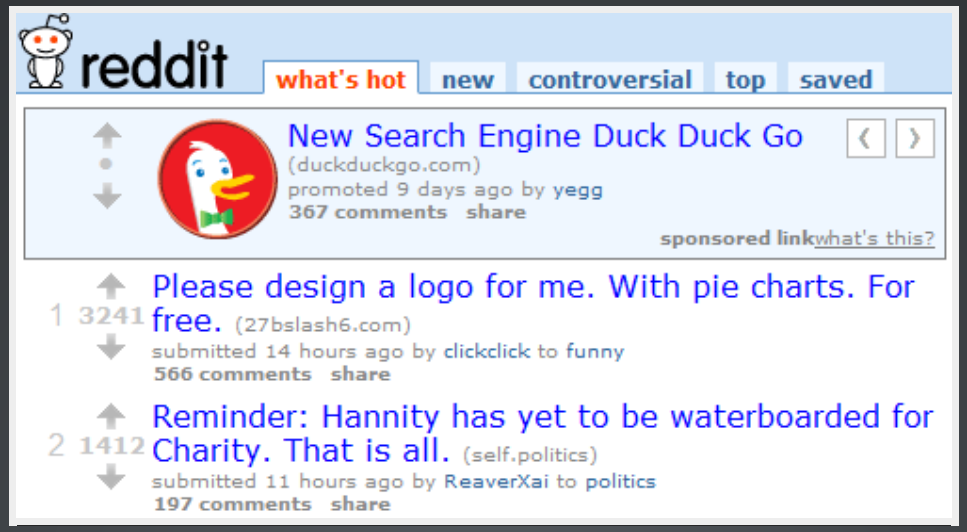

reddit
what's hot
new
controversial
top
saved
New Search Engine Duck Duck Go
(duckduckgo.com)
promoted 9 days ago by yegg
367 comments share
sponsored linkwhat's this?
1 3241
Please design a logo for me. With pie charts. For free. (27bslash6.com)
submitted 14 hours ago by clickclick to funny
566 comments share
2 1412
Reminder: Hannity has yet to be waterboarded for Charity. That is all. (self.politics)
submitted 11 hours ago by ReaverXai to politics
197 comments share

[–] **bitofnewsbot** 2715 points 8 hours ago

Article summary:

- It looked a lot like the U.S. shutdown of today, or the 17 previous U.S. shutdowns.
- Australia's 1975 shutdown ended pretty differently, though, than they do here in America.
- Queen Elizabeth II's official representative in Australia, Governor-General Sir John Kerr, simply dismissed the prime minister.
- At 4:50 p.m., Kerr dissolved the rest of Parliament, essentially firing everyone, with a formal proclamation that ended with the words "God Save the Queen."
- This means that, legally speaking, the 1975 Australian government funding crisis ended because Queen Elizabeth II dismissed everyone in the government.

Powered by TextTeaser API

permalink

[–] **greenbowl** 1219 points 7 hours ago

Did anyone notice that this summary is posted by a bot?

permalink **parent**

SUMMARIZING TEXT IN PYTHON

by Matt Gallivan

```
def summarize(text, num_summaries=3):
    '''Create small summaries of a larger text.'''
    summaries = []
    # ...
    return summaries[:num_summaries]
```

HOW?

TEXTRANK

1. BREAK TEXT INTO PIECES.
2. CONNECT THE PIECES IN A GRAPH.
3. FIND THE MOST IMPORTANT PIECE.

1. BREAK TEXT INTO PIECES

SENTENCES!

```
import re

def sentences(text):
        '''Break text blob into sentences.'''
        ends = re.compile('[.?!]')
        return ends.split(text)
```

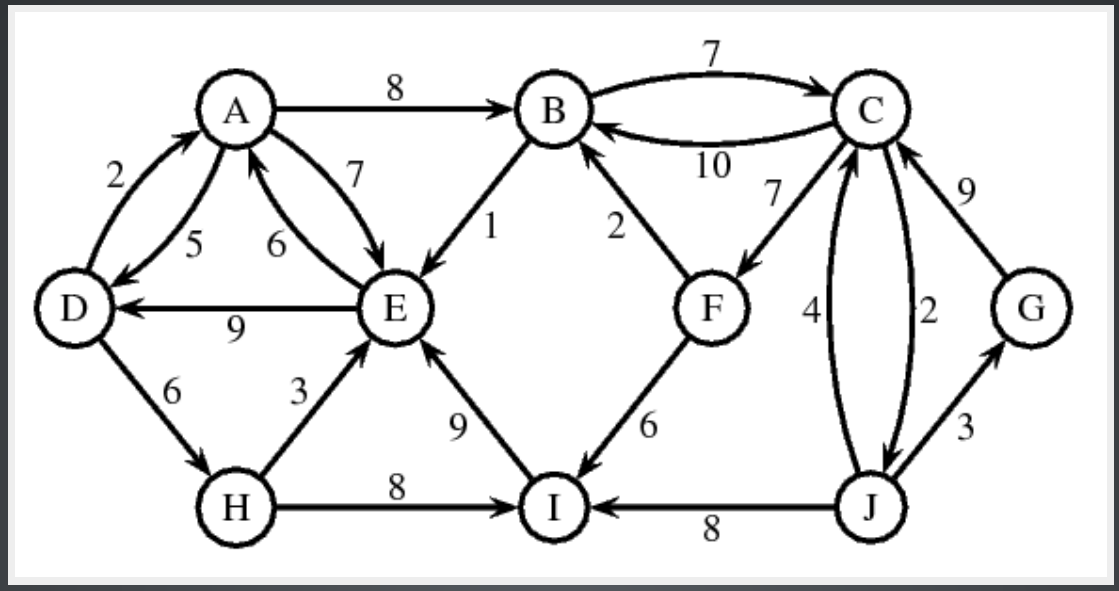
BETTER SENTENCES!

```
from nltk.tokenize import sent_tokenize

def sentences(text):
        '''Break text blob into sentences.'''
        return sent_tokenize(text)
```

2. CONNECT THE PIECES IN A GRAPH

GRAPH?


A
B
C
D
E
F
G
H
I
J
8
7
10
2
7
5
6
1
2
7
9
4
2
9
6
3
9
6
3
8
8

EDGES

```
def connect(nodes):
        '''Return a list of edges connecting the nodes,
       where the edges are given a weight based on their
       similarity.'''
       return [(start, end, similarity(start, end))
                           for start in nodes
                           for end in nodes
                           if start is not end]
```

```
def similarity(c1, c2):
        '''Return the amount of similarity between two chunks.'''
       return len(common_words(c1, c2)) /
              (log(len(words(c1))) + log(len(words(c2))))
```

3. FIND THE MOST IMPORTANT PIECE

PAGERANK

WHAT WEBSITES ARE THE MOST IMPORTANT?

```
# INPUT:
nodes = get_all_websites()
edges = connect_all_websites_that_link_to_each_other(nodes)

# OUTPUT
pagerank(nodes, edges) # =>

{
    'www.google.com' : 40000000000000000,
    'www.yahoo.com' : 25,
    'www.reddit.com' : 23,
    'news.ycombinator.com' : 14,
    # ...
}
```

PAGERANK

USE SENTENCES INSTEAD OF WEBSITES!

```
# INPUT:
nodes = sentences(text)
edges = connect(nodes)

# OUTPUT
pagerank(nodes, edges) # =>

{
    'A really good summary sentence!' : 243,
    'Pretty good at summarizing - maybe too specific though' : 165,
    'Not that great.' : 142,
    # ...
}
```

PAGERANK

```
import networkx as nx

def rank(nodes, edges):
    '''Return a dictionary containing the scores for each vertex.'''
    graph = nx.DiGraph()
    graph.add_nodes_from(nodes)
    graph.add_weighted_edges_from(edges)
    return nx.pagerank(graph)
```

PUTTING IT TOGETHER

```
def summarize(text, num_summaries=3):
    '''Create small summaries of a larger text.'''
    nodes = sentences(text)
    edges = connect(nodes)
    scores = rank(nodes, edges)
    return sorted(scores, key=data.get)[:num_summaries]
```

ADOBE CUSTOMER DATA

'VERY RECENTLY, ADOBE'S SECURITY TEAM DISCOVERED SOPHISTICATED ATTACKS ON OUR NETWORK, INVOLVING ILLEGAL ACCESS OF CUSTOMER INFORMATION AS WELL AS SOURCE CODE FOR NUMEROUS ADOBE PRODUCTS', ARKIN SAID IN A BLOG POST.

TWITTER IPO

THE COMPANY UNSEALED THE DOCUMENTS ON THURSDAY, DISCLOSING THAT IT GENERATED 317 MILLION IN REVENUE IN 2012 AND THAT IT HAD MORE THAN 218 MILLION ACTIVE USERS AS OF THE END OF JUNE, UP 44 PERCENT FROM A YEAR EARLIER.

TWITTER HAS REVEALED ITS HIGHLY ANTICIPATED STOCK OFFERING, WITH THE HUGELY POPULAR MESSAGING PLATFORM STATING THAT IT SEEKS TO RAISE $1BN.

GREEN EGGS AND HAM

I DO NOT LIKE THEM IN A HOUSE.

I DO NOT LIKE THEM IN A BOX.

I WILL NOT EAT THEM IN A HOUSE.

I DO NOT LIKE THEM WITH A FOX.

I DO NOT LIKE THEM WITH A MOUSE.

AND I WOULD EAT THEM WITH A GOAT... AND I WILL EAT THEM, IN THE RAIN.

AND I WILL EAT THEM, IN THE RAIN.

IN CONCLUSION...

1. MACHINE LEARNING DOESN'T HAVE TO BE HARD
2. THERE ARE A TON OF LIBRARIES TO HELP YOU
3. MACHINE LEARNING IS GOOD