

Project 2

March 19, 2013

Next Few Weeks

Sun	Mon	Tues	Wed	Thurs	Fri	Sat
3/17	3/18	3/19	3/20	3/21	3/22	3/23
		Project 2 Out		HW2-4 Due HW2-5 Out		
3/24	3/25	3/26	3/27	3/28	3/29	3/30
Spring Break! Woouoooo!						
3/31	4/1	4/2	4/3	4/4	4/5	4/6
		HW2-5 Due		Proposal Due		
4/7	4/8	4/9	4/10	4/11	4/12	4/14
				Project 2 Due		

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- Example Project 2 Proposal

Project Gutenberg



search book catalog

- Book Search
- Catalog
- Bookshelves

search site

- Main Page
- Categories
- News
- Contact Info

donate

Project Gutenberg needs your donation!

Free eBooks by Project Gutenberg

From Project Gutenberg, the first producer of free ebooks.

[Mobile Site](#) · [Book search](#) · [Bookshelves by topic](#) · [Top downloads](#) · [Recently added](#) · [Report errors](#)

Some of Our Latest Books



Welcome

Project Gutenberg offers over 38,000 free ebooks: choose among free epub books. free kindle books. download them or read them online.

Project Gutenberg



Free eBooks by Project Gutenberg

From Project Gutenberg, the first producer of free ebooks.

search book catalog

1. Find a book. Any book.
2. How large is the Plain Text UTF-8 File?
 1. Mb = Megabyte
 2. Kb = Kilobyte
3. Find a book that is < 1Mb. Download it.

1024 Kb = 1Mb

donate

Project Gutenberg needs your donation!

Project Gutenberg offers over 38,000 free ebooks: choose among free epub books. free kindle books. download them or read them online.

Project Gutenberg



Free eBooks by Project Gutenberg

From Project Gutenberg, the first producer of free ebooks.

search book catalog

Look at the function
`removeLicenseFromProjectGutenberg`
in `DataImport.py`

search site

- Main Page
- Categories
- News
- Contact Info

donate

Project Gutenberg needs your donation!



Welcome

Project Gutenberg offers over 38,000 free ebooks: choose among free epub books. free kindle books. download them or read them online.

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- Example Project 2 Proposal

Webster's Unabridged Dictionary

The Online Plain Text English Dictionary

OPTED is a public domain English word list dictionary, based on the public domain portion of "The Project Gutenberg Etext of Webster's Unabridged Dictionary" which is in turn based on the 1913 US Webster's Unabridged Dictionary. (See [Project Gutenberg](#))

This version has been extensively stripped down and set out as one definition per line. All the Gutenberg EText tags and formatting have been removed by computer. Version 0.03 is a new processing of v0.47 of the websters dictionary and it has considerably fewer errors. Also the definition limit of 255 chars has been removed to give full justice of some of the more majestic of the originals. Some important errors in the parts-of-speech fields have been corrected and a lot of inflections/ alternatives and plurals that were missed due to software bugs in v0.01 and 0.02 are now included properly.

The dictionary is set as a word list with definitions, using minimal HTML markup. The only tags used are <P>, and <I> and these serve to delimit the words (between s) the part of speech or type (between <I>s) and the definitions (The rest of the line). Each entry is between a <P>, </P> pair. This will facilitate computer processing. The text was prepared on a macintosh, so the few accented and unlauded characters appear best if your browser is set to Western MacRoman encoding (this should look like an unlauded u : ü). If this causes problems and I get enough responses, I'll look into producing an ISO 8859-1 or even a Unicode version.

The dictionary can be viewed (with patience) directly online as you would a normal printed dictionary, otherwise a user can download the pages and process them in some way on their own machine. The only usage conditions are that if the material is redistributed, the content (not the formatting) remain in the public domain (ie free) and that the content be easily accessible in non-encoded plain text format at no cost to the end user. The origin of the content should also be acknowledged, including OPTED, Project Gutenberg and the 1913 edition of Webster's Unabridged Dictionary. If the material is to be included in commercial products, Project Gutenberg should be contacted first. There are no restrictions for personal or research uses of this material.

OPTED v0.03 by Letter(size)

Second computer generated version:

[A\(1.1M\)](#) | [B\(1005k\)](#) | [C\(1.6M\)](#) | [D\(1M\)](#) | [E\(809k\)](#) | [F\(784k\)](#) | [G\(564k\)](#) | [H\(686k\)](#) | [I\(833k\)](#) | [J\(172k\)](#) | [K\(172k\)](#) | [L\(637k\)](#) | [M\(931k\)](#) | [N\(343k\)](#) | [O\(466k\)](#) | [P\(1.5M\)](#) | [Q\(147k\)](#) | [R\(931k\)](#) | [S\(2.1M\)](#) | [T\(1005k\)](#) | [U\(343k\)](#) | [V\(343k\)](#) | [W\(490k\)](#) | [X\(49k\)](#) | [Y\(74k\)](#) | [Z\(74k\)](#)

Webster's Unabridged Dictionary

The Online Plain Text English Dictionary

OPTED is a public domain English word list dictionary, based on the public domain portion of "The Project Gutenberg Etext of Webster's Unabridged Dictionary" which is in turn based on the 1913 US Webster's Unabridged Dictionary. (See Project Gutenberg)

1. According to the homepage, what does each line contain?
2. What letter is the **smallest** file?
 1. Mb = Megabyte
 2. Kb = Kilobyte
3. Click on it. Right-click and select View Page Source...

1024 Kb = 1Mb

Second computer generated version.

[A\(1.1M\)](#) | [B\(1005k\)](#) | [C\(1.6M\)](#) | [D\(1M\)](#) | [E\(809k\)](#) | [F\(784k\)](#) | [G\(564k\)](#) | [H\(686k\)](#) | [I\(833k\)](#) | [J\(172k\)](#) | [K\(172k\)](#) | [L\(637k\)](#) | [M\(931k\)](#) | [N\(343k\)](#) | [O\(466k\)](#) | [P\(1.5M\)](#) | [Q\(147k\)](#) | [R\(931k\)](#) | [S\(2.1M\)](#) | [T\(1005k\)](#) | [U\(343k\)](#) | [V\(343k\)](#) | [W\(490k\)](#) | [X\(49k\)](#) | [Y\(74k\)](#) | [Z\(74k\)](#)

Webster's Unabridged Dictionary

The Online Plain Text English Dictionary

OPTED is a public domain English word list dictionary, based on the public domain portion of "The Project Gutenberg Etext of Webster's Unabridged Dictionary" which is in turn based on the 1913 US Webster's Unabridged Dictionary. (See [Project Gutenberg](#))

This version has been extensively stripped down and set out as one definition per line. All the Gutenberg EText tags and formatting have been removed by computer. Version 0.03 is a new processing of v0.47 of the websters dictionary and it has considerably fewer errors. Also the definition limit of 255 chars has been removed to give full justice of some of the more majestic of the originals. Some important errors in the parts-of-speech fields have been corrected and a lot of inflections/ alternatives and plurals that were missed due to software bugs in v0.01 and 0.02 are now included properly.

Look at the function
`getWebsterDictionary`
in `DataImport.py`

and process them in some way on their own machine. The only usage conditions are that if the material is redistributed, the content (not the formatting) remain in the public domain (ie free) and that the content be easily accessible in non-encoded plain text format at no cost to the end user. The origin of the content should also be acknowledged, including OPTED, Project Gutenberg and the 1913 edition of Webster's Unabridged Dictionary. If the material is to be included in commercial products, Project Gutenberg should be contacted first. There are no restrictions for personal or research uses of this material.

OPTED v0.03 by Letter(size)

Second computer generated version:

[A\(1.1M\)](#) | [B\(1005k\)](#) | [C\(1.6M\)](#) | [D\(1M\)](#) | [E\(809k\)](#) | [F\(784k\)](#) | [G\(564k\)](#) | [H\(686k\)](#) | [I\(833k\)](#) | [J\(172k\)](#) | [K\(172k\)](#) | [L\(637k\)](#) | [M\(931k\)](#) | [N\(343k\)](#) | [O\(466k\)](#) | [P\(1.5M\)](#) | [Q\(147k\)](#) | [R\(931k\)](#) | [S\(2.1M\)](#) | [T\(1005k\)](#) | [U\(343k\)](#) | [V\(343k\)](#) | [W\(490k\)](#) | [X\(49k\)](#) | [Y\(74k\)](#) | [Z\(74k\)](#)

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- Example Project 2 Proposal

The American Presidency Project



John Woolley and Gerhard Peters

HOME DATA DOCUMENTS ELECTIONS MEDIA LINKS

Presidential Debates • 1960 - 2012

Click on
Republican Candidates Debate in
Mesa, AZ

	February 22nd, 2012	Republican Candidates Debate in Mesa, Arizona
	January 26th, 2012	Republican Candidates Debate in Jacksonville, Florida
	January 23rd, 2012	Republican Candidates Debate in Tampa, Florida
	January 19th, 2012	Republican Candidates Debate in Charleston, South Carolina
	January 16th, 2012	Republican Candidates Debate in Myrtle Beach, South Carolina
2012	January 8th, 2012	Republican Candidates Debate in Concord, New Hampshire
	January 7th, 2012	Republican Candidates Debate in Manchester, New Hampshire
	December 15th, 2011	Republican Candidates Debate in Sioux City, Iowa
	December 10th, 2011	Republican Candidates Debate in Des Moines, Iowa
	November 22nd, 2011	Republican Candidates Debate in Washington, DC

- Pub
- Stat
- Addre
- Inau
- Wee
- Fire
- Nev
- Exe
- Prod
- Signing Statements
- Press Briefings
- Statements of Administration Policy
- Economic Report of the President
- Debates
- Convention Speeches
- Party Platforms
- 2012 Election Documents
- 2008 Election Documents
- 2004 Election Documents
- 1960 Election Documents
- 2009 Transition

The American Presidency Project

The screenshot shows the website's interface. At the top is a navigation bar with links: HOME, DATA, DOCUMENTS, ELECTIONS, MEDIA, and LINKS. Below this is a search bar with the text "John Woolley and Gerhard Peters" and a search icon. The main content area is titled "Presidential Debates • 1960 - 2012". A large text box in the center of the page reads: "Look at the function getTranscript in DataImport.py". Below this, a table lists presidential debates for the year 2012.

Date	Event
February 22nd, 2012	Republican Candidates Debate in Mesa, Arizona
January 26th, 2012	Republican Candidates Debate in Jacksonville, Florida
January 23rd, 2012	Republican Candidates Debate in Tampa, Florida
January 19th, 2012	Republican Candidates Debate in Charleston, South Carolina
January 16th, 2012	Republican Candidates Debate in Myrtle Beach, South Carolina
January 8th, 2012	Republican Candidates Debate in Concord, New Hampshire
January 7th, 2012	Republican Candidates Debate in Manchester, New Hampshire
December 15th, 2011	Republican Candidates Debate in Sioux City, Iowa
December 10th, 2011	Republican Candidates Debate in Des Moines, Iowa
November 22nd, 2011	Republican Candidates Debate in Washington, DC

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- Example Project 2 Proposal

Project 2 Rubric

Category	# Points Earned
Proposal	15
Degree of Difficulty	15
Execution	25
Code Quality	15
Website	15
Analysis	15
TOTAL	100

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- **Example Project 2 Proposal**

Anna Ritz Project 2 Proposal

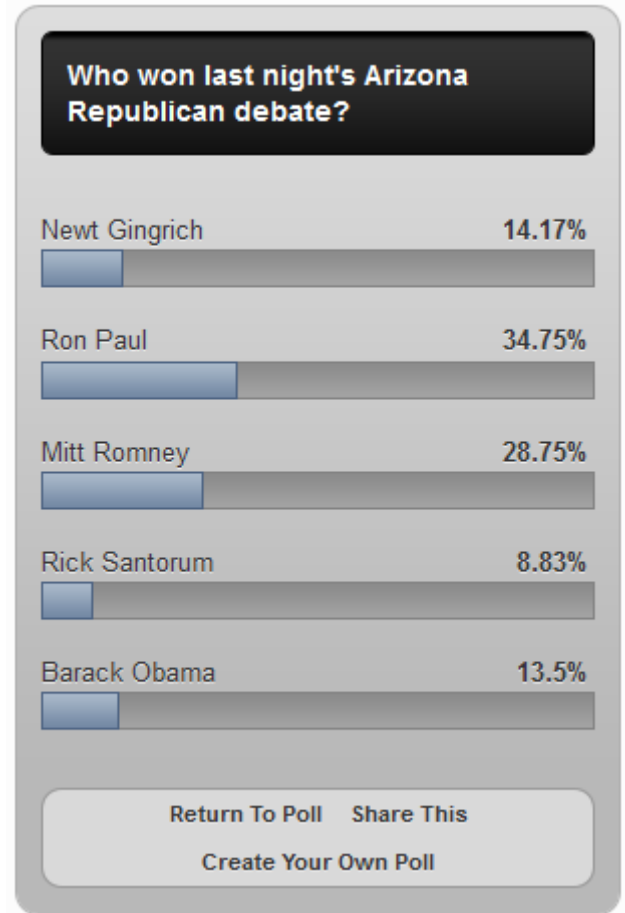
Background: After each debate, there's lots of talk about who "won" it, i.e.

http://www.washingtonpost.com/blogs/the-fix/post/arizona-republican-debate-winners-and-losers/2012/02/22/gIQAsKkVUR_blog.html

I will define the "winner" as the person who received applause the most frequently during the debate.

Claim: I claim that in the AZ debate, Romney "won" and Santorum "lost" – that is, Romney received applause the most and Santorum received applause the least.

....



http://blogs.phoenixnewtimes.com/va-lleyfever/2012/02/who_won_last_nights_arizona_re.php

Skeleton Code

Skeleton Code

```
# Anna Ritz
# Project 2
# Skeleton Code

## This program assesses how "popular" each republican candidate is
## by counting the number of [applause] tags in a
## Republican debate.

# CONSTANT VARIABLES (will NEVER change values) are in ALL CAPS
# If you put these variables OUTSIDE all functions, then you
# can access them in ANY function.
CANDIDATES = ['GINGRICH', 'PAUL', 'ROMNEY', 'SANTORUM']
DEBATE_FILE = 'AZDebate.txt'

def assessPopularity():
    '''Assesses the popularity of the candidates in the AZ debate.
    INPUTS: none
    OUTPUTS: none'''

    # Step 1: Read the debate file
    myString = readFile()

    # Step 2: For each candidate, assess popularity
    for cand in CANDIDATES:
        countApplause(cand, myString)

    return

def readFile():
    '''Reads DEBATE_FILE and returns a string.
    INPUTS: none
    OUTPUTS: String of the debate'''

    return '' # returns an empty string for now.

def countApplause(candidate, debateString):
    '''Assesses the popularity of the candidate in the debateString.
    INPUTS: candidate (String) - name of candidate
```

Anna Ritz
Project 2 Proposal

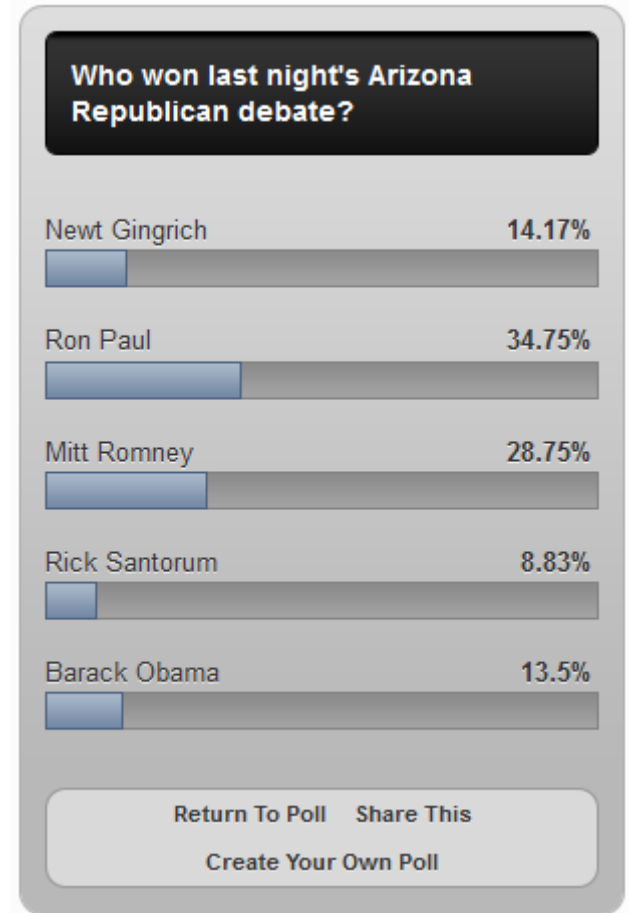
...

Claim: I claim that in the AZ debate, Romney “won” and Santorum “lost” – that is, Romney received applause the most and Santorum received applause the least.

....

Backup Plan: ???

Increasing Degree of Difficulty: ???



http://blogs.phoenixnewtimes.com/va-lleyfever/2012/02/who_won_last_nights_arizona_re.php

Today: Project 2

- Data Sources
 - Project Gutenberg
 - English Dictionary
 - Debate Transcripts
- Project 2 Description
- Example Project 2 Proposal

Next Few Weeks

Sun	Mon	Tues	Wed	Thurs	Fri	Sat
3/17	3/18	3/19	3/20	3/21	3/22	3/23
		Project 2 Out		HW2-4 Due HW2-5 Out		
3/24	3/25	3/26	3/27	3/28	3/29	3/30
Spring Break! Woouoooo!						
3/31	4/1	4/2	4/3	4/4	4/5	4/6
		HW2-5 Due		Proposal Due		
4/7	4/8	4/9	4/10	4/11	4/12	4/14
				Project 2 Due		