
**11-442 / 11-642:
Search Engines**

Reading Summaries

Jamie Callan
Carnegie Mellon University
callan@cs.cmu.edu

Reading Summary Policy

Your reading summary must be original

- Original words
- Original organization of ideas

You may not use someone else's words or organization of ideas

- "Plagiarism", "Cheating", "Academic integrity violation"

Copy and Paste is Plagiarism

Textbook

A first take at building an inverted index

To gain the speed benefits of indexing at retrieval time, we have to build the index in advance. The major steps in this are:

1. Collect the documents to be indexed:

Friends, Romans, countrymen.

So let it be with Caesar

2. Tokenize the text, turning each document into a list of tokens:

Friends Romans countrymen So

3. Do linguistic preprocessing, producing a list of normalized tokens, which are the indexing terms:

friend roman countryman so

4. Index the documents that each term occurs in by creating an inverted index, consisting of a dictionary and postings

Reading summary

The major steps in this are:

1. Collect the documents to be indexed.
2. Tokenize the text, turning each document into a list of tokens.
3. Do linguistic preprocessing, producing a list of normalized tokens, which are the indexing terms.
4. Index the documents that each term occurs in by creating an inverted index, consisting of a dictionary and postings.

3

© 2019, Jamie Callan

Copy, Paste, and Revise is Also Plagiarism ("Revise" == "Hide, Deceive")

Textbook

A first take at building an inverted index

To gain the speed benefits of indexing at retrieval time, we have to build the index in advance. The major steps in this are:

1. Collect the documents to be indexed:

Friends, Romans, countrymen.

So let it be with Caesar

2. Tokenize the text, turning each document into a list of tokens:

Friends Romans countrymen So

3. Do linguistic preprocessing, producing a list of normalized tokens, which are the indexing terms:

friend roman countryman so

4. Index the documents that each term occurs in by creating an inverted index, consisting of a dictionary and postings

Reading Summary Version 1

The major steps in this are:

1. Collect the documents to be indexed.
2. Tokenize the text, turning each document into a list of tokens.
3. Do linguistic preprocessing, producing a list of normalized tokens, which are the indexing terms.
4. Index the documents that each term occurs in by creating an inverted index, consisting of a dictionary and postings.

4

Reading Summary Version 2

- The important steps in building an index are
1. Obtain the documents.
 2. Process the document text into tokens.
 3. Lexical analysis on the tokens to create index terms.
 4. Create an inverted index that contains a term dictionary and postings.

© 2019, Jamie Callan

Reading Summaries Need to Be Your Thoughts

Copying suggests that you do not understand the material

- If you understood it, you could just write it in your own words

Revising someone else's text just hides the source of the ideas

- It pretends that someone else's ideas are your ideas
- This makes it a worse offense, not better

5

© 2019, Jamie Callan

Reading Summaries Need to Be Your Thoughts

Show us what you know ...

- Not what you can look up in a book

Close the book or browser before you write the reading summary

Writing reading summaries from memory is good practice

- E.g., for exams and job interviews

Use the reading summaries to practice remembering

- “Low stakes” practice (i.e., only a small portion of the grade)

6

© 2019, Jamie Callan

Grading Scale

✓+ 110

Our way of saying ‘nice job!’

But ... a small bonus – not worth the extra effort

✓ 100

The best use of your time ... optimize for this score

✓ – 80

A small penalty – not a big deal

A warning that maybe you don’t understand the material
as well as you thought

7

© 2019, Jamie Callan

What To Submit (Submit on Canvas)

January 22, before class

- A reading summary for Week 1 (Jan 15-17) 1 page
- A reading summary for Week 2 (Jan 22-24) 1 page
- Two separate documents

January 29, before class

- A reading summary for Week 3 (Jan 29-31) 1 page

February 5, before class

- ...

Use Canvas to submit your reading summaries

8

© 2019, Jamie Callan