Accomplished by python
Data Set:
S_N=["2016-04-07.json", "2016-04-08.json"…]
N = [google, news, time, newyork times]
2016-MM-DD = {"date": "2016-04-10", "news": ["Why Obama Thinks the Senates Inaction Is Dangerous to Democracy", "Donald Trumps Confidence Game Has Been Years In the Making", … …, "Former Saints Player Will Smith Shot and Killed"]}
In "news", there are a number of titles that were collected from news website main page.
Approach 1: K-gram
Approach 2: TF-IDF
Approach 3:

1. Preprocess data; extract all news titles without sign, all lowercase and so on.
2. Filter all stop-words (consist partly from nltk.corpus.stopwords and partly from the result analysis and added arbitrarily.)
3. Process all words in each titles into a dictionary with tuples:
   dict = {("Obama","NOUN"), ("Wins", "VERB")…}
4. For each title, select out all noun and verb words.
5. Compare each two title across two Sets, for each title, if there are k words exists on both set's titles. Those titles will be recognized as similar titles.

Since news titles writing principle determines it must be short, simple but meaningful. So nouns and verbs will be really powerful to indicates if two titles are report the same news. So I choose k = 1, which means if two titles has one or more key words are same, they are similar titles.

The result for 2016-04-10 between three news website is as follows:

	Similar titles amount	Total titles amount	Ratio
New york times & google news	31	56 & 213	55.36%
Time & google news	36	41 & 213	87.80%
New york times & Time	11	56 & 41	26.83%

Approach 4:

1. Preprocess data; extract all news titles without sign, all lowercase and so on.
2. Filter all stop-words (consist partly from nltk.corpus.stopwords and partly from the result analysis and added arbitrarily.)
3. For each title, split titles into words list.
4. For each title words list in set1, compare it with all title word list in set2, if there are more than k words are same, assume them as similar titles.

Here I choose k = 2, which means if two titles has two or more key words are same, they are similar titles.

The result for 2016-04-10 between three news website is as follows:

	Similar titles amount	Total titles amount	Ratio
New york times & google news	31	56 & 213	16.07%
Time & google news	36	41 & 213	41.46%
New york times & Time	11	56 & 41	9.76%

Similar titles # [0, 2, 3, 4, 8, 9, 10, 11, 12, 34, 16, 17, 22, 25, 28, 29, 31](17 in total)
Thus, 41.46% news that reported on time.com was covered on news.google.com.
Detailed result of similar titles is like:

04/10/2016	**www.time.com**	**news.google.com**
1	Donald Trump's Bad Weekend	Donald Trump Loses Weekend Delegate Fight in 5 States
2	Obama: No Political Influence in Clinton Email Probe	Obama: 'No political influence' in Clinton email probe
3	Danny Willett Wins the Masters After Jordan Spieth Collapses	Danny Willett Wins Masters as Jordan Spieth Collapses
		Jordan Spieth Gracious in Defeat After Masters Collapse
		Masters notes: Olympic goals for Willett, tough finishes for DJ, Langer
		Danny Willett takes full advantage of Jordan Spieth's astonishing collapse
		Danny Willett wins
		Willett wins after shocking Spieth collapse
		Winner's Bag: Danny Willett at
		Danny Willett: All you need to know about
4	Brussels Terror Group Planned Another France Attack	Terror suspect Abrini admits he was 'man in hat' at Brussels airport
…	….	…
17	The Panama Papers Love Connection	Malta's opposition demands resignation of PM over Panama Papers
		The Panama Papers prove Mr. Sanders was wrong about a trade pact with Panama
		Panama Papers

Approach 4

	2016-04-08	2016-04-09	2016-04-10	2016-04-11
New york times & google news	16.39 %	20.97%	23.21%	
Time & google news	30.23%	51.52&	12.19 %	
New york times & Time	15.15%	21.21%	9.09%	

Approach 3

	2016-04-08	2016-04-09	2016-04-10	2016-04-11
New york times & google news	9.83 %	11.29%	16.07%	
Time & google news	20.93%	36.37%	7.32%	
New york times & Time	6.06%	12.12%	3.03%	

Approach 2

Approach 1