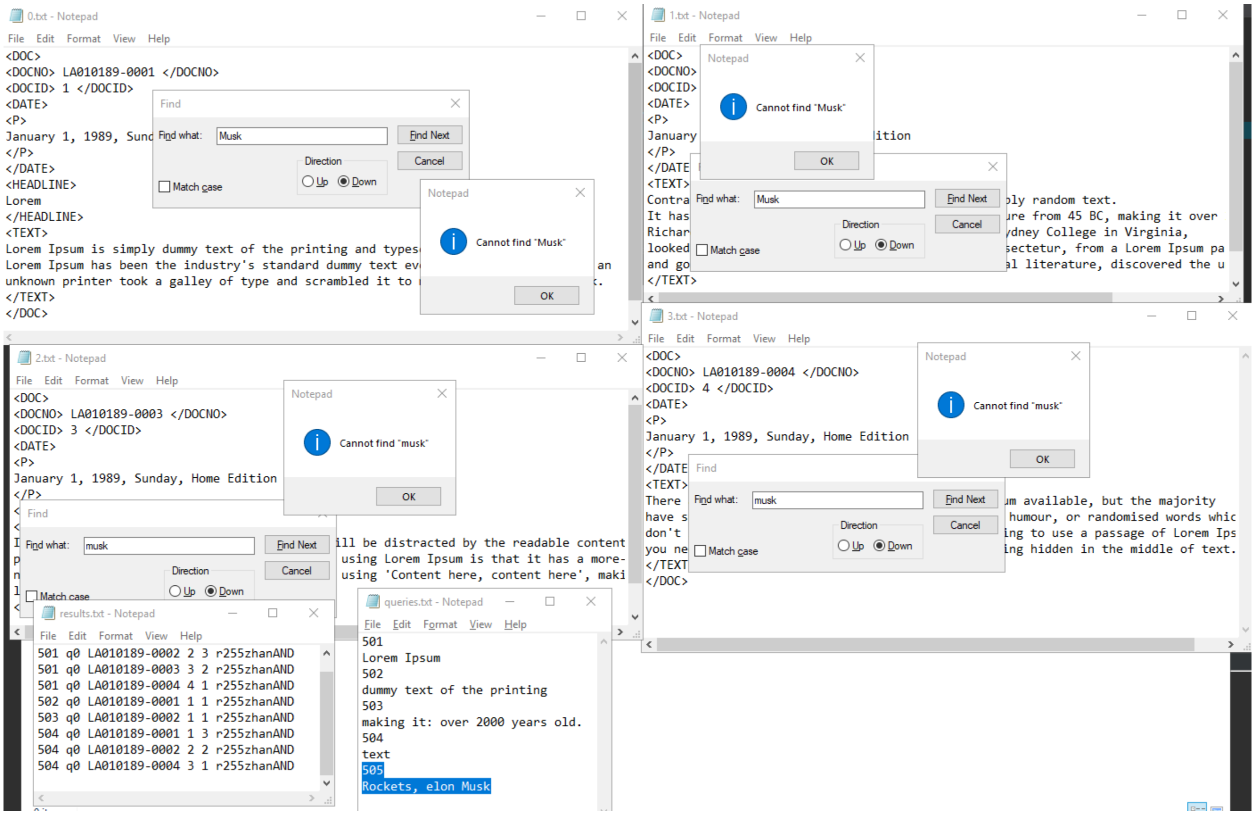
results.txt - Notepad

```
501 q0 LA010189-0001 1 4 r255zhanAND
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

9 items

```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old.
504
text
505
Rockets, elon Musk
```

Figure 9)



0.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sund
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is simply dummy text of the printing and types
Lorem Ipsum has been the industry's standard dummy text ev                  an
unknown printer took a galley of type and scrambled it to                   .
</TEXT>
</DOC>
```

Find — Find what: Musk — Direction ○ Up ● Down — □ Match case

Notepad: Cannot find "Musk"

1.txt - Notepad

```
<DOC>
<DOCNO>
<DOCID>
<DATE>
<P>
January                         ition
</P>
</DATE
<TEXT>
Contra                                ply random text.
It has                                ure from 45 BC, making it over
Richar                                ydney College in Virginia,
looked                                sectetur, from a Lorem Ipsum pa
and go                                al literature, discovered the u
</TEXT>
```

Notepad: Cannot find "Musk"

Find — Find what: Musk — Direction ○ Up ● Down — □ Match case

2.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
                                      ill be distracted by the readable content
                                      using Lorem Ipsum is that it has a more-
                                      using 'Content here, content here', maki
```

Notepad: Cannot find "musk"

Find — Find what: musk — Direction ○ Up ● Down — □ Match case

3.txt - Notepad

```
<DOC>
<DOCNO> LA010189-0004 </DOCNO>
<DOCID> 4 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE
<TEXT>
There                                 m available, but the majority
have s                                humour, or randomised words whic
don't                                 ing to use a passage of Lorem Ips
you ne                                ing hidden in the middle of text.
</TEXT
</DOC>
```

Notepad: Cannot find "musk"

Find — Find what: musk — Direction ○ Up ● Down — □ Match case

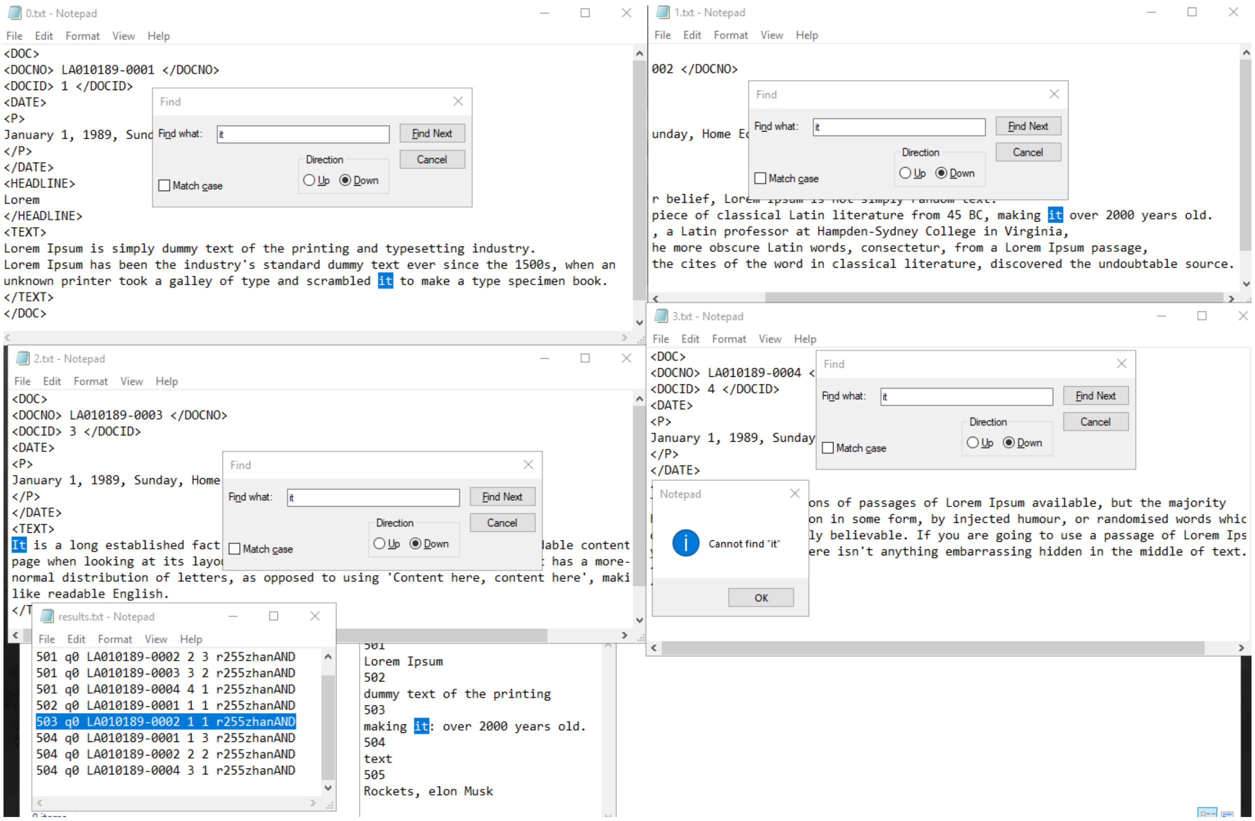
results.txt - Notepad

```
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

queries.txt - Notepad

```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old.
504
text
505
Rockets, elon Musk
```

Figure 10)



```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sund
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is simply dummy text of the printing and typesetting industry.
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an
unknown printer took a galley of type and scrambled it to make a type specimen book.
</TEXT>
</DOC>
```

Find what: it

```
002 </DOCNO>

unday, Home E

r belief, Lorem
piece of classical Latin literature from 45 BC, making it over 2000 years old.
, a Latin professor at Hampden-Sydney College in Virginia,
he more obscure Latin words, consectetur, from a Lorem Ipsum passage,
the cites of the word in classical literature, discovered the undoubtable source.
```

Find what: it

```
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home
</P>
</DATE>
<TEXT>
It is a long established fact                                        able content
page when looking at its layo                                        has a more-
normal distribution of letters, as opposed to using 'Content here, content here', maki
like readable English.
</T
```

Find what: it

```
<DOC>
<DOCNO> LA010189-0004 <
<DOCID> 4 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday
</P>
</DATE>
                       ons of passages of Lorem Ipsum available, but the majority
                       on in some form, by injected humour, or randomised words whic
                       ly believable. If you are going to use a passage of Lorem Ips
                       ere isn't anything embarrassing hidden in the middle of text.
```

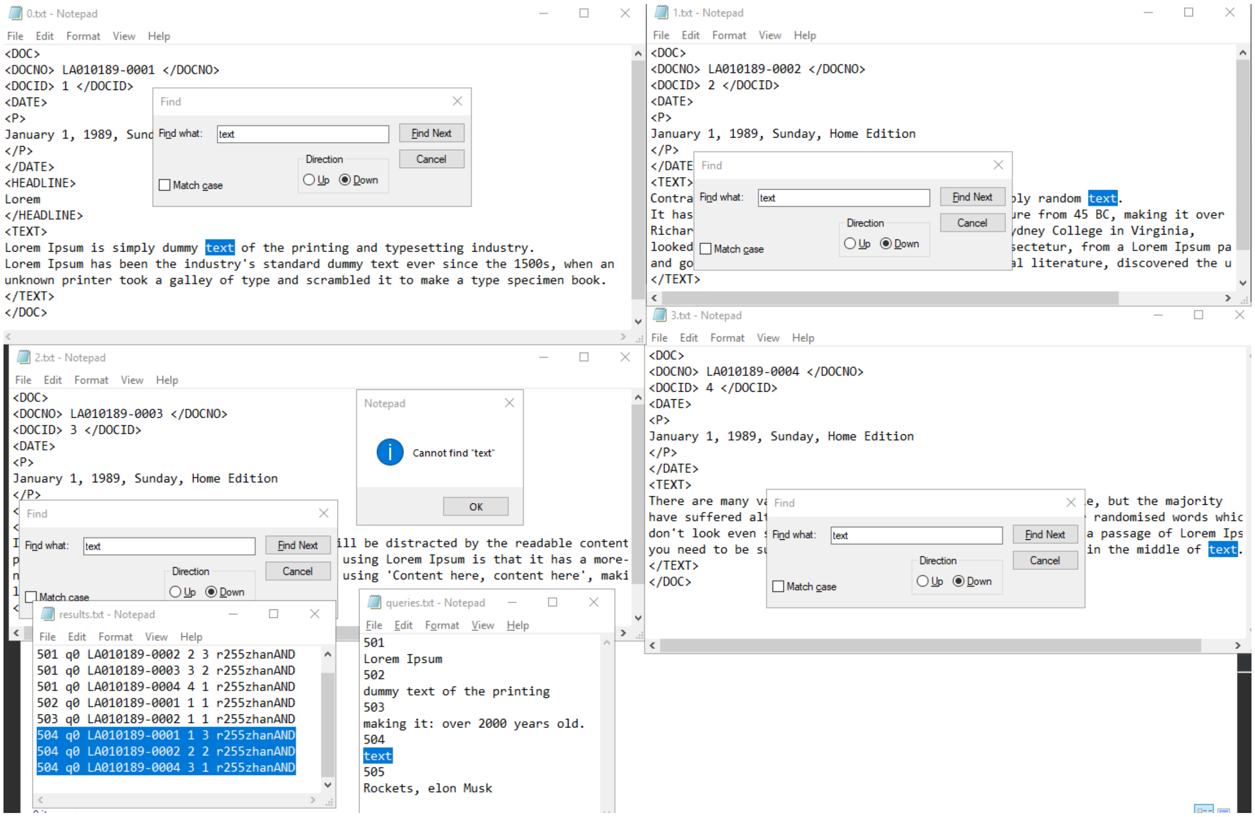
Find what: it

Cannot find "it"

```
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old.
504
text
505
Rockets, elon Musk
```

Figure 11)



```
<DOC>
<DOCNO> LA010189-0001 </DOCNO>
<DOCID> 1 </DOCID>
<DATE>
<P>
January 1, 1989, Sund
</P>
</DATE>
<HEADLINE>
Lorem
</HEADLINE>
<TEXT>
Lorem Ipsum is simply dummy text of the printing and typesetting industry.
Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an
unknown printer took a galley of type and scrambled it to make a type specimen book.
</TEXT>
</DOC>
```

Find what: text

```
<DOC>
<DOCNO> LA010189-0002 </DOCNO>
<DOCID> 2 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE
<TEXT>
Contra                                 ply random text.
It has                                 ure from 45 BC, making it over
Richar                                 ydney College in Virginia,
looked                                 sectetur, from a Lorem Ipsum pa
and go                                 al literature, discovered the u
</TEXT>
```

Find what: text

```
<DOC>
<DOCNO> LA010189-0003 </DOCNO>
<DOCID> 3 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
<
<
I                                       ill be distracted by the readable content
p                                       using Lorem Ipsum is that it has a more-
n                                       using 'Content here, content here', maki
l
<
```

Cannot find "text"

Find what: text

```
<DOC>
<DOCNO> LA010189-0004 </DOCNO>
<DOCID> 4 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<TEXT>
There are many va                             e, but the majority
have suffered alt                             randomised words whic
don't look even s                             a passage of Lorem Ips
you need to be su                             in the middle of text.
</TEXT>
</DOC>
```

Find what: text

```
501 q0 LA010189-0002 2 3 r255zhanAND
501 q0 LA010189-0003 3 2 r255zhanAND
501 q0 LA010189-0004 4 1 r255zhanAND
502 q0 LA010189-0001 1 1 r255zhanAND
503 q0 LA010189-0002 1 1 r255zhanAND
504 q0 LA010189-0001 1 3 r255zhanAND
504 q0 LA010189-0002 2 2 r255zhanAND
504 q0 LA010189-0004 3 1 r255zhanAND
```

```
501
Lorem Ipsum
502
dummy text of the printing
503
making it: over 2000 years old.
504
text
505
Rockets, elon Musk
```

Problem 3)

Topic 401:

The first ten documents were:

1) 401 q0 LA122990-0070 1 14 r255zhanAND
2) 401 q0 LA040389-0047 2 13 r255zhanAND
3) 401 q0 LA040490-0003 3 12 r255zhanAND
4) 401 q0 LA021890-0100 4 11 r255zhanAND
5) 401 q0 LA052190-0065 5 10 r255zhanAND
6) 401 q0 LA100889-0019 6 9 r255zhanAND
7) 401 q0 LA082690-0052 7 8 r255zhanAND
8) 401 q0 LA090490-0093 8 7 r255zhanAND
9) 401 q0 LA050590-0114 9 6 r255zhanAND
10) 401 q0 LA050789-0068 10 5 r255zhanAND

1) This document with the headline "Israel facing $35 billion tab to settle soviets" is NOT relevant to the topic. Although the document contains the topic's words, this document talks about the financial crisis that Israel economy faces which does not have elaborate on the language and cultural differences delaying the integration of foreign minorities in Germany. Germany was only mentioned as an example to be one of the countries to help finance immigration. Therefore this document is not relevant.
2) This document with the headline "NATO'S Success forcing changes; on 40th anniversary; alliance is moving towards new strategy" is NOT relevant to the topic. This document does contain all the key words of the topic but is focused on how the "North Atlantic Treaty Organization has been the most successful defense alliance of the century". Germany was mentioned only be the holding short-range nuclear missiles. Therefore this document is not relevant.
3) This document with the headline "Soviets toughens secession rules' independence: The new law is immediately rejected by representatives of the restive Baltic republics" talks about the "pressures for independence building" up in the Soviet empire and does not focus on the immigration problems with Germany and therefore is not relevant to this topic.
4) This document with the headline "Evil lies in system, not in the race; Germany: it's not the predisposition of a people but the quest of a dictator for total control that leads to aggression and war" in my opinion is relevant to the topic of foreign minorities in Germany. This document talks about the history of Germany under Hilter's regime and how other countries "share [the] reluctance to see the reconstruction of a powerful Germany in the heart of Europe" due to the World War. The document describes how the system that Germany was previously using: "totalitarian politics" that separates the cultural differences and therefore result in immigration difficulties that led to Germany's dark past. In other words, the system that Germany previously had did not welcome immigrants well and treated their own race above all else. Therefore this document is relevant.
5) The document with the headline "Romanians go to polls, warily test democracy" is not relevant to this topic. This document is about how Romanians are testing democracy to be their government system and details about the process and factors that are happening. Germany was only mentioned as an example of how a "West German polling firm" proved extremely accurate in forecasting the outcome and therefore this topic is not relevant.
6) This document with the headline "Endpapers: international pen goes to Canada" is not relevant to this topic. The document does not focus on the lack of integration in a significant way in Germany but rather talks about a large geographically diverse international meeting of writers and goes in details. Therefore this document is not relevant.
7) This document with the headline "Motorcyclists draw stares along the silk road" is not relevant to the language and cultural differences integration with foreigners in Germany. This document talks about how a band of motorcyclists sped past a guy in this "horse-drawn cart" on silk road. There was a mention "110 people in a motorcycle caravan flying foreign flags" which does not contribute any value to how the lack of integration in Germany with immigration and therefore is not relevant.
8) This document with the headline "Europeans have much to lose in the gulf puzzle; France, which has pinned its Mideast policy on Baghdad for two decades, probably has the most at stake in the region" is not relevant to how languages and cultural differences impact integration of foreigners in Germany. This document talk about how

“the United States carries the biggest stick in the show down” again the “Iraqi leader Saddam Hussein”. Therefore this document is not relevant.

9) This document with the headline “Independence declared by Latvian Parliament” is not relevant to the immigration integration difficulties of immigrants in Germany. This document goes in detail of how Latvia’s parliament is going toward independence. Therefore, this document is not relevant.
10) This document with the headline “No melting pot; Europe busy closing door to foreigners” is opinionated whether it is relevant or not. However, it’s more on the un-relevant side. This document talks mainly about immigration problems as a whole and its core detail is not about Germany. They merely have a short example of how “200,000 East Europeans arriving to resettle in West Germany” but does not express details about integration with cultural difference and therefore this document is not relevant.

Out of these 10 documents, only one showed relevancy and that was “4)”. Therefore the precision is 1/10 = 0.1.

Topic 403:

The first ten documents were:

1) 403 q0 LA033089-0013 1 44 r255zhanAND
2) 403 q0 LA082890-0074 2 43 r255zhanAND
3) 403 q0 LA111589-0004 3 42 r255zhanAND
4) 403 q0 LA051490-0120 4 41 r255zhanAND
5) 403 q0 LA033089-0019 5 40 r255zhanAND
6) 403 q0 LA042189-0027 6 39 r255zhanAND
7) 403 q0 LA110490-0091 7 38 r255zhanAND
8) 403 q0 LA111989-0048 8 37 r255zhanAND
9) 403 q0 LA101890-0267 9 36 r255zhanAND
10) 403 q0 LA050290-0050 10 35 r255zhanAND

1) The document with the headline “Nutritionally speaking: Researchers assail lack of calcium in Women’s diets” is relevant to this topic. This document discusses the “best insurance against osteoporosis” and what it is. Osteoporosis is the “gradual loss of bone mineral from the skeleton” and the document also promotes some dietary intakes by mentioning that “drinking milk is not an option” because it will significantly meet the daily requirement and elaborates more on the importance of “dark leafy vegetables”. Therefore this document is relevant.
2) This document with the headline “Studying Side Effects of Hormone Therapy” is not relevant to the topic. Although the document does mention how hormone replacement therapy “is a drug treatment… [that] can reduce a woman’s risk of osteoporosis,” the document does not mention any dietary recommendations nor nutrition/mineral metabolism to aid the decrease in bone mass. Therefore this document is not relevant.
3) This document with the headline “UCSD to study the tie between hormones, heart disease” is not relevant to this topic because they do not elaborate on osteoporosis but merely to use the term as an example to explain other concepts. The document is mainly focused on “if commonly used postmenopausal hormones therapies can reduce woman’s risk of heart disease” and therefore is not relevant.
4) This document with the headline “People’s pharmacy: fluoride held unlikely to increase breast cancer risk” is not relevant for this topic. The document mentions osteoporosis to have an “experimental treatment” that “looks very promising”. However, the document’s core focus is not around that but the concern if fluoride causes cancer. Therefore this document is not relevant.
5) This document with the headline “Disease of dietary excess; Americans eating themselves to death is not relevant to this topic. The document does go into details with dietary and the importance of nutrition but does not go in depth with osteoporosis and suggest ways to decrease the likelihood of depreciating bone mass. Therefore this document is not relevant.
6) This document with the headline “Clotheslines: Saint lauren puts a new slant on slits” is not relevant to the topic. This document is about clothing and mentioning osteoporosis is a question to what type of clothing to the people who have that condition should wear. Therefore, this document is not relevant.

7) This documents with the headline “growth hormone reverses signs of age” is not relevant to the discussion of dietary intakes to aid the prevention of osteoporosis. The main topic of this document is to discuss the about growth hormones and how it impacts signs of age. Therefore this document is not relevant.
8) This document with the headline “How healthy was primitive man” is not relevant to the topic. This document discusses men’s health in general and does not discuss prevention of osteoporosis. Therefore this document is not relevant.
9) This document with the headline “People’s Pharmacy: careful consultation needed in planning one’s own death” is not relevant to this topic. The mere mention of osteoporosis was included in a question of how a user is worried about that condition but “don’t want to take estrogen hormones for the fear of breast cancer”. The document is mainly about how a user should plan his or her own death. Therefore this document is not relevant.
10) This document with the headline “Health care for the aged: invest now or pay later” is about how aging of America will “escalate health-care costs” and is not relevant to the topic of the prevention in osteoporosis. The document mentions that osteoporosis is a factor of why the increase cost in health care every year but does not have discussion with the disturbance of nutrition that may cause this condition. Therefore this document is not relevant.

The precision of these 10 documents would be 1/10 = 0.1. The first document “1)” was relevant.