

**P1.**

In order to extract raw html from my text file of at least 1000 URIs, I used a python library called “newspaper. It greatly facilitated the html and text (article) extraction process. I also had to use sys to handle a Unicode error, however, I am not sure if it completely resolved the issue. Some research showed that this was bad practice, but it helped enough to generate some results.

Notes: Not all of my URIs were fully redirected. With that being said, this will make for difficulty/ error in my ranking later on. Additionally, my code broke after about 256 URIs were processed. I am not entirely sure what caused it, but I proceeded with the assignment to the best of my abilities.

**Resources:**

<http://newspaper.readthedocs.io/en/latest/>

<https://pypi.python.org/pypi/newspaper>

<http://stackoverflow.com/questions/28917686/python-read-urls-from-file-and-print-to-file>

<http://stackoverflow.com/questions/21129020/how-to-fix-unicodedecodeerror-ascii-codec-cant-decode-byte>

text.py

twitter\_streaming.py

```
1  # encoding=utf8
2  from bs4 import BeautifulSoup
3  from newspaper import Article
4  import urllib
5  import requests
6  import sys
7
8  reload(sys)
9  sys.setdefaultencoding('utf8')
10
11  count = 1
12
13  with open('urls.txt', 'r') as f: #text file containing the URLs
14
15
16      for url in f:
17          if url.strip() != "":
18              pFile = 'processed_'+str(count)+'.txt'
19              rFile = 'raw_'+str(count)+'.txt'
20
21              with open(rFile, 'a') as fout:
22                  print(u"URL Line: {}".format(url.encode('utf-8')))
23
24                  # you might want to remove endlines and whitespaces from
25                  # around the URL, which what strip() does
26                  article = Article(url.strip())
27                  article.download()
28                  article.parse()
29
30                  # write/append to file
31                  fout.write(article.html)
32                  print(u"Raw HTML: {}".format(article.html.encode('utf-8')))
33
34              with open(pFile, 'a') as fout:
35                  print(u"URL Line: {}".format(url.encode('utf-8')))
36
37                  # you might want to remove endlines and whitespaces from
38                  # around the URL, which what strip() does
39                  article = Article(url.strip())
40                  article.download()
41                  article.parse()
```

```
# write/append to file










fout.write(article.title)
fout.write(article.text)


print(u>Title: {}".format(article.title.encode('utf-8'))))

# print authors only if there are authors to show.
if len(article.authors) == 0:
    print('No author!')
else:
    for author in article.authors:
        print(u"Author: {}".format(author.encode('utf-8'))))

print("Text of the article:")
print(article.text.encode('utf-8'))
count+=1
```

## Sample raw html:

Name	Date modified	Type	Size
 raw_1.txt	2/25/2017 8:34 PM	Text Document	32 KB
 raw_2.txt	2/25/2017 8:35 PM	Text Document	139 KB
 raw_3.txt	2/25/2017 8:35 PM	Text Document	157 KB
 raw_4.txt	2/25/2017 8:35 PM	Text Document	12 KB
 raw_5.txt	2/25/2017 8:35 PM	Text Document	160 KB
 raw_6.txt	2/25/2017 8:35 PM	Text Document	44 KB
 raw_7.txt	2/25/2017 8:35 PM	Text Document	105 KB
 raw_8.txt	2/25/2017 8:35 PM	Text Document	36 KB
 raw_9.txt	2/25/2017 8:35 PM	Text Document	385 KB

 raw\_1.txt - Notepad

File Edit Format View Help












```
<!DOCTYPE html>
<html lang="en-US" prefix="og: http://ogp.me/ns# fb: http://ogp.me/ns/fb# article: http://ogp.me/ns/art:
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="profile" href="http://gmpg.org/xfn/11">
<link rel="pingback" href="https://1bitcoins.net/xmlrpc.php">
<title>There is Bitcoin and&#8230; there is Blockchain BS! | 1 Bitcoins</title>

<!-- SEO Ultimate (http://www.seodesignsolutions.com/wordpress-seo/) -->
<meta property="og:type" content="article" />
<meta property="og:title" content="There is Bitcoin and&#8230; there is Blockchain BS!" />
<meta property="og:url" content="https://1bitcoins.net/there-is-bitcoin-and-there-is-blockchain-
<meta property="article:published_time" content="2017-02-26" />
<meta property="article:modified_time" content="2017-02-26" />
<meta property="article:author" content="https://1bitcoins.net/author/admin/" />
<meta property="article:tag" content="Bitcoin News" />
<meta property="article:tag" content="News" />
<meta property="article:tag" content="Reddit" />
<meta property="article:tag" content="247 Bitcoin" />
<meta property="article:tag" content="Bitcoin" />
<meta property="article:tag" content="Bitcoin News" />
<meta property="article:tag" content="Bitcoins" />
<meta property="article:tag" content="BTC" />
<meta property="article:tag" content="Reddit" />
<meta property="og:site_name" content="1 Bitcoins" />
<meta name="twitter:card" content="summary" />

<!-- /SEO Ultimate -->

<link rel='dns-prefetch' href='//fonts.googleapis.com' />
<link rel='dns-prefetch' href='//s.w.org' />
<link rel="alternate" type="application/rss+xml" title="1 Bitcoins &raquo; Feed" href="https://1bitcoin:
<link rel="alternate" type="application/rss+xml" title="1 Bitcoins &raquo; Comments Feed" href="https://,
<link rel="alternate" type="application/rss+xml" title="1 Bitcoins &raquo; There is Bitcoin and&#8230; there is Blockchain BS!
<script type="text/javascript">
    window._wpemojiSettings = {"baseUrl":"https://s.w.org/images/core/emoji/2.
    !function(a,b,c){function d(a){var b,c,d,e,f=String.fromCharCode;if(!k||!k.fill
    (i[h]),c.supports.everything=c.supports.everything&&c.supports[i[h]],"flag"!==i[h]&&(c.supports.everyth:
    </script>
<style type="text/css">
```

### Sample processed files:

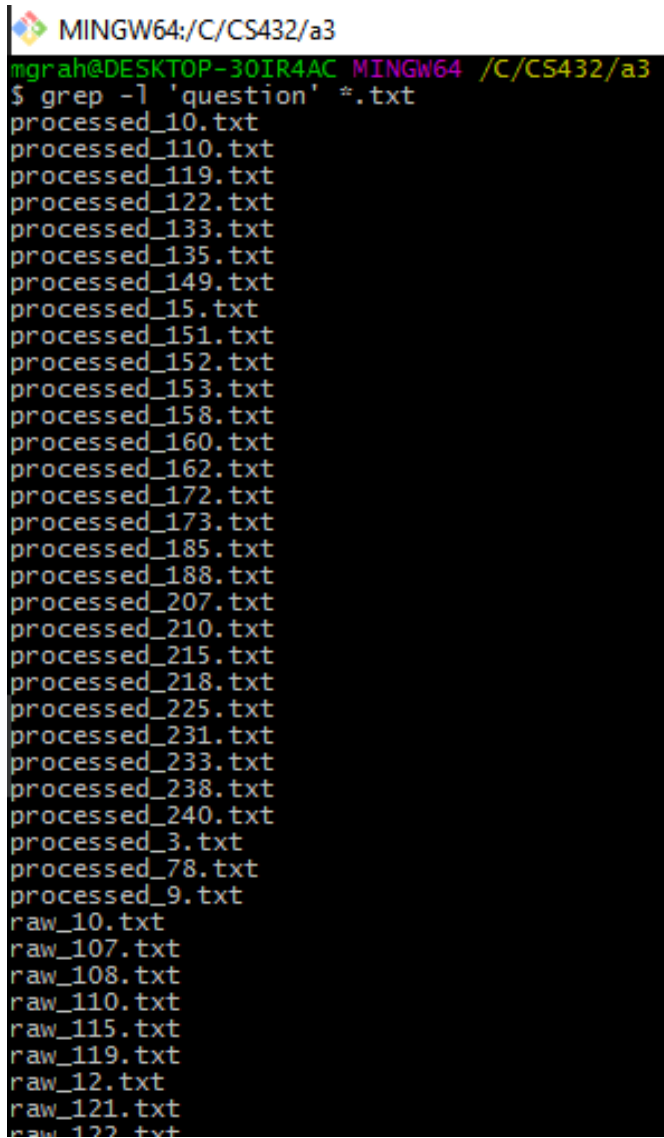
Name	Date modified	Type	Size
 processed_1.txt	2/25/2017 8:35 PM	Text Document	1 KB
 processed_2.txt	2/25/2017 8:35 PM	Text Document	1 KB
 processed_3.txt	2/25/2017 8:35 PM	Text Document	4 KB
 processed_4.txt	2/25/2017 8:35 PM	Text Document	1 KB
 processed_5.txt	2/25/2017 8:35 PM	Text Document	7 KB
 processed_6.txt	2/25/2017 8:35 PM	Text Document	3 KB
 processed_7.txt	2/25/2017 8:35 PM	Text Document	2 KB
 processed_8.txt	2/25/2017 8:35 PM	Text Document	2 KB
 processed_9.txt	2/25/2017 8:35 PM	Text Document	2 KB
 processed_10.txt	2/25/2017 8:35 PM	Text Document	7 KB
 processed_3.txt - Notepad		— □ ×	

File Edit Format View Help

McCain cuts off questions about Saudi donationSen. John McCain cut off  
PHOENIX - Sen. John McCain on Friday cut off questions from reporters a  
McCain denied that he had any link to the McCain Institute for Internat  
"I'm proud that the institute is named after me, but I have nothing to  
The Saudi donation was first reported Thursday by Bloomberg Politics.  
McCain laughed off comparisons to the Clinton Foundation, whose fundrai  
"Look at what the Clinton institute has raised in the form of money," M  
McCain cut off questions about the donations after a few minutes.  
"I've answered all the questions," he said. "I've answered all the ques  
The McCain Institute web site shows several longtime McCain supporters  
IRS documents for the McCain Institute Foundation show the foundation's  
One of the institute's marquee events, the Sedona Forum, features McCai  
"The institute carries John McCain's name, it advertises itself as purs  
"He may not have any kind of direct control over the institute, but it'  
Foreign governments are barred from making campaign donations to U.S. p  
"Anyone who's going to chip in a million dollars or more to the McCain  
As chairman of the powerful Senate Armed Services Committee and a leadi  
"They want to see the U.S. involved in Syria. They want to see our nego  
Former State Sen. Kelli Ward, one of McCain's opponents in the August R  
"I hate to see foreign governments playing in our political process," W

## P2.

Then I used grep to search the documents for the term “question.”

A screenshot of a Windows command prompt window with a black background and white text. The title bar at the top reads 'MINGW64: C:/CS432/a3'. The prompt is 'mgraph@DESKTOP-30IR4AC MINGW64 /C/CS432/a3'. The command entered is '\$ grep -l 'question' \*.txt'. The output lists 35 files: processed\_10.txt through processed\_240.txt, processed\_3.txt, processed\_78.txt, processed\_9.txt, raw\_10.txt, raw\_107.txt, raw\_108.txt, raw\_110.txt, raw\_115.txt, raw\_119.txt, raw\_12.txt, raw\_121.txt, and raw\_122.txt.

```
MINGW64: C:/CS432/a3
mgraph@DESKTOP-30IR4AC MINGW64 /C/CS432/a3
$ grep -l 'question' *.txt
processed_10.txt
processed_110.txt
processed_119.txt
processed_122.txt
processed_133.txt
processed_135.txt
processed_149.txt
processed_15.txt
processed_151.txt
processed_152.txt
processed_153.txt
processed_158.txt
processed_160.txt
processed_162.txt
processed_172.txt
processed_173.txt
processed_185.txt
processed_188.txt
processed_207.txt
processed_210.txt
processed_215.txt
processed_218.txt
processed_225.txt
processed_231.txt
processed_233.txt
processed_238.txt
processed_240.txt
processed_3.txt
processed_78.txt
processed_9.txt
raw_10.txt
raw_107.txt
raw_108.txt
raw_110.txt
raw_115.txt
raw_119.txt
raw_12.txt
raw_121.txt
raw_122.txt
```

Then I opened the text files in Word so I could search for how many times the search word occurs in each document. Word also shows you the word count at the bottom left of the document.

Sample search:

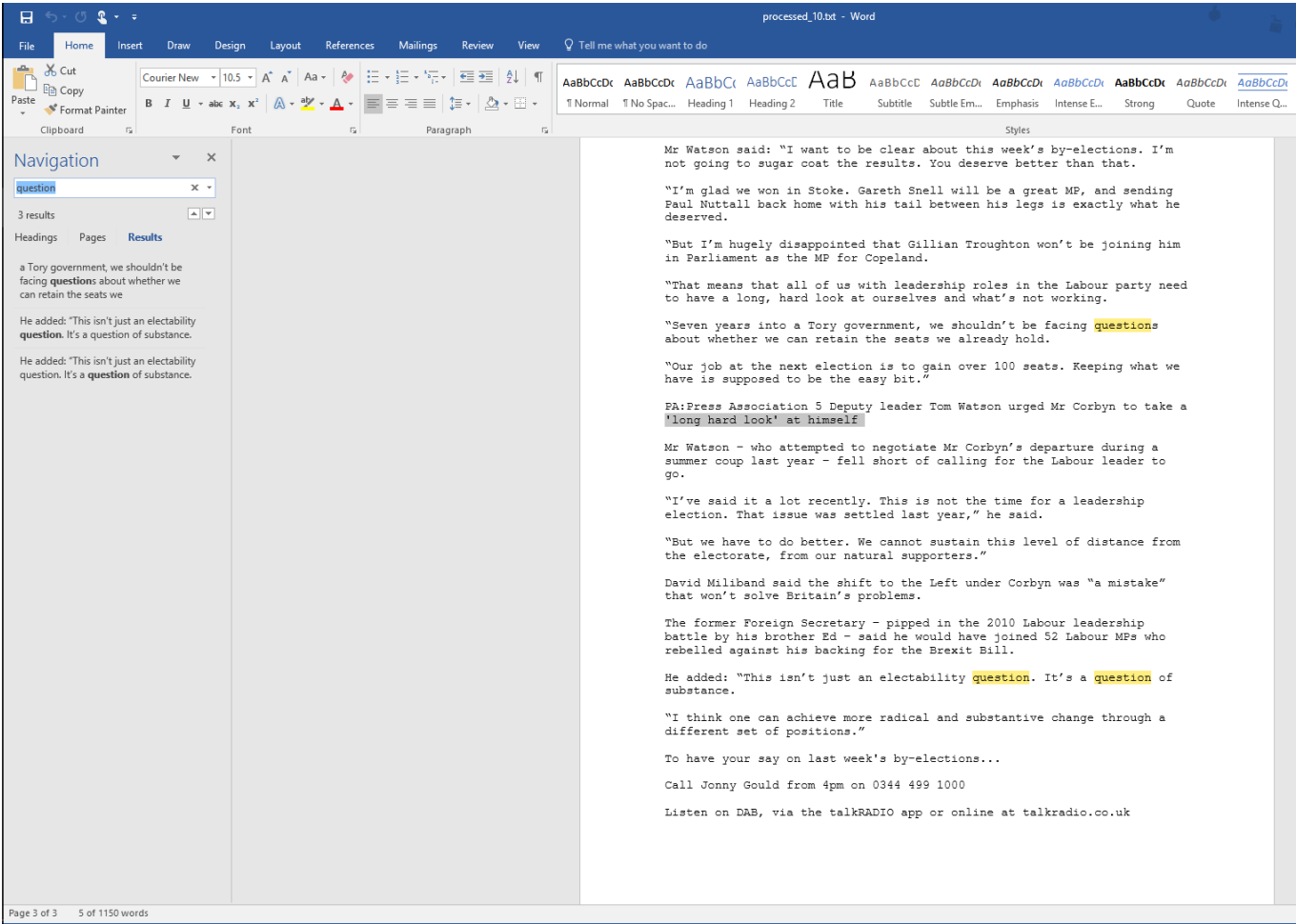


Table:

A	B	C	D	E	F	G	H
	Document with Hits	# of Hits	Total # of Words in Document	T/F	IDF	IDF(term)	TF-IDF
	processed_10.txt	3	1150	0.002609	25.5	4.6724	0.066522
	processed_110.txt	1	801	0.001248	25.5		0.031835
	processed_119.txt	1	986	0.001014	25.5		0.025862
	processed_122.txt	2	393	0.005089	25.5		0.129771
	processed_133.txt	5	588	0.008503	25.5		0.216837
	processed_135.txt	1	814	0.001229	25.5		0.031327
	processed_149.txt	2	1208	0.001656	25.5		0.042219
	processed_15.txt	1	968	0.001033	25.5		0.026343
	processed_151.txt	1	1121	0.000892	25.5		0.022748
	processed_152.txt	1	396	0.002525	25.5		0.064394
Total Docs		255					
Docs with Hits (Max = 10)		10					

**P3.**

**URI #1:** <https://www.thesun.co.uk/news/2956495/theresa-may-labour-off-the-map-jeremy-corbyn/>

**Rank:** 7/10

**Sample (URI #1):**

## Check PAGE RANK of Web site pages Instantly

In order to [check pagerank](#) of a single web site, web page or domain name, please submit the URL of that web site, web page or domain name to the form below and click "Check PR" button.

Check PR

Web Page URL: <https://www.thesun.co.uk>

The Page Rank: **7/10**

(the page rank value is 7 from 10 possible points)

**URI #2:** <https://www.forbes.com/sites/ewanspence/2017/02/24/apple-news-headlines-iphone-new-leak-touchid-macbook-pro/#12789ca71738>

**Rank:** 8/10

**URI #3:** [https://sg.news.yahoo.com/duterte-fans-foes-mass-manila-rallies-140800577.html?soc\\_src=social-sh&soc\\_trk=tw](https://sg.news.yahoo.com/duterte-fans-foes-mass-manila-rallies-140800577.html?soc_src=social-sh&soc_trk=tw)

**Rank:** 6/10

**URI #4:** <http://www.star-telegram.com/news/local/community/northeast-tarrant/article135021414.html>

**Rank:** 6/10

**URI #5:** <https://blogs.blogtowinn.com/4-pros-and-cons-of-hiring-a-content-marketing-firm-to-write-your-company-blog-business-2-community-blog/>

**Rank:** 0/10

**URI #6:** <https://apple.news/AaaWu4kGyRXiCkIXw13w2A>

**Rank:** 0/10



**URI #7:** [https://tonic.vice.com/en\\_us/article/were-all-guinea-pigs-in-a-failed-decades-long-diet-experiment](https://tonic.vice.com/en_us/article/were-all-guinea-pigs-in-a-failed-decades-long-diet-experiment)

**Rank:** 0/10

**URI #8:** [http://www.npr.org/2017/02/25/517257273/trump-will-be-first-president-in-36-years-to-skip-white-house-correspondents-din?utm\\_medium=RSS&utm\\_campaign=news](http://www.npr.org/2017/02/25/517257273/trump-will-be-first-president-in-36-years-to-skip-white-house-correspondents-din?utm_medium=RSS&utm_campaign=news)

**Rank:** 8/10

**URI #9:** <http://www.nationalreview.com/article/445237/sweden-crime-rates-statistics-immigration-trump-fox-news>

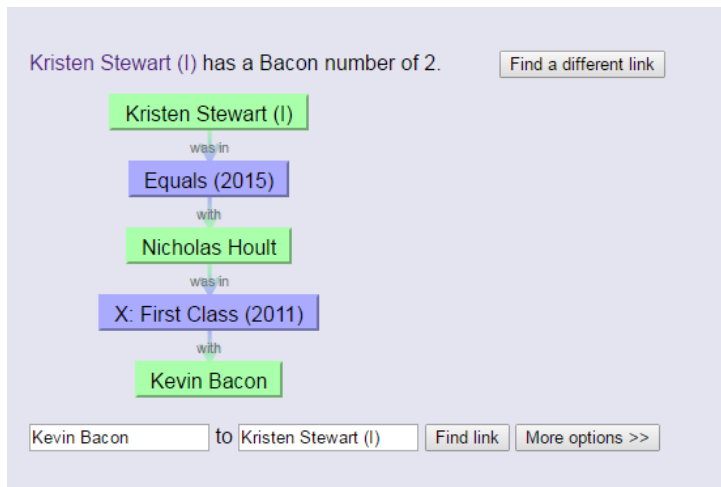
**Rank:** 7/10

**URI #10:** <http://gopurpleaces.com/news/2017/2/25/baseball-drops-9-2-decision-to-southern-miss.aspx>


**Rank:** 0/10

P6.

Stewart has a Bacon number of 2:



She wrote a paper with Bhautik Joshi.

**Cornell University Library**

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > cs > arXiv:1701.04928

[\(Help\)](#) [Advanced search](#)

Computer Science > Computer Vision and Pattern Recognition

## Bringing Impressionism to Life with Neural Style Transfer in Come Swim

[Bhautik Joshi](#), [Kristen Stewart](#), [David Shapiro](#)  
(Submitted on 18 Jan 2017)

Neural Style Transfer is a striking, recently-developed technique that uses neural networks to artistically redraw an image in the style of a source style image. This paper explores the use of this technique in a production setting, applying Neural Style Transfer to redraw key scenes in 'Come Swim' in the style of the impressionistic painting that inspired the film. We document how the technique can be driven within the framework of an iterative creative process to achieve a desired look, and propose a mapping of the broad parameter space to a key set of creative controls. We hope that this mapping can provide insights into priorities for future research.

Comments: 3 pages, 6 figures, paper is a case study of how Neural Style Transfer can be used in a movie production context

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**

ACM classes: I.3.3; I.4.0

Cite as: [arXiv:1701.04928 \[cs.CV\]](#)  
(or [arXiv:1701.04928v1 \[cs.CV\]](#) for this version)

**Download:**

- PDF
- Other formats (license)

Current browse context:  
cs.CV  
< prev | next >  
new | recent | 1701

Change to browse by:  
cs

References & Citations

- NASA ADS

1 blog link (what is this?)


DBLP - CS Bibliography  
listing | bibtex  
[Bhautik Joshi](#)  
[Kristen Stewart](#)  
[David Shapiro](#)

Bookmark (what is this?)

Joshi collaborated with Sébastien Ourselin on a paper about simulation.

ResearchGate

Join for freeLog in



Bhautik Joshi

Adobe Systems Inc, San Jose

Biomedical Engineering, Computer Engineering, Data Structures

PhD

il 4.90

Overview


Contributions




Info

Stats

Research

Article: **Bringing Impressionism to Life with Neural Style Transfer in Come Swim**

 Full-text available · Article · Jan 2017




 Bhautik Joshi  Kristen Stewart  David Shapiro

[Download full-text](#)

0 Citations 10 Reads

Article: **A quantitative assessment of approaches to mesh generation for surgical simulation**

Article · Oct 2008 · Engineering With Computers

 Bhautik Joshi  Andriy Fedorov +1 more author...  Sébastien Ourselin

[Read](#)

3 Citations 25 Reads

# A quantitative assessment of approaches to mesh generation for surgical simulation

**Article** in Engineering With Computers 24(4):417-430 · October 2008 with 24 Reads

DOI: 10.1007/s00366-008-0088-z · Source: DBLP



1st [Bhautik Joshi](#)  
i10 4.9 · Adobe Systems Inc



2nd [Andrey Fedorov](#)  
i10 28.32 · Brigham and Women's Hospital

Last [Sébastien Ourselin](#)

[Show more authors](#)

10 Publications 60 Citations 0 Projects

**Bhautik Joshi**



Institution  
Adobe Systems Inc

Department  
Engineering

[Message author](#)

meshes as models for anatomy. These meshes are used in modelling schemes. A variety of mesh generation schemes are used to generate meshes from segmented anatomical volumes. Each scheme, however, has its own characteristics. When choosing a scheme, one must consider the quality, patient-specific meshes that provide a good approximation of the anatomy within the bounds required for real-

time interaction, there is also a limit imposed on the number of elements in the mesh generated. To the authors' knowledge, there has been little work directly assessing the suitability of mesh generators for surgical simulation. This paper seeks to address this issue by assessing the use of six mesh generators in a surgical simulation scenario, and examining how they affect simulation precision. This paper aims to perform these comparisons against high-resolution reference meshes, where we examine the precision of meshes from the same mesh generator at different levels of complexity.

He has an Erdős number of 4:

[Home](#) [Preferences](#) [Free Tools](#) [About](#) [Librarians](#) [Terms of Use](#) [Blog](#)

AMERICAN MATHEMATICAL SOCIETY  
**MathSciNet**  
Mathematical Reviews  
ISSN 2167-5163

Old Dominion Univ

[Search MSC](#) [Collaboration Distance](#) [Current Journals](#) [Current Publications](#)

**MR Erdos Number = 4**

<a href="#">Sébastien Ourselin</a>	coauthored with	<a href="#">Katherine J. Dobson</a>	<a href="#">MR3224124</a>
<a href="#">Katherine J. Dobson</a>	coauthored with	<a href="#">Kees Joost Batenburg</a>	<a href="#">MR3390321</a>
<a href="#">Kees Joost Batenburg</a>	coauthored with	<a href="#">Robert Tijdeman</a>	<a href="#">MR2833908</a>
<a href="#">Robert Tijdeman</a>	coauthored with	<a href="#">Paul Erdős<sup>1</sup></a>	<a href="#">MR0937987</a>

[Change First Author](#) [Change Second Author](#) [New Search](#)

Free Tool [Help](#) [Support Mail](#)

AMERICAN MATHEMATICAL SOCIETY

Mirror Sites [Providence, RI USA](#)

© Copyright 2017, American Mathematical Society  
[Privacy Statement](#)

By extension, Stewart has an Erdős number of: 6

Therefore, she has an Erdős-Bacon number of: 8

(6+2 = 8)