

## **MSCI 541 Homework 1**

Thomas Kleinknecht

### **Question 1**

It was explained in lecture that with an online retrieval system, changes to the retrieval algorithm could be tested on real users as they do their searches. This paper:

Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu. 2012. Trustworthy online controlled experiments: five puzzling outcomes explained. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). ACM, New York, NY, USA, 786-794

<http://dx.doi.org/10.1145/2339530.2339653>

discusses "carryover effects" in section 3.5. In Figure 9 of the paper, the authors show a bad carryover effect. Knowing that such outcomes are possible with online experiments, if you were an engineer working on a new search engine ranking algorithm, what might you do to minimize the chance that they occur?

In section 3.5.4 of this paper they discuss two methods for mitigation of the impact of the carryover effect. You could use local randomization, which in practice would be to re-randomized a few of the bucket between each other each time in order to normalize the effect their respective previous experiment had on the users. The only thing about this approach is that you don't have the same continuity between groups so the benefit of consistent groups given by the bucket system is somewhat lost.

Another strategy may be to pair these studies with lab studies which can be carefully controlled to isolate how users change their behaviour when switching from platform A to B or starting with B and moving to A. This may allow you to carefully measure the carryover effect which is occurring so that you are more capable of spotting it in your bucket experiment results and using the known effects to adjust these metrics.

### **Question 2**

A) The precision of the top 5 results for ranker A is %, or 40%.

B) The precision of the top 5 results for ranker A is %, or 40%.

C) The use of binary relevance, by its definition, has no way of measuring how relevant or useful a given result is for a user, outside of the fact that it passes an arbitrary definition of relevance. Satisfying a user's information needs, as is shown by this example, is more than just showing any set of relevant results, it is about showing the most relevant results. Binary relevance does not allow us to measure and improve on this criteria, hiding the true quality difference of these rankings.

Set based retrieval measures of quality encourage less or more results to be given, even if this may not be what provides the user with the best opportunity to get the information they need. If the search engine only returns one result every time which it is sure is relevant, its precision measure will be 100%. Such is the case with the recall measure when the search engine just returns everything. Its use, specifically in this example, where we are taking a set of first 5 results to measure precision, leaves us with no way of measuring whether or not giving 20 results, but with a precision of 30% instead of 40%, would give the user more information which is important to them. This is because set based retrieval gives no value to the order of the results as it just views the set as unordered. Just as with binary relevance,

set-based retrieval measures offer us no way to distinguish between these two results, as beyond these 5 returned results we have no measure of the relevance of the results which we do not measure.

D) An example of the issue of binary relevance would be if you are a user searching for "University Campus Map", however you are a University of Waterloo student on campus for the first time in September of your first year. Any map for a the campus of a university in the world would be relevant to this search, as it would be a campus map as was searched. Both the campus map for U of T and University of Waterloo would be deemed as relevant. To the user however, the Waterloo map is much more useful.

As was briefly mentioned in my answer to part C, an example of the shortcomings of a set-based measure for information retrieval quality is in the case of the recall measure. This measures total number of relevant results found with respect to the total number of relevant results in the collection. An unordered list of every entry in the collection would most likely provide absolutely no value to the user, but this would, by definition, give a recall measure of 100%.

### **Question 3**

I have not included my response to this question as Professor Smucker has posted to Campuswire saying that this would not be marked.

### **Question 4**

#### Approach

In IndexEngine, my approach was to unzip the file one line at a time as suggested, iterating through and saving each raw document one at a time then extracting the metadata information. After this is complete, the program exports the docno to the DOCNOs.txt file which is created upon execution of the program in the latimes-index directory. This will be used when the GetDoc program is run. Next, the program checks to see whether the YY, MM and DD folders for this doc have been created and creates them if not (a YY/MM/DD structure is used for storing documents). The metadata and raw document are then stored separately in their respective folders under the document's date for ease of future access. The metadata folder consists of four lines of text, each containing solely the value of that field so that this file will not need to be parsed in the future. They are ordered as they are to be displayed in GetDoc, docno, internal id, date, then headline.

For the GetDoc program, it is relatively simple due to the way things were set up in IndexEngine. Firstly, the DOCNOs.txt file is read in line by line, and an ArrayList is built to access the docno for a given internal id which serves as the index, as well as a hashmap in order to access the internal id given the docno as a key value. With these mappings created, the date can be extracted from the docno, allowing easy traversal of the document directory, where the raw document can be extracted and stored as a string, as well as the docno, internal id, date and headline, by extracting the first and only 4 lines of the metadata file. These values are then printed to the console with their appropriate titles, and the program is complete.

#### Programming Language and Instructions

The programs were written in Java, specifically java version "13.0.1". In order to build and run the program, the program will need to be cloned from the repository, and can be run in the command line if you have java downloaded on your computer (or after downloading).

First, ensure you have the proper directories and files set up. You will need the gzipped LA Times portion of the TREC volumes 4 and 5 collection downloaded, and its file path. Also, you will need to create a “latimes-index” directory and note the file path.

Next, IndexEngine must be run before GetDoc. In order to run IndexEngine, navigate to the cloned repository, and compile the code using the command:

```
javac IndexEngine.java
```

Now you can run IndexEngine with following command:

```
java IndexEngine "/path/to/latimes.gz" "path/to/latimes-index"
```

These two arguments should be enclosed in quotations, and contain the file path of your gzipped latimes file and your latimes-index directory.

This program may take a few minutes to run, so please be patient. When this is complete, you can now run the GetDoc program to retrieve the metadata and raw document for any document given its docno or internal id.

To do so, compile the GetDoc program with the following command:

```
javac GetDoc.java
```

To run the GetDoc program, use either of the following:

```
java GetDoc "path/to/latimes-index" id 7
```

Or

```
java GetDoc "path/to/latimes-index" docno LA010189-001
```

With the path to your latimes-index directory, either “id” or “docno” to represent if you are searching by internal id or docno, and then the respective internal id or docno. This will return the metadata and raw document for this article and print this result to the terminal.

### Testing/Program Results:

#### **IndexEngine:**

##### Command Line Verification:

If less than two arguments are provided:

```
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java IndexEngine.java
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java IndexEngine "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes.gz"
Please provide a path to the latimes.gz file and a path to a directory where the documents and metadata will be stored as arguments to the IndexEngine program. Ensure each path enclosed in quotations.
```

If latimes.gz file path does not exist or latimes-index file path does not exist:

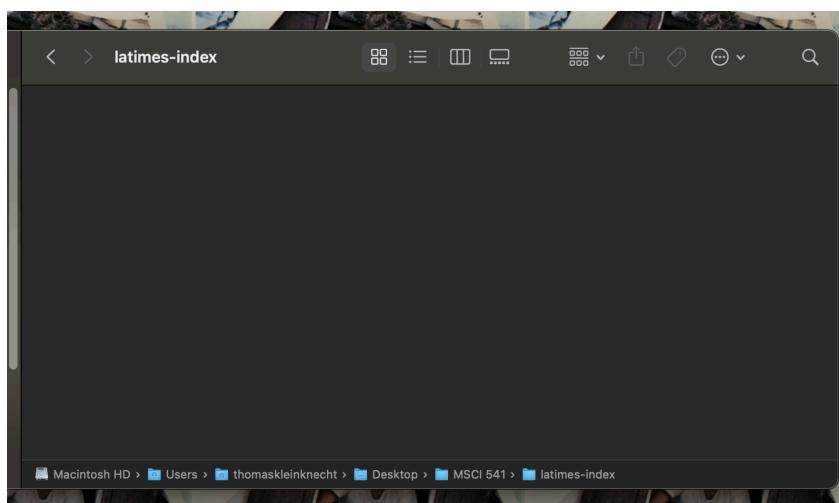
```
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java IndexEngine "/Users/thomaskleinknecht/Desktop/MSCI 542/latimes.gz" "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes-index"
Please provide the proper path to the latimes.gz file. This directory does not exist.
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java IndexEngine "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes.gz" "/Users/thomaskleinknecht/Desktop/MSCI 51/latimes-index"
Please provide the proper path to the latimes-index file. This directory does not exist.
```

If directory already exists/has already been created:



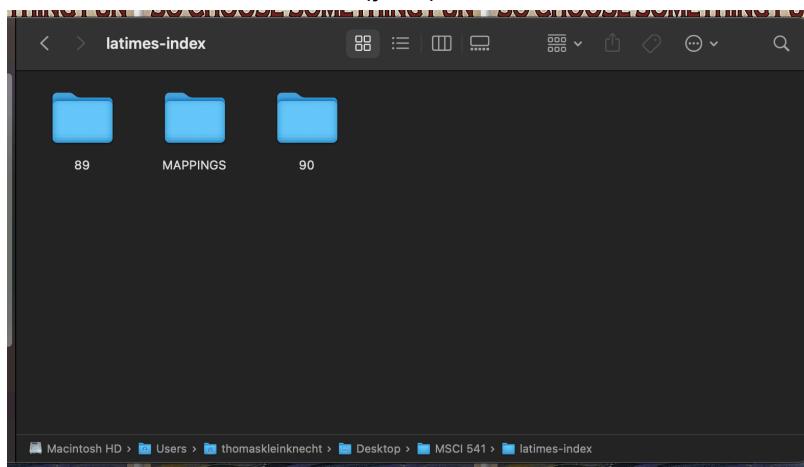
### General Functionality:

Before:

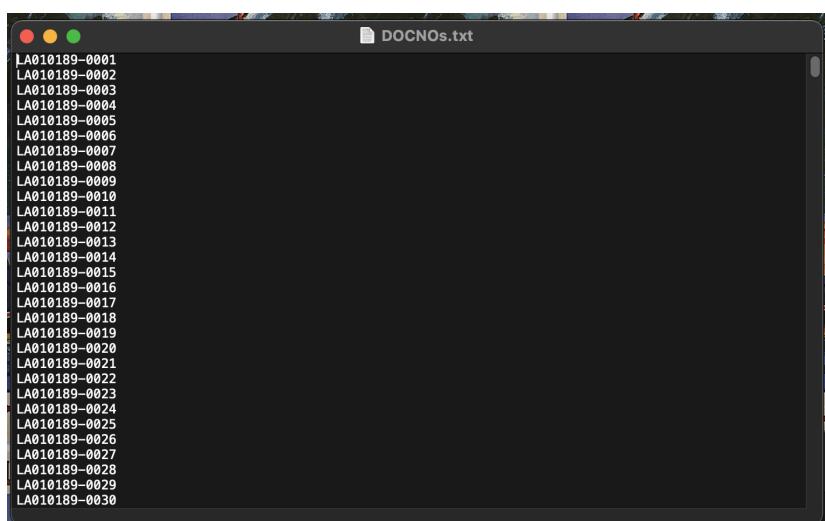
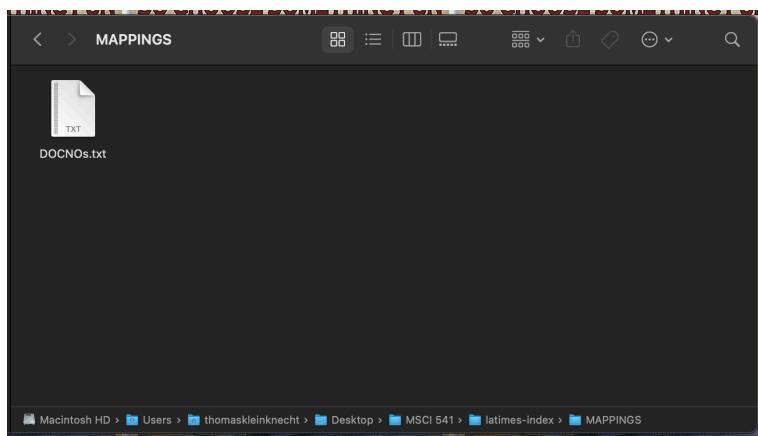


After:

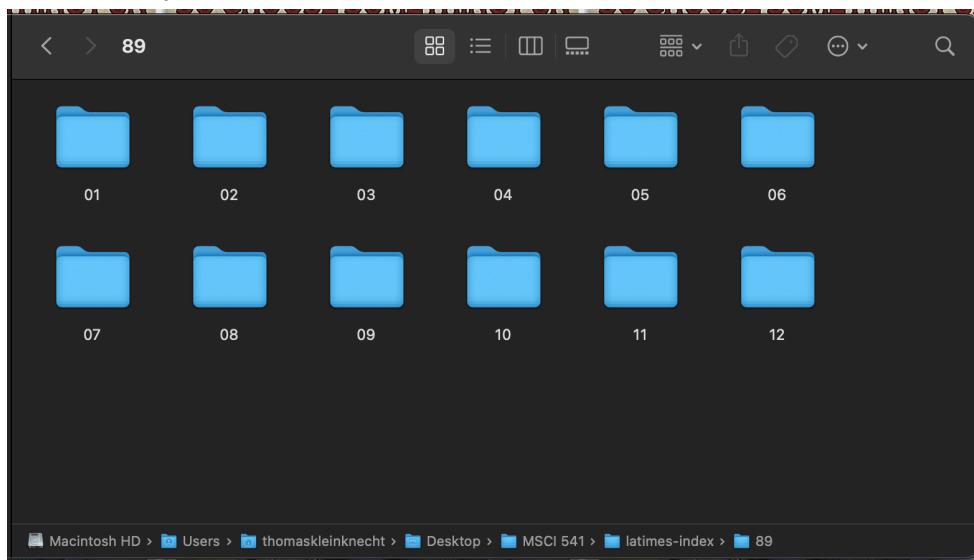
MAPPINGS Folder and YY (year) folders created:



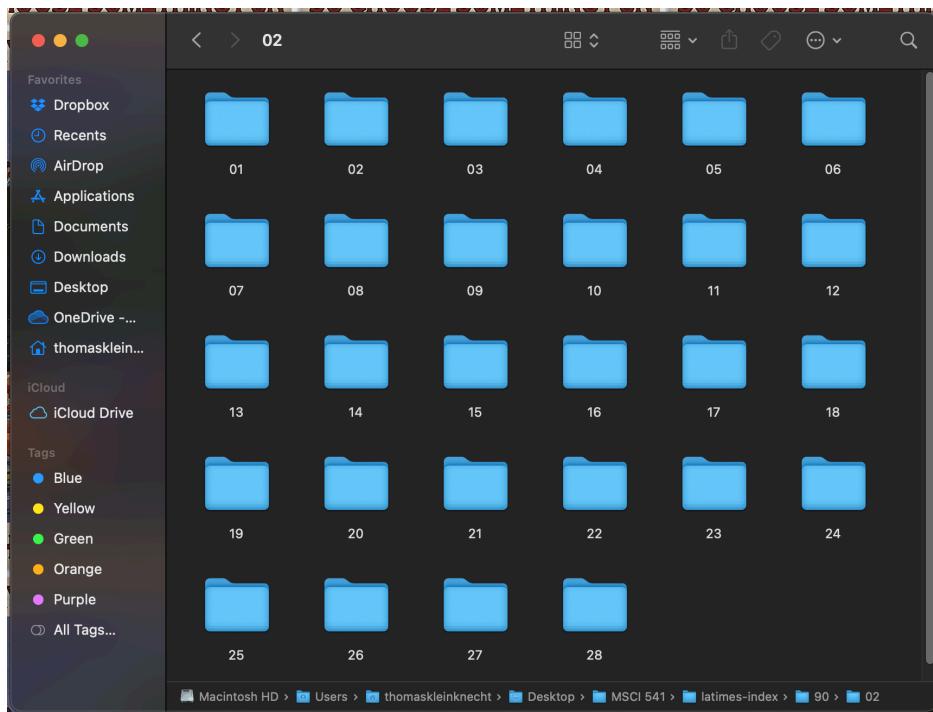
In MAPPINGS, contains DOCNOs.txt file with a list of all docnos, the line which they are on - 1 represents their internal ID:



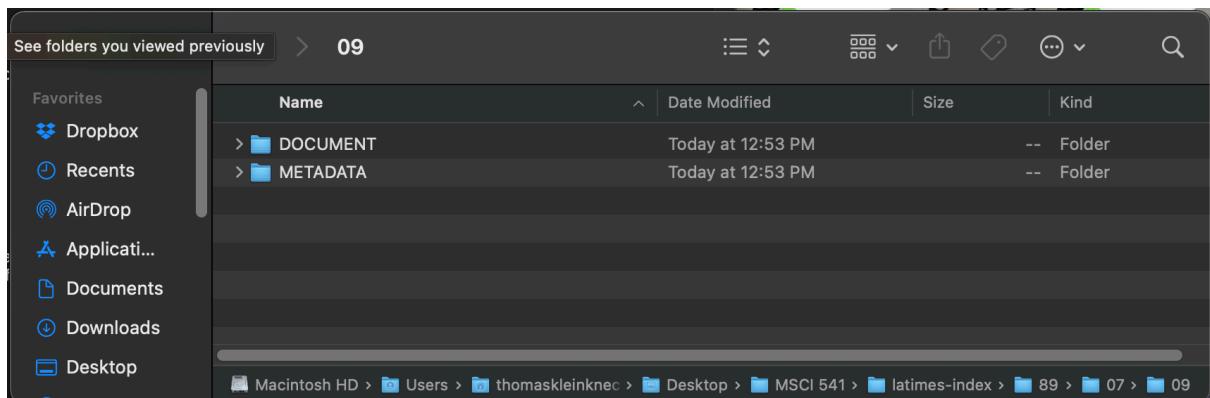
Under each year, MM (month) folders created for each month:



Under each month, DD folders created for each day:



Under each day, and METADATA and DOCUMENT folder is created, housing each documents metadata and raw document that falls on that day. All files are named by DOCNO.txt:



Name	Date Modified	Size	Kind
LA070989-0001.txt	Today at 12:53 PM	7 KB	Plain Text
LA070989-0002.txt	Today at 12:53 PM	5 KB	Plain Text
LA070989-0003.txt	Today at 12:53 PM	8 KB	Plain Text
LA070989-0004.txt	Today at 12:53 PM	5 KB	Plain Text
LA070989-0005.txt	Today at 12:53 PM	14 KB	Plain Text
LA070989-0006.txt	Today at 12:53 PM	9 KB	Plain Text
LA070989-0007.txt	Today at 12:53 PM	5 KB	Plain Text
LA070989-0008.txt	Today at 12:53 PM	785 bytes	Plain Text
LA070989-0009.txt	Today at 12:53 PM	506 bytes	Plain Text
LA070989-0010.txt	Today at 12:53 PM	2 KB	Plain Text
LA070989-0011.txt	Today at 12:53 PM	23 KB	Plain Text
LA070989-0012.txt	Today at 12:53 PM	7 KB	Plain Text
LA070989-0013.txt	Today at 12:53 PM	6 KB	Plain Text
LA070989-0014.txt	Today at 12:53 PM	8 KB	Plain Text
LA070989-0015.txt	Today at 12:53 PM	5 KB	Plain Text
LA070989-0016.txt	Today at 12:53 PM	4 KB	Plain Text
LA070989-0017.txt	Today at 12:53 PM	6 KB	Plain Text
LA070989-0018.txt	Today at 12:53 PM	4 KB	Plain Text
LA070989-0019.txt	Today at 12:53 PM	8 KB	Plain Text
LA070989-0020.txt	Today at 12:53 PM	10 KB	Plain Text
LA070989-0021.txt	Today at 12:53 PM	2 KB	Plain Text
LA070989-0022.txt	Today at 12:53 PM	1 KB	Plain Text
LA070989-0023.txt	Today at 12:53 PM	2 KB	Plain Text
LA070989-0024.txt	Today at 12:53 PM	741 bytes	Plain Text
LA070989-0025.txt	Today at 12:53 PM	1 KB	Plain Text
LA070989-0026.txt	Today at 12:53 PM	1 KB	Plain Text
LA070989-0027.txt	Today at 12:53 PM	1 KB	Plain Text
LA070989-0028.txt	Today at 12:53 PM	899 bytes	Plain Text
LA070989-0029.txt	Today at 12:53 PM	5 KB	Plain Text
LA070989-0030.txt	Today at 12:53 PM	9 KB	Plain Text
LA070989-0031.txt	Today at 12:53 PM	1 KB	Plain Text
LA070989-0032.txt	Today at 12:53 PM	3 KB	Plain Text

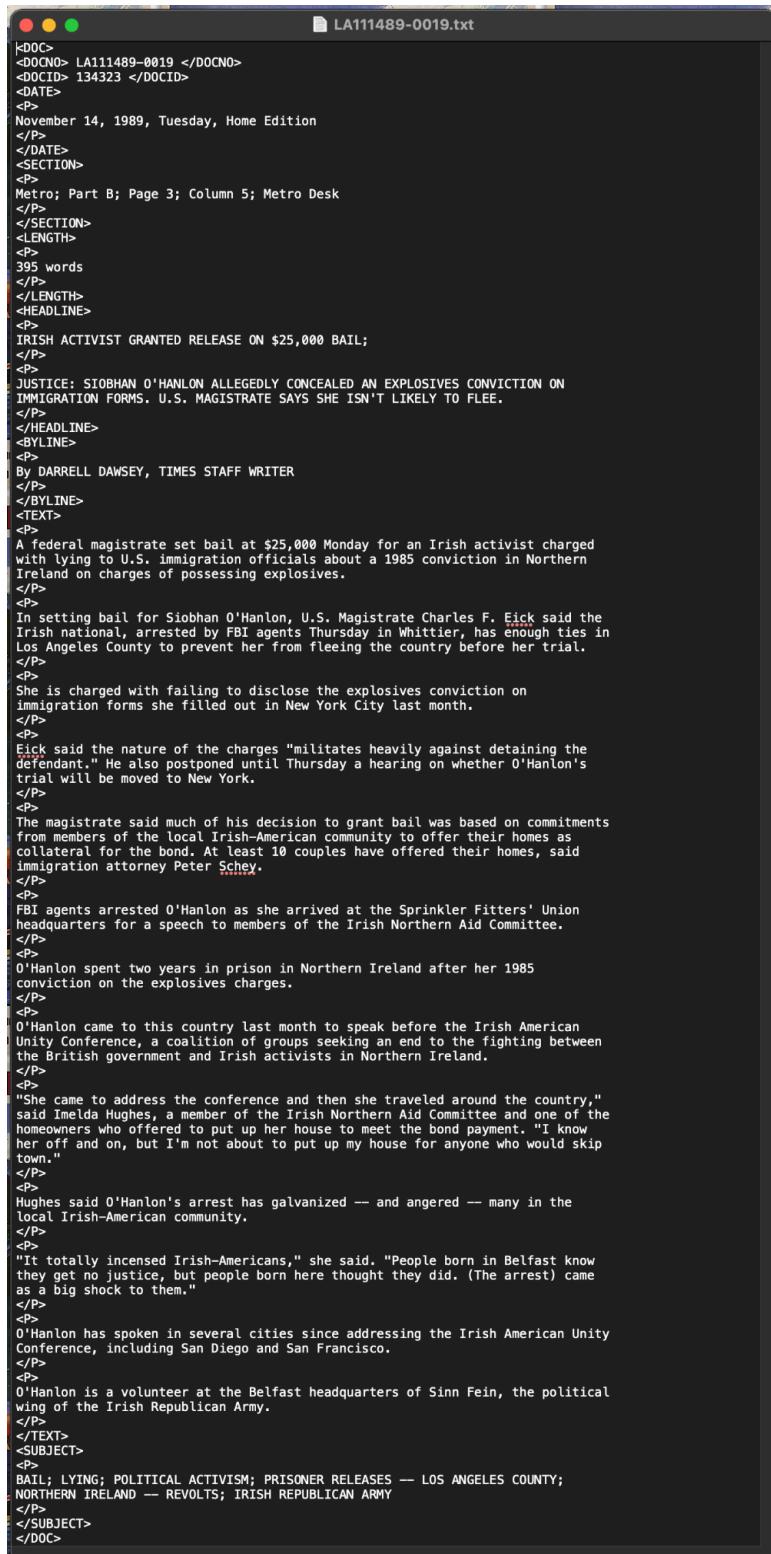
Macintosh HD > Users > thomas > Desktop > MSCI 5 > latimes > 89 > 07 > 09 > DOCUMENT

Name	Date Modified	Size	Kind
LA070989-0001.txt	Today at 12:53 PM	132 bytes	Plain Text
LA070989-0002.txt	Today at 12:53 PM	51 bytes	Plain Text
LA070989-0003.txt	Today at 12:53 PM	123 bytes	Plain Text
LA070989-0004.txt	Today at 12:53 PM	144 bytes	Plain Text
LA070989-0005.txt	Today at 12:53 PM	149 bytes	Plain Text
LA070989-0006.txt	Today at 12:53 PM	146 bytes	Plain Text
LA070989-0007.txt	Today at 12:53 PM	48 bytes	Plain Text
LA070989-0008.txt	Today at 12:53 PM	33 bytes	Plain Text
LA070989-0009.txt	Today at 12:53 PM	33 bytes	Plain Text
LA070989-0010.txt	Today at 12:53 PM	60 bytes	Plain Text
LA070989-0011.txt	Today at 12:53 PM	165 bytes	Plain Text
LA070989-0012.txt	Today at 12:53 PM	116 bytes	Plain Text
LA070989-0013.txt	Today at 12:53 PM	194 bytes	Plain Text
LA070989-0014.txt	Today at 12:53 PM	136 bytes	Plain Text
LA070989-0015.txt	Today at 12:53 PM	152 bytes	Plain Text
LA070989-0016.txt	Today at 12:53 PM	124 bytes	Plain Text
LA070989-0017.txt	Today at 12:53 PM	151 bytes	Plain Text
LA070989-0018.txt	Today at 12:53 PM	160 bytes	Plain Text
LA070989-0019.txt	Today at 12:53 PM	130 bytes	Plain Text
LA070989-0020.txt	Today at 12:53 PM	167 bytes	Plain Text
LA070989-0021.txt	Today at 12:53 PM	128 bytes	Plain Text
LA070989-0022.txt	Today at 12:53 PM	54 bytes	Plain Text
LA070989-0023.txt	Today at 12:53 PM	65 bytes	Plain Text
LA070989-0024.txt	Today at 12:53 PM	62 bytes	Plain Text
LA070989-0025.txt	Today at 12:53 PM	54 bytes	Plain Text
LA070989-0026.txt	Today at 12:53 PM	54 bytes	Plain Text
LA070989-0027.txt	Today at 12:53 PM	54 bytes	Plain Text
LA070989-0028.txt	Today at 12:53 PM	50 bytes	Plain Text
LA070989-0029.txt	Today at 12:53 PM	90 bytes	Plain Text
LA070989-0030.txt	Today at 12:53 PM	163 bytes	Plain Text
LA070989-0031.txt	Today at 12:53 PM	210 bytes	Plain Text
LA070989-0032.txt	Today at 12:53 PM	328 bytes	Plain Text

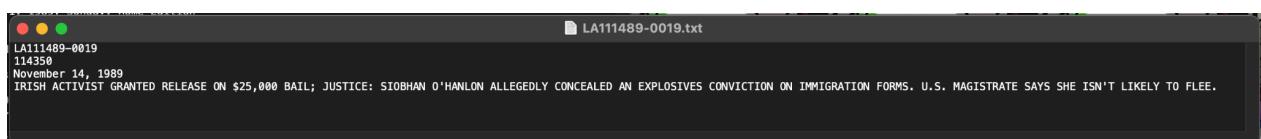
Macintosh HD > Users > thomas > Desktop > MSCI 5 > latimes > 89 > 07 > 09 > METADATA

## Example metadata and respective raw doc files:

1. LA111489.txt (notice the headline is appended to one line and the tags are all removed)



```
<DOC>
<DOCNO> LA111489-0019 </DOCNO>
<DOCID> 134323 </DOCID>
<DATE>
<P>
November 14, 1989, Tuesday, Home Edition
</P>
</DATE>
<SECTION>
<P>
Metro; Part B; Page 3; Column 5; Metro Desk
</P>
</SECTION>
<LENGTH>
<P>
395 words
</P>
</LENGTH>
<HEADLINE>
<P>
IRISH ACTIVIST GRANTED RELEASE ON $25,000 BAIL;
</P>
<P>
JUSTICE: SIOBHAN O'HANLON ALLEGEDLY CONCEALED AN EXPLOSIVES CONVICTION ON
IMMIGRATION FORMS. U.S. MAGISTRATE SAYS SHE ISN'T LIKELY TO FLEE.
</P>
</HEADLINE>
<BYLINE>
<P>
By DARRELL DAWSEY, TIMES STAFF WRITER
</P>
</BYLINE>
<TEXT>
<P>
A federal magistrate set bail at $25,000 Monday for an Irish activist charged
with lying to U.S. immigration officials about a 1985 conviction in Northern
Ireland on charges of possessing explosives.
</P>
<P>
In setting bail for Siobhan O'Hanlon, U.S. Magistrate Charles F. Eick said the
Irish national, arrested by FBI agents Thursday in Whittier, has enough ties in
Los Angeles County to prevent her from fleeing the country before her trial.
</P>
<P>
She is charged with failing to disclose the explosives conviction on
immigration forms she filled out in New York City last month.
</P>
<P>
Eick said the nature of the charges "militates heavily against detaining the
defendant." He also postponed until Thursday a hearing on whether O'Hanlon's
trial will be moved to New York.
</P>
<P>
The magistrate said much of his decision to grant bail was based on commitments
from members of the local Irish-American community to offer their homes as
collateral for the bond. At least 10 couples have offered their homes, said
immigration attorney Peter Schey.
</P>
<P>
FBI agents arrested O'Hanlon as she arrived at the Sprinkler Fitters' Union
headquarters for a speech to members of the Irish Northern Aid Committee.
</P>
<P>
O'Hanlon spent two years in prison in Northern Ireland after her 1985
conviction on the explosives charges.
</P>
<P>
O'Hanlon came to this country last month to speak before the Irish American
Unity Conference, a coalition of groups seeking an end to the fighting between
the British government and Irish activists in Northern Ireland.
</P>
<P>
"She came to address the conference and then she traveled around the country," said
Imelda Hughes, a member of the Irish Northern Aid Committee and one of the
homeowners who offered to put up her house to meet the bond payment. "I know
her off and on, but I'm not about to put up my house for anyone who would skip
town."
</P>
<P>
Hughes said O'Hanlon's arrest has galvanized — and angered — many in the
local Irish-American community.
</P>
<P>
"It totally incensed Irish-Americans," she said. "People born in Belfast know
they get no justice, but people born here thought they did. (The arrest) came
as a big shock to them."
</P>
<P>
O'Hanlon has spoken in several cities since addressing the Irish American Unity
Conference, including San Diego and San Francisco.
</P>
<P>
O'Hanlon is a volunteer at the Belfast headquarters of Sinn Fein, the political
wing of the Irish Republican Army.
</P>
</TEXT>
<SUBJECT>
<P>
BAIL; LYING; POLITICAL ACTIVISM; PRISONER RELEASES -- LOS ANGELES COUNTY;
NORTHERN IRELAND -- REVOLTS; IRISH REPUBLICAN ARMY
</P>
</SUBJECT>
</DOC>
```



```
LA111489-0019
114350
November 14, 1989
IRISH ACTIVIST GRANTED RELEASE ON $25,000 BAIL; JUSTICE: SIOBHAN O'HANLON ALLEGEDLY CONCEALED AN EXPLOSIVES CONVICTION ON IMMIGRATION FORMS. U.S. MAGISTRATE SAYS SHE ISN'T LIKELY TO FLEE.
```

## 2. LA090990.txt

```
LA090990-0007.txt

<DOC>
<DOCNO> LA090990-0007 </DOCNO>
<DOCID> 273321 </DOCID>
<DATE>
<P>
September 9, 1990, Sunday, Home Edition
</P>
</DATE>
<SECTION>
<P>
Magazine; Page 4; Magazine Desk
</P>
</SECTION>
<LENGTH>
<P>
41 words
</P>
</LENGTH>
<HEADLINE>
<P>
BROTHERS MENENDEZ
</P>
</HEADLINE>
<TEXT>
<P>
To teach competitiveness without character, conscience and compassion is to
invite tragedy. To compound this, by teaching that the pursuit of valuables is
more important than the pursuit of values is to assure tragedy.
</P>
<P>
PHOEBE LAMBETH
</P>
<P>
Orange
</P>
</TEXT>
<TYPE>
<P>
Letter to the Editor
</P>
</TYPE>
</DOC>
```

```
LA090990-0007.txt

LA090990-0007
89639
September 9, 1990
BROTHERS MENENDEZ
```

3. Document with no headline. Headline is saved as an empty string:

The screenshot shows a Mac OS X desktop with two windows for the file "LA010189-0007.txt".

The top window is a text editor displaying the following XML code:

```
<DOC>
<DOCNO> LA010189-0007 </DOCNO>
<DOCID> 13 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<SECTION>
<P>
TV Times; Page 5; Television Desk
</P>
</SECTION>
<LENGTH>
<P>
24 words
</P>
</LENGTH>
<TEXT>
<P>
May I add yet another hurrah for the "Live With Regis and Kathie Lee" show? I
find it stimulating and refreshing.
</P>
<P>
J. K. Matthews, Pomona
</P>
</TEXT>
<TYPE>
<P>
Letter to the Editor
</P>
</TYPE>
</DOC>
```

The bottom window is a smaller text viewer showing the extracted text:

```
LA010189-0007
6
January 1, 1989
```

## GetDoc:

### Command Line Verification:

1. Three arguments are not provided
2. Path to latimes-index does not exist
3. Second argument does not specifically contain “docno” or “id”
4. Internal id provided is not valid and does not exist
5. DOCNO provided is not valid and does not exist

```
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinnecht$ java GetDoc
Please provide a path to the latimes-index directory, either docno or id, then the respective docno or id. Ensure path is enclosed in quotes.
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinnecht$ java GetDoc "Users/thomaskleinnecht/Desktop/MSCI 541/latimes-index" int 1
Please provide the proper path to the latimes-index directory. This directory does not exist.
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinnecht$ java GetDoc "/Users/thomaskleinnecht/Desktop/MSCI 541/latimes-index" int 1
Please specify either docno or id in the second argument.
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinnecht$ java GetDoc "/Users/thomaskleinnecht/Desktop/MSCI 541/latimes-index" id 100
000000
The internal id provided is invalid. Please enter a valid internal id.
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinnecht$ java GetDoc "/Users/thomaskleinnecht/Desktop/MSCI 541/latimes-index" docno
hgdksjhfgdskjfng
The docno provided is invalid. Please enter a valid docno.
```

### General Functionality:

Using internal id to retrieve a document:

```

Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java GetDoc "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes-index" id 77543
docno: LA080590-0097
internal id: 77543
date: August 5, 1990
headline: MORE ROSEANNE
raw document:
<DOC>
<DOCNO> LA080590-0097 </DOCNO>
<DOCID> 257374 </DOCID>
<DATE>
<P>
August 5, 1990, Sunday, San Diego County Edition
</P>
</DATE>
<SECTION>
<P>
Metro; Part B; Page 2; Column 1; Metro Desk
</P>
</SECTION>
<LENGTH>
<P>
127 words
</P>
</LENGTH>
<HEADLINE>
<P>
MORE ROSEANNE
</P>
</HEADLINE>
<TEXT>
<P>
With respect to Roseanne Barr's national anthem escapade, I am truly
embarrassed by the people who booed her.
</P>
<P>
However, embarrassment turned to confusion when President Bush publicly
concluded that her singing was a disgrace.
</P>
<P>
Isn't it a greater disgrace for Bush's son to be involved in the billion-dollar
Silverado Banking, Savings & Loan Assn. scandal? To me, that is a real
profound disgrace.
</P>
<P>
At least Roseanne knew the words to the "Star-Spangled Banner." Do we have such
short memories that we can't recall Pia Zadora's miserable attempt to sing the
national anthem over network TV? She forgot the words, or didn't know them.
</P>
<P>
My hat is off to you, Rosie. You sang your best and I, for one, appreciate your
effort.
</P>
<P>
CYNTHIA McMILLAN
</P>
<P>
San Diego
</P>
</TEXT>
<TYPE>
<P>
Letter to the Editor
</P>
</TYPE>
</DOC>
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ █

```

## Using DOCNO to retrieve a document:

```

Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java GetDoc "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes-index" docno LA110689-0021
docno: LA110689-0021
internal id: 111331
date: November 6, 1989
headline: QUOTABLE
raw document:
<DOC>
<DOCNO> LA110689-0021 </DOCNO>
<DOCID> 130773 </DOCID>
<DATE>
<P>
November 6, 1989, Monday, San Diego County Edition
</P>
</DATE>
<SECTION>
<P>
Calendar; Part F; Page 1; Column 1; Entertainment Desk
</P>
</SECTION>
<LENGTH>
<P>
20 words
</P>
</LENGTH>
<HEADLINE>
<P>
QUOTABLE
</P>
</HEADLINE>
<TEXT>
<P>
"Actors are wicked creatures. They cannot do without cliches."
</P>
<P>
-- actress Lia Akhejakova, in Soviet Theatre magazine
</P>
</TEXT>
</DOC>
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ █

```

## Retrieving a document with no headline:

```
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ java GetDoc "/Users/thomaskleinknecht/Desktop/MSCI 541/latimes-index" id 7
docno: LA010189-0008
internal id: 7
date: January 1, 1989
headline:
raw document:
<DOC>
<DOCNO> LA010189-0008 </DOCNO>
<DOCID> 14 </DOCID>
<DATE>
<P>
January 1, 1989, Sunday, Home Edition
</P>
</DATE>
<SECTION>
<P>
TV Times; Page 5; Television Desk
</P>
</SECTION>
<LENGTH>
<P>
42 words
</P>
</LENGTH>
<TEXT>
<P>
Thanks to ABC and Marlo Thomas for sharing a holiday gift with us. "Free to Be . . . a Family" was a touching reminder that truly we are one human family and the planet must be preserved for our precious children.
</P>
<P>
Pamela Popovich, Redondo Beach
</P>
</TEXT>
<TYPE>
<P>
Letter to the Editor
</P>
</TYPE>
</DOC>
Thomas-MacBook:msci-541-f23-hw1-thomask902 thomaskleinknecht$ █
```

## Sources/Citations

1. Discussed question 4 with classmate Matthew Erxleben on high level strategy and requirements of the program
2. For structure/methods on how to unzip the file line by line in java I referenced <https://stackoverflow.com/questions/1080381/gzipinputstream-reading-line-by-line>
3. Most ideas for structure and overall strategy were from class as well as classmates questions on campuswire