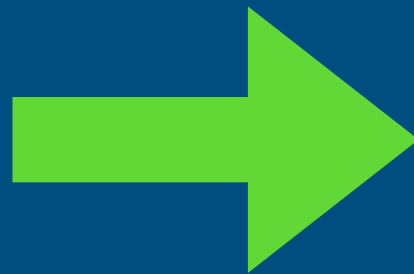


Singapore Govt Press Statements

Outreach in the Indonesian Media

Objective



Jakarta Post (English)

Merdeka

Inilah

Metrotvnews

Kompas

Tribunnews

Reputable, uses Google Custom Search Engine

To analyse to what extent our press statements have been covered by the Indonesian media

Time Frame: 01-01-2016 to 24-07-2017

Steps

Scraping

- 65% of my time / codes

Translating and
Organising Data

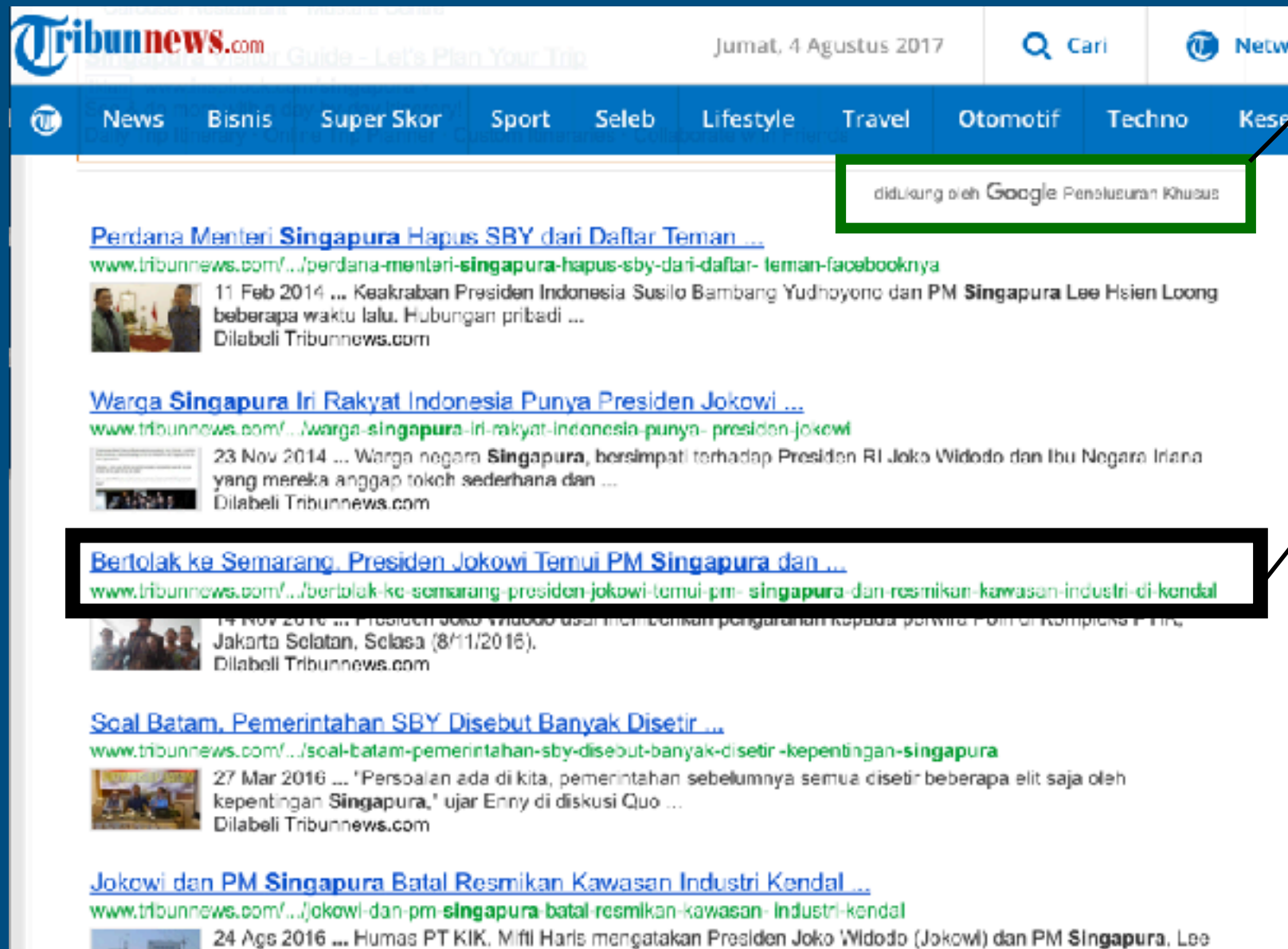
- 30% of my time

Comparing the Documents
Using Sklearn

Scraping (1)

Indon News Site Uses Google Custom Search Engine (GCSE)

Use GCSE API to get the link and title of each article



Go to the url

Extract the date and text of the article using BeautifulSoup

Scraping (2)

SG Government press statements - use own search engines

The screenshot shows the search interface of the Singapore Government press statements website. On the left, there are navigation links: Singapore Headlines, Special Events, and Photo Gallery. The main search area has a 'Keyword' field with 'Indonesia' entered. Below it, the 'Search' section includes checkboxes for 'Press Statements & Releases' (checked), 'Speeches', and 'Transcripts'. The 'Date' section has radio buttons for 'Most recent' (selected), 'Past 24 hrs', 'Month', and 'From' (which is currently selected). The 'From' and 'To' date pickers show 'Jun 2016' and 'Jul 2017' respectively. A 'Search' button is at the bottom of the search section. Below the search section, there is a 'Search result' table with a black border around the first row.

Date	Statement
28-Jul-2017	MFA Press Statement: Visit by Senior Minister of State, Ministry of Defence and Ministry of Foreign Affairs, Dr Mohamed Maliki Osman to Pekanbaru and Jambi City, the Republic of Indonesia, 25 to 28 July 2017
24-Jul-2017	MFA Press Statement: Visit by Senior Minister of State, Ministry of Defence and Ministry of Foreign Affairs, Dr Mohamed Maliki Osman to Pekanbaru and Jambi City, the Republic of Indonesia, 25 to 28 July 2017
17-Jul-2017	MFA Spokesman's Comments in response to media queries on the missing Singaporean diver in the waters of Gili Lawa Laut near Komodo Islands, Indonesia
25-May-2017	MFA Press Statement: Condolence letters from Singapore leaders on the bomb blast in East Jakarta, Indonesia, on 24 May 2017
25-May-2017	MFA Spokesman's Comments on the bomb blast in East Jakarta, Indonesia, on 24 May 2017
25-Apr-2017	MFA Press Statement: Visit by Minister for Foreign Affairs, Dr Vivian Balakrishnan to Batam, Indonesia, 25 April 2017

Extract the date, title, url of the article using BeautifulSoup

Go to the url

Extract the body of the article using BeautifulSoup

Organising the Data

```
In [181]: pd.options.display.max_rows = 200  
df_idnSG
```

Out[181]:

	title	link	snippet	body	date
83	Hari Ketiga ASEAN School di Singapura, Perenan...	http://pekanbaru.tribunnews.com/2017/07/18/har...	6 hari yang lalu ... Hari ketiga pelaksanaan k...	Laporan Wartawan Tribun Pekanbaru, Palti Siaha...	2017-07-18
84	Mau Terbang ke Singapura, Mathew Thomas Dicega...	http://metropolitan.inilah.com/read/detail/239...	6 hari yang lalu ... Ia menjelaskan, pagi ini ...	INILAHCOM, Jakarta - Selain menetapkan putra...	2017-07-18
85	Timnas U-18 Indonesia Terhindar dari Posisi Ju...	http://kupang.tribunnews.com/2017/07/18/timnas...	6 hari yang lalu ... POS KUPANG.COM - Tim sepa...	POS KUPANG.COM - Tim sepak bola nasional U-18 ...	2017-07-18
86	Mikro Seharusnya Dirjen Pajak Bisa Intip Dan...	http://ekonomi.metrotvnews.com/mikro/RkjPOA3N-...	6 hari yang lalu ... Indonesia Kini Bisa Intip...	Metrotvnews.com, Jakarta: Direktorat Jenderal ...	2017-07-18
87	Sudah Mau Naik Pesawat Menuju Singapura Pagi T...	http://wartakota.tribunnews.com/2017/07/18/sud...	6 hari yang lalu ... Axel Matthew Thomas (19) ...	WARTA KOTA, SEMANGGI - Axel Matthew Thomas (19...	2017-07-18
88	Putra Jeremy Thomas dicegah saat mau terbang k...	https://www.merdeka.com/peristiwa/putra-jeremy...	6 hari yang lalu ... Let's be smart. Putra Jer...	Merdeka.com - Axel Mathew telah ditetapkan ter...	2017-07-18
89	Axel jadi tersangka, Jeremy Thomas batal pergi...	https://www.merdeka.com/peristiwa/axel-jadi-te...	6 hari yang lalu ... Tadi pagi, Axel hendak pe...	Merdeka.com - Axel Matthew putra sulung dari a...	2017-07-18
90	Azzahra Raih Enam Emas di Kejuaraan Renang ASE...	http://pekanbaru.tribunnews.com/2017/07/18/azz...	6 hari yang lalu ... Perenang andalan Riau Azz...	Laporan Wartawan Tribun Pekanbaru, Palti Siaha...	2017-07-18

Convert the date to datetime format

Removing articles from irrelevant categories - e.g sports, entertainment, travel

Select only articles which appear on the date of the SG press releases and the date thereafter

Translating the text

```
In [181]: pd.options.display.max_rows = 200  
df_idnSG
```

Out[181]:

	title	link	snippet	body	date
83	Hari Ketiga ASEAN School di Singapura, Perenan...	http://pekanbaru.tribunnews.com/2017/07/18/har...	6 hari yang lalu ... Hari ketiga pelaksanaan k...	Laporan Wartawan Tribun Pekanbaru, Palti Siaha...	2017-07-18
84	Mau Terbang ke Singapura, Mathew Thomas Dicega...	http://metropolitan.inilah.com/read/detail/239...	6 hari yang lalu ... Ia menjelaskan, pagi ini ...	INILAHCOM, Jakarta - Selain menetapkan putra...	2017-07-18
85	Timnas U-16 Indonesia Terhindar dari Posisi Ju...	http://kupang.tribunnews.com/2017/07/18/timnas...	6 hari yang lalu ... POS KUPANG.COM - Tim sepa...	POS KUPANG.COM - Tim sepak bola nasional U-16 ...	2017-07-18
86	Mikro Seharusnya Dirjen Pajak Bisa Intip Dan...	http://ekonomi.metrotvnews.com/mikro/RkJPOA3N-...	6 hari yang lalu ... Indonesia Kini Bisa Intip...	Metrotvnews.com, Jakarta: Direktorat Jenderal ...	2017-07-18
87	Sudah Mau Naik Pesawat Menuju Singapura Pagi T...	http://wartakota.tribunnews.com/2017/07/18/sud...	6 hari yang lalu ... Axel Matthew Thomas (19) ...	WARTA KOTA, SEMANGGI - Axel Matthew Thomas (19...	2017-07-18
	Putra Jeremv Thomas				

```
In [209]: df_rel_noJpost
```

Out[209]:

	title	link	snippet	body	date
0	Hari Ketiga ASEAN School di Singapura, Perenan...		ketiga	The third day of ASEAN School championships in...	2017-07-18
1	Mau Terbang ke Singapura, Mathew Thomas Dicega...	http://metropolitan.inilah.com/read/detail/239...	6 hari yang lalu ... Ia menjelaskan, pagi ini ...	In addition to assigning the son of actor Je...	2017-07-18
2	Timnas U-16 Indonesia Terhindar dari Posisi Ju...	http://kupang.tribunnews.com/2017/07/18/timnas...	6 hari yang lalu ... POS KUPANG.COM - Tim sepa...	POS KUPANG.COM - U-16 national football team d...	2017-07-18
3	Mikro Seharusnya Dirjen Pajak Bisa Intip Dan...	http://ekonomi.metrotvnews.com/mikro/RkJPOA3N-...	6 hari yang lalu ... Indonesia Kini Bisa Intip...	Metrotvnews.com, Jakarta: The Directorate Gene...	2017-07-18
4	Sudah Mau Naik Pesawat Menuju Singapura Pagi T...	http://wartakota.tribunnews.com/2017/07/18/sud...	6 hari yang lalu ... Axel Matthew Thomas (19) ...	Warta KOTA, SEMANGGI - Axel Matthew Thomas (19...	2017-07-18

Comparing the docs (1)

SG Press Statements with
'Indonesia' from 01-01-2016
to 24-07-2017

Indonesian News with the
word 'Singapore' from
01-01-2016 to 24-07-2017

82

1

Index	Similarity Score
-------	------------------

1.
----	------

2.
----	------

3.
----	------

.	
---	--

.	
---	--

.	
---	--

.	
---	--

.	
---	--

.	
---	--

127.
------	-------

128.
------	-------

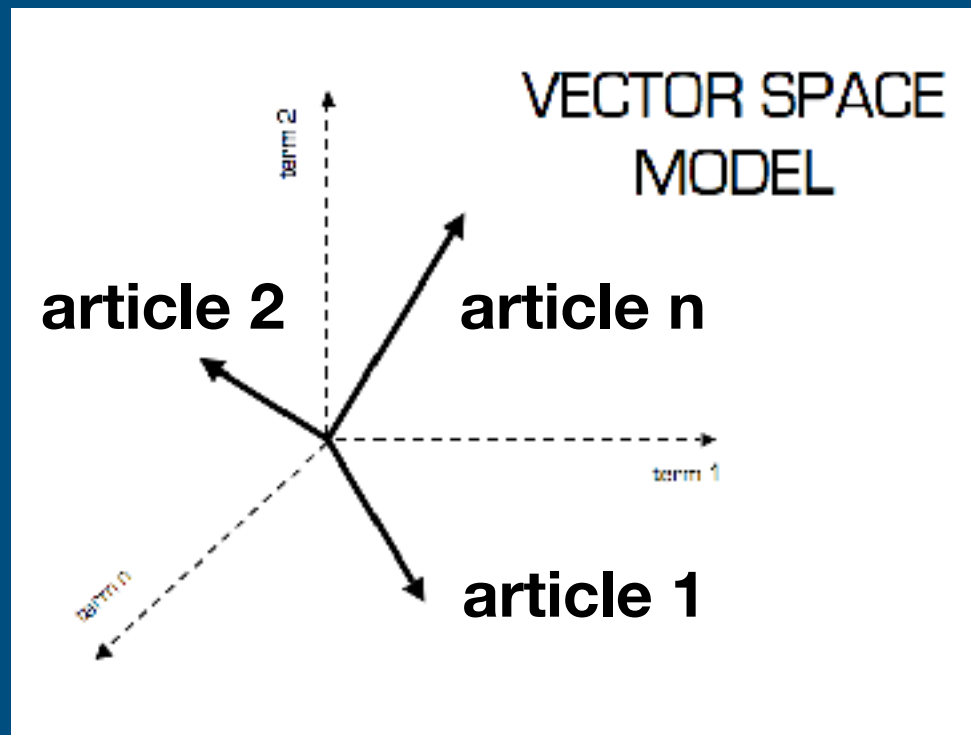
Take one

Compare with
each of the
Indonesian
articles

Calculate Similarity Score

Comparing the docs (2)

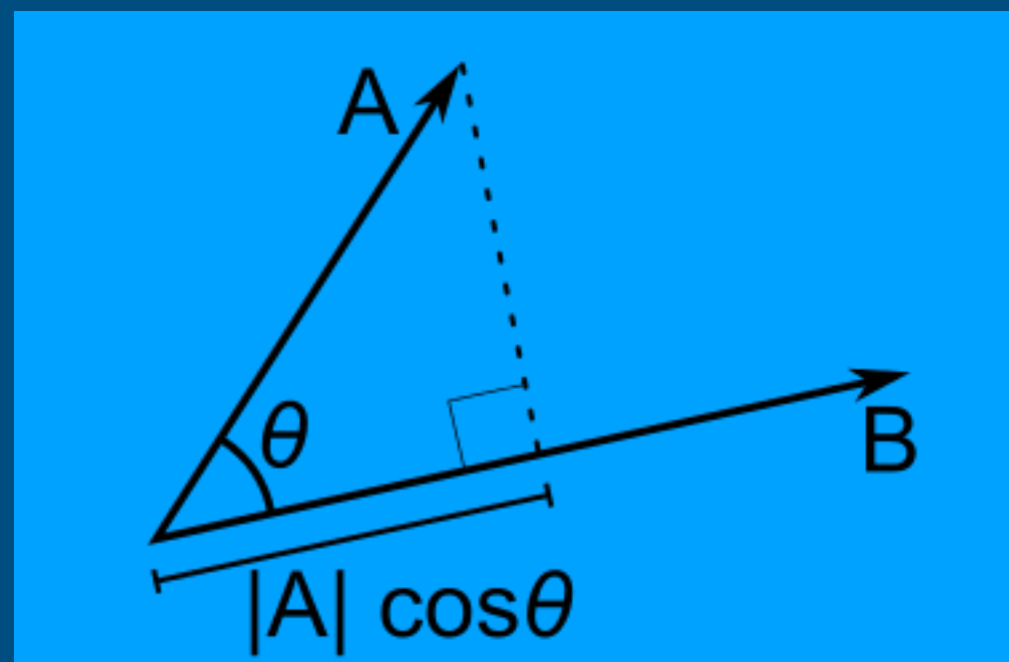
Lesson 12 and 13: Sklearn - CountVectorizer, tf-idf, cosine-similarity



Imagine each article as a vector in N-dim
 $N\text{-dim} = \{\text{number of terms}\}$

Each vector is normalised - i.e magnitude 1

Direction of the vector is determined by the
“term-frequency * inverse document frequency
of each term” of every term in the article



The more similar the article, the
smaller the angle

If angle is 0, $\cos 0 = 1$

To determine “degree of similarity”,
train the data

Results

```
In [8]: df_SGstatement[df_SGstatement.similarity != ''][['date', 'title', 'similarity', 'idnIndex']]
```

Out[8]:

	date	title	similarity	idnIndex
3	2017-07-12	SINGAPORE IS READY TO HAVE AN AEOI RELATIONSHI...	[0.469825612539, 0.292979489258]	[34, 9]
9	2017-05-25	MFA Press Statement: Condolence letters from S...	[0.378049168735, 0.214133953562]	[99, 106]
18	2017-04-25	MFA Press Statement: Visit by Minister for For...	[0.448266158004, 0.315823334478]	[75, 52]
24	2017-03-06	MFA Press Statement: Visit by Deputy Prime Min...	[0.323663971221, 0.23653833858]	[89, 88]
34	2016-12-08	MFA Press Statement: Condolence letters from P...	[0.725894277803, 0.254928964727]	[99, 106]
36	2016-12-07	MFA Spokesman's Comments on the earthquake in ...	[0.427787953723, 0.0584689414219]	[99, 41]
39	2016-11-15	PM Lee Hsien Loong at Joint Press Conference f...	[0.32024721843, 0.316656526725]	[117, 119]
42	2016-11-14	MFA Press Statement: Singapore-Indonesia Leade...	[0.403800207628, 0.347692089295]	[104, 112]

Through training data - Two articles are considered similar when the cosine angle is more than 0.32

Out of the 82 SG Press Statements, only 7 of them were reported before

Learning Points

- A) Similarity testing using Gensim library did not yield useful results - it gave me high similarity scores even though the articles were not similar**
- B) SG Press Statements are not useful tools to influence the Indonesian media**
- C) Natural language processing tools to determine:**
 - (a) the top ten dates on news about Singapore in the last 1.5 years**
 - (b) The top ten Names and Organisations mentioned in the articles**
- D) Steps can be repeated for larger time frames and for other Indonesian news websites that have their own search engines**