

# Unsupervised Method Summarizing News Articles

Alex Shum, Robert Corell-Johanns

December 11, 2014

## Introduction

Summarizing a news article is the process of reducing the original document text into a few brief key points. A good summarization should be both concise and retain all the important elements from the original text. There are two main classes of methods for summarization: abstractive and extractive[1]. Abstractive summarization methods generate new text that will paraphrase the article. These techniques seek to create summaries close to what a person might write to generalize an article. By contrast, extractive techniques do not generate any text and will simply extract the most relevant sentences from the article. Extractive methods can only return summaries containing text that was present in the original article. The methods and results presented here will be limited to extractive techniques.

Among extractive techniques there are many approaches to rank sentences and phrases within an article. Common extractive tasks include key phrase detection and sentence extraction – finding the most relevant keyword or sentence. Our goal is to summarize the overall content of a news article and our focus will be implementing an extractive method to find the most relevant sentences to summarize a news article. The method implemented in this project is TextRank; TextRank is an unsupervised method similar to the PageRank algorithm for web pages. Included below is a brief overview of extractive methods, an explanation of the TextRank algorithm, implementation issues, our results, and our conclusions.

## Background

One of the first uses for automatic summarization was to automate finding keywords or key phrases for journal papers. An early solution to this problem was a supervised technique; key phrases were assigned by readers for various collections of documents and a decision tree algorithm was used to classify text as a key phrase or not [3]. This supervised method would look at the length of the text, frequency, if the text contains nouns or verbs and a few other features to train a model to predict if a string of text is a key phrase or not. A decision tree method is able to determine which features are important for key phrases but unfortunately the supervised nature of this method requires both training data and domain knowledge for feature selection.

Another possible approach to automatic summarization is to use latent semantic analysis [4]. The document can be represented as an  $m \times n$  matrix with sentences representing the  $m$  unique word and  $n$  sentences in the document. The columns of this matrix represent the sentences in the document. Distance between sentences are given by the cosine distance:  $d(\vec{s}_1, \vec{s}_2) = \frac{\vec{s}_1 \cdot \vec{s}_2}{|\vec{s}_1||\vec{s}_2|}$ . The most relevant sentences should have the shortest average distance to other sentences. Alternatively, latent semantic analysis utilizes a singular value decomposition and it is also possible to rank sentences based on the highest singular value.

More recently a few graph based methods based on the PageRank algorithm have been proposed for this problem. The PageRank algorithm represents web pages as vertices in a graph with edges representing links between pages. PageRank ranks web pages based on the link structure; important web pages are linked to by other important web pages [6]. This idea is extended to sentences in a document; important sentences are similar to other important sentences. TextRank and LexRank are two algorithms that rank sentences based on this idea [2][5]. These two algorithms represent sentences in a document as vertices and edges based on a distance metric between sentences. LexRank uses unweighted edges between sentences if the two sentences meets some minimum distance threshold using cosine distance. TextRank uses weighted edges based on the shared number of words normalized by the log lengths of the sentences. We chose TextRank based on its unsupervised nature and the absence of parameters to tweak. We give a more detailed description of the TextRank algorithm in the next section.

# Introduction to TextRank

The TextRank method is based on the PageRank algorithm. The original PageRank algorithm is based around the idea of a random web surfer [9]; how likely are they to end up at a particular page by following hyperlinks? The original PageRank algorithm worked on graphs with unweighted edges and the PageRank of a particular page was given by  $rank(v_i) = (1 - d) + d \sum_{j \in in(v_i)} \frac{rank(v_j)}{|out(v_j)|}$  where  $v_i$  and  $v_j$  represent web pages,  $in(v_j)$  represents the set of vertices that links to page  $i$ ,  $out(v_j)$  represents the pages that link from page  $j$ , and  $d$  represents the dampening factor. The dampening factor represents the web surfer jumping to a random page.

TextRank represents sentences as vertices and edges between sentences is weighted by the following distance function:  $d(s_i, s_j) = \frac{|w_k | w_k \in s_i \& w_k \in s_j |}{\log(|s_i|) + \log(|s_j|)}$ .  $s_i$  and  $s_j$  represent sentence  $i$  and sentence  $j$ . Each sentence is a set of sets and the distance metric is the cardinality of the intersection between two sentence sets divided by the sum of the logs of the sentence lengths. This distance function for edge weights introduces a separate issue: the original PageRank algorithm is for unweighted graphs. In order to get around this issue a weighted variation of PageRank is used:  $rank(s_i) = (1 - d) + d \sum_{j \in In(s_i)} \frac{w_{ji} rank(s_j)}{\sum_{s_k \in out(s_j)} rank(s_k)}$ .  $s_i$  and  $s_j$  represent sentence  $i$  and sentence  $j$  and  $w_{ji}$  represents the edge weight between sentence  $j$  and  $i$ . It turns out with this slight modification we can iteratively calculate the page rank of sentences in a document.

## Implementation and Java Code

Before running the TextRank algorithm some basic text preprocessing was done. Punctuation was stripped from the text and stop words, words with little semantic meaning, were removed from the text using the Lucene library [7]. The text was parsed into individual sentences using the openNLP library. The adjacency matrix was calculated between all sentences. The adjacency matrix was normalized so that the columns sum to one and then the weighted pagerank is calculated iteratively.

Our Java implementation of TextRank is split into seven java class files. The most important file is the SummaryBot.java file that defines a SummaryBot object. A SummaryBot is meant to take in a string or a text file containing an article and summarize it down to just a few sentences. To do this it first extracts the text from the file and then sends it to a ISentenceProcessor object. The SentenceProcessor object is to separate the text into sentences and remove unnecessary or irrelevant words. Initially sentence processing was done solely in openNLP with OpenNLPSentenceProcessor.java. However, openNLP does not have stop word removal and we used part-of-speech tagging to accomplish the same task. Due to computation time and better stop word removal we implemented the sentence processing with both libraries and this is contained in LuceneSentenceProcessor.java.

After being processed, the sentences are passed to the CreateCommonMatrix method in MiscUtils.java which creates an adjacency matrix  $A$  where  $a_{i,j}$  represents the number of words shared between sentences  $i$  and  $j$ . Then the matrix is sent to the IRanker object which uses some method to rank the sentences based on values in the matrix. We implemented using the PageRanker class which iteratively calculates the weighted page rank to accomplish this task. Finally, with the results from the Ranker object, the SummaryBot is able to print the best sentences for summarizing the article. All of the java source code can be found at our repository: <http://www.github.com/ALShum/textsummary>.

As a final heuristic, SummaryBot returns the top sentences in the same order that they appear in the original article text. When sentences were returned based strictly on highest PageRank, it was quite common for the top sentence to make a reference to something that appeared earlier in the article. By returning sentences in the same order as the text in the article it increases the likelihood that the summary sentences will follow the same logical ordering as the original text.

## Results

Since this is not a supervised algorithm testing the results was not straightforward. We did not have the resources to compare TextRank's results with human results. There are a few numerical measures used to evaluation performance based on N-gram statistics. Our goal was to have more interpretable results and we compare our implementation of TextRank against "the competition" – other available text summarizers. The following is a list of summarizers used:

1. Java TextRank implementation.

2. <http://freesummarizer.com/>
3. <http://autosummarizer.com/>
4. <http://textcompactor.com/>

For evaluation purposes we randomly choose the following six news articles from various sources:

1. <http://www.reuters.com/article/2014/12/04/us-vatican-economy-pellidUSKCN0JI1CG20141204?feedType=RSS&feedName=worldNews>
2. <http://thinkprogress.org/climate/2014/12/04/3599528/cruise-lines-sewage-2014/>
3. [http://articles.economictimes.indiatimes.com/2014-12-02/news/56648982\\_1\\_syria-british-man-jihadis](http://articles.economictimes.indiatimes.com/2014-12-02/news/56648982_1_syria-british-man-jihadis)
4. <http://www.thelocal.de/20141204/merkel-speaks-out-against-net-neutrality>
5. <http://www.bbc.com/news/world-europe-30322198>
6. <http://www.aljazeera.com/news/middleeast/2014/11/us-led-air-raids-target-syria-rebel-groups-20141116123.html>

We use our implementation of TextRank against other online text summarizers for the articles mentioned above. We check the similarity in terms of the number of matched sentences between the above summarizers and TextRank. Additionally we try to identify which summary includes the most context and main points from the main article. Complete results can be found at our repository: <http://www.github.com/ALShum/textsummary/blob/master/Results.txt>. The following is a summary of the results. Summarizer 1 refers to our Java implementation of TextRank and summarizers 2, 3, 4 refer respectively to the online summarizers listed above. Articles 1 through 6 refer to the articles listed above respectively. For TextRank we always return the top 5 sentences that summarize the entire news article. The online summarizers return a varying number of sentences depending on the news article. For context purposes we include the first three paragraphs of each news article as well as the summaries from the respective summarizers.

## Article 1

The Vatican's economy minister has said hundreds of millions of euros were found "tucked away" in accounts of various Holy See departments without having appeared in the city-state's balance sheets.

In an article for Britain's Catholic Herald Magazine to be published on Friday, Australian Cardinal George Pell wrote that the discovery meant overall Vatican finances were in better shape than previously believed.

"In fact, we have discovered that the situation is much healthier than it seemed, because some hundreds of millions of euros were tucked away in particular sectional accounts and did not appear on the balance sheet," he wrote.

## Summarizer 1

- The Vatican's economy minister has said hundreds of millions of euros were found "tucked away" in accounts of various Holy See departments without having appeared in the city-state's balance sheets.
- In an article for Britain's Catholic Herald Magazine to be published on Friday, Australian Cardinal George Pell wrote that the discovery meant overall Vatican finances were in better shape than previously believed.
- "It is important to point out that the Vatican is not broke ... the Holy See is paying its way, while possessing substantial assets and investments," Pell said, according to an advance text made available on Thursday.
- Pell did not suggest any wrongdoing but said Vatican departments had long had "an almost free hand" with their finances and followed "long-established patterns" in managing their affairs.
- In his article, Pell said the reforms were "well under way and already past the point where the Vatican could return to the 'bad old days'."

### Summarizer 2

- The Vatican's economy minister has said hundreds of millions of euros were found "tucked away" in accounts of various Holy See departments without having appeared in the city-state's balance sheets.
- Pell's office sent a letter to all Vatican departments last month about changes in economic ethics and accountability.
- Each department will have to enact "sound and efficient financial management policies" and prepare financial information and reports that meet international accounting standards.

### Summarizer 3

- The Vatican's economy minister has said hundreds of millions of euros were found tucked away in accounts of various Holy See departments without having appeared in the city-states balance sheets.
- Pell did not suggest any wrongdoing but said Vatican departments had long had an almost free hand with their finances and followed long-established patterns in managing their affairs.
- Pell's office sent a letter to all Vatican departments last month about changes in economic ethics and accountability.
- As of Jan. 1, each department will have to enact sound and efficient financial management policies and prepare financial information and reports that meet international accounting standards.

### Summarizer 4

- "The Holy See is paying its way, while possessing substantial assets and investments," Pell said, according to an advance text made available on Thursday.
- Pell did not suggest any wrongdoing but said Vatican departments had long had "an almost free hand" with their finances and followed "long-established patterns" in managing their affairs.
- "Very few were tempted to tell the outside world what was happening, except when they needed extra help," he said, singling out the once-powerful Secretariat of State as one department that had especially jealously guarded its independence.
- "It was impossible for anyone to know accurately what was going on overall," said Pell, head of the new Secretariat for the Economy that is independent of the now downgraded Secretariat of State.
- Pell is an outsider from the English-speaking world transferred by Pope Francis from Sydney to Rome to oversee the Vatican's often muddled finances after decades of control by Italians.

### Overall Results

TextRank was able to match one sentence with summarizers 2, two with summarizer 3, and one from summarizer 4. However, even though TextRank only matched a few sentences, the summary generated still has the important main points of the article. In fact TextRank provided a bit more context than the other summarizers in including the second sentence identifying George Pell, author of a relevant article for a Catholic magazine. The other summarizers mention Pell but miss the earlier references that introduce this source in the article.

### Article 2

Cruise ships released over one billion gallons of sewage into the ocean in 2014, according to a new report from Friends of the Earth.

The analysis, which worked off of federal data, did show that some of the 16 cruise lines assessed are slowly becoming more environmentally friendly. But according to the press release from environmental group Friends of the Earth (FOE), over 40 percent of the 167 ships in operation still operate using waste treatment technology that's more than 35 years old. "Such antiquated treatment systems leave harmful levels of fecal matter, bacteria, heavy metals and other contaminants in the water," FOE noted.

Federal law requires that cruise ships only dump treated wastewater if they are within three nautical miles of shore. But beyond that point, it's essentially a free-for-all.

### Summarizer 1

- But according to the press release from environmental group Friends of the Earth (FOE), over 40 percent of the 167 ships in operation still operate using waste treatment technology that's more than 35 years old.
- This is an industry worth billions of dollars that could install the most advanced sewage treatment and air pollution reduction technology available, said Marcie Keever, the oceans and vessels program director at FOE. We're encouraged that some cruise lines are taking incremental steps to improve their performance, but the entire industry must stop hiding behind weak regulations and take action to make sure the oceans their ships travel remain as clear as the photos in cruise brochures.
- But for the 2014 report, all 16 cruise lines apparently refused to respond to FOE's request for information on their pollution reduction efforts forcing the group to rely on federal data.
- Outside of the transparency issue, Disney Cruise Lines is apparently the most environmentally friendly, scoring an A for sewage treatment, a B- for air pollution reduction, and an A for water quality compliance.
- According to FOE, data from EPA also shows that each day an average cruise ship is at sea it emits more sulfur dioxide than 13 million cars and more soot than one million cars.

### Summarizer 2

- FOE also cites data from the Environmental Protection Agency (EPA) which shows an average cruise ship with 3,000 passengers and crew produces about 21,000 gallons of sewage a day enough to fill 10 backyard swimming pools in a week.
- Up through 2013, the report card had graded the cruise lines on three metrics: the quality of their sewage treatment technology, how much they reduce their air pollution through the use of cleaner fuels and plugging into onshore power, and how well they comply with water quality standards especially those established by the State of Alaska to protect its coasts.
- But for the 2014 report, all 16 cruise lines apparently refused to respond to FOE's request for information on their pollution reduction efforts forcing the group to rely on federal data.
- Outside of the transparency issue, Disney Cruise Lines is apparently the most environmentally friendly, scoring an A for sewage treatment, a B- for air pollution reduction, and an A for water quality compliance.

### Summarizer 3

- FOE also cites data from the Environmental Protection Agency which shows an average cruise ship with 3,000 passengers and crew produces about 21,000 gallons of sewage a day enough to fill 10 backyard swimming pools in a week.
- Were encouraged that some cruise lines are taking incremental steps to improve their performance, but the entire industry must stop hiding behind weak regulations and take action to make sure the oceans their ships travel remain as clear as the photos in cruise brochures.
- The centerpiece of FOEs analysis was their Cruise Ship Report Card, which has come out every year since 2009.
- Up through 2013, the report card had graded the cruise lines on three metrics: the quality of their sewage treatment technology, how much they reduce their air pollution through the use of cleaner fuels and plugging into onshore power, and how well they comply with water quality standards especially those established by the State of Alaska to protect its coasts.
- Outside of the transparency issue, Disney Cruise Lines is apparently the most environmentally friendly, scoring an A for sewage treatment, a B- for air pollution reduction, and an A for water quality compliance.

#### Summarizer 4

- Cruise ships released over one billion gallons of sewage into the ocean in 2014, according to a new report from Friends of the Earth.
- The analysis, which worked off of federal data, did show that some of the 16 cruise lines assessed are slowly becoming more environmentally friendly.
- “We’re encouraged that some cruise lines are taking incremental steps to improve their performance, but the entire industry must stop hiding behind weak regulations and take action to make sure the oceans their ships travel remain as clear as the photos in cruise brochures.”
- The centerpiece of FOE’s analysis was their Cruise Ship Report Card, which has come out every year since 2009.
- Up through 2013, the report card had graded the cruise lines on three metrics: the quality of their sewage treatment technology, how much they reduce their air pollution through the use of cleaner fuels and plugging into onshore power, and how well they comply with water quality standards - especially those established by the State of Alaska to protect its coasts.

#### Overall Results

TextRank identified two of the same sentences as summarizer 2, one with summarizer 3, and no sentences from summarizer 4. Although the summarizer includes some of the main points of the article including mention of the EPA reports, it misses the overall summary sentence. Summarizer 4’s first sentence is a good overview of the article that the other summarizers miss. However, TextRank includes context about the “Friends of Earth (FOE)”. The other summarizers mention FOE in their summary sentences but without earlier context a reader would not understand the meaning of this term.

#### Article 3

Jihadis, who travelled to Iraq and Syria to join the Islamic State (IS) are begging to be allowed to return home, complaining they were forced to do menial tasks, media reported Tuesday.

Foreign recruits have been writing to their parents with feeble excuses that range from having to clean toilets and to their iPod was no longer working or that it was getting too cold, the media reported.

A series of weepy messages leaked to the French daily Le Figaro revealed a number of young French Muslim converts are having second thoughts about signing up to IS.

#### Summarizer 1

- A third fighter said he was “sick” of his time with the militant group, adding: “They make me do the washing up.”
- Tanvir Sheikh, father of one of Majeed’s friends who was still missing in Iraq, said he felt betrayed by his son.
- Sheikh said his son Fahad had a job offer in Kuwait but instead he decided to travel to Iraq to join the extremists.
- It came days after a court heard how a British man was jailed for conspiring to attend a terror training camp in Syria after he returned home after whining to his girlfriend about his treatment abroad.
- Radicalised foreigners have been drawn to IS, which has conducted a series of mass executions and other atrocities since launching its offensive in Iraq and Syria in June, the report said.

#### Summarizer 2

- Jihadis, who travelled to Iraq and Syria to join the Islamic State (IS) are begging to be allowed to return home, complaining they were forced to do menial tasks, media reported Tuesday.
- Foreign recruits have been writing to their parents with feeble excuses that range from having to clean toilets and to their iPod was no longer working or that it was getting too cold, the media reported.
- A series of weepy messages leaked to the French daily Le Figaro revealed a number of young French Muslim converts are having second thoughts about signing up to IS.

- “I’m fed up to the back teeth. My iPod no longer works out here. I have got to come home,” one of the foreign IS recruits was quoted as saying.

### Summarizer 3

- Jihadis, who travelled to Iraq and Syria to join the Islamic State are begging to be allowed to return home, complaining they were forced to do menial tasks, media reported Tuesday.
- Dozens of other recruits, who are reportedly now working with French lawyers, have collected texts and messages that suggest the jihadis felt cheated into joining IS.
- One lawyer said that the longer they remain in Iraq and Syria, the more chance they have of becoming time bombs when they return.
- The report also mentioned the Indian national, Majeed, who went to Syria to fight for the militant group and was later arrested by the National Investigation Agency for terror-related offences when he arrived back home.
- The brothers brought back bullets and took the pictures on their mobile phones to remember their time in the jihadi camp.

### Summarizer 4

- One lawyer said that the longer they remain in Iraq and Syria, the more chance they have of becoming “time bombs” when they return.
- The report also mentioned the Indian national, Majeed, who went to Syria to fight for the militant group and was later arrested by the National Investigation Agency (NIA) for terror-related offences when he arrived back home.
- Majeed told NIA officers he was sidelined by the jihadis for whom he fetched water and performed other lowly tasks such as cleaning toilets, instead of taking part in the deadly offensive like he wanted.
- Al Qaeda announced in September a new chapter of its extremist movement charged with waging jihad in South Asia.
- Tanvir Sheikh, father of one of Majeed’s friends who was still missing in Iraq, said he felt betrayed by his son.

### Overall Results

TextRank identified zero of the same sentences as summarizers 2 and 3 and identified one sentence in common with summarizer 4. TextRank missed the first summary sentence from summarizer 2 and 3 but includes some of the examples of why foreigners are regretting joining ISIS. Overall this article was relatively difficult to summarize and both TextRank and summarizer 4 did fairly poor on this article.

### Article 4

Merkel said that some key services for the digital economy would require reliable transmission quality and should therefore be treated differently than other data.

At the Vodafone-hosted Digitising Europe conference in Berlin, she called for a splitting of services, “one for free internet, and the other for special services”, adding that it was up to Brussels to negotiate how it would work.

“An innovation-friendly internet means that there is a guaranteed reliability for special services,” she said. “These can only develop when predictable quality standards are available”.

### Summarizer 1

- Merkel said that some key services for the digital economy would require reliable transmission quality and should therefore be treated differently than other data.
- Social Democratic Party (SPD) MEP Petra Kammerevert told The Local that this would make it more difficult to find a common European position on net neutrality.

- “If Merkel goes into negotiations with the position she’s outlined today, it will be very difficult for the European Council to find a common position,” she said.
- The Council, composed of the heads of state and government of all the EU member countries, must find a common negotiating position to deal with the European Parliament, which voted against a European Commission plan for regulations that would allow a two-tier internet in April.
- “This is not the net neutrality we want, but a move towards the creation of a two-tier network where content becomes preferred based on who pays for it to make it so”, said Markus Beckedahl, Berlin-based founder of netzpolitik.

## Summarizer 2

- Merkel said that some key services for the digital economy would require reliable transmission quality and should therefore be treated differently than other data.
- At the Vodafone-hosted Digitising Europe conference in Berlin, she called for a splitting of services, “one for free internet, and the other for special services”, adding that it was up to Brussels to negotiate how it would work.
- Merkel added that these special services would run over existing internet infrastructure.
- Social Democratic Party (SPD) MEP Petra Kammerevert told The Local that this would make it more difficult to find a common European position on net neutrality.
- “If Merkel goes into negotiations with the position she’s outlined today, it will be very difficult for the European Council to find a common position,” she said.

## Summarizer 3

- At the Vodafone-hosted Digitising Europe conference in Berlin, she called for a splitting of services, one for free internet, and the other for special services , adding that it was up to Brussels to negotiate how it would work.
- Social Democratic Party MEP Petra Kammerevert told The Local that this would make it more difficult to find a common European position on net neutrality.
- If Merkel goes into negotiations with the position shes outlined today, it will be very difficult for the European Council to find a common position, she said.
- The Council, composed of the heads of state and government of all the EU member countries, must find a common negotiating position to deal with the European Parliament, which voted against a European Commission plan for regulations that would allow a two-tier internet in April.
- If you poke holes into net neutrality the way Chancellor Merkel suggests, then its no longer democratic, he said.

## Summarizer 4

- Merkel said that some key services for the digital economy would require reliable transmission quality and should therefore be treated differently than other data.
- At the Vodafone-hosted Digitising Europe conference in Berlin, she called for a splitting of services, “one for free internet, and the other for special services”, adding that it was up to Brussels to negotiate how it would work.
- “An innovation-friendly internet means that there is a guaranteed reliability for special services,” she said. ”These can only develop when predictable quality standards are available”.
- Merkel added that these special services would run over existing internet infrastructure.
- Social Democratic Party (SPD) MEP Petra Kammerevert told The Local that this would make it more difficult to find a common European position on net neutrality.

## Overall Results

TextRank was able to match 3 of the sentences from summarizer 2 and two of the sentences from summarizer 3 and 4.



## Article 5

President Vladimir Putin has warned Russians of hard times ahead and urged self-reliance, in his annual state-of-the nation address to parliament.

Russia has been hit hard by falling oil prices and by Western sanctions imposed in response to its interventions in the crisis in neighbouring Ukraine.

The rouble, once a symbol of stability under Mr Putin, suffered its biggest one-day decline since 1998 on Monday.

### Summarizer 1

- He insisted the "tragedy" in Ukraine's south-east had proved that Russian policy had been right but said Russia would respect its neighbour as a brotherly country.
- It is not our design or desire that we see a Russia isolated through its own actions," Mr Kerry said.
- So it's telling that Mr Putin chose to stress his unwavering hard line on the crisis in Ukraine: what happened in Kiev was an "illegal coup" and Crimea, which Russia annexed, is like "holy land" for Russia and will always be treated that way.
- Stressing that Russia remained "open for the world", Mr Putin suggested Western sanctions should be seen as a stimulus.
- Mr Putin said: "There is no doubt they would have loved to see the Yugoslavia scenario of collapse and dismemberment for us - with all the tragic consequences it would have for the peoples of Russia.

### Summarizer 2

- Russia has been hit hard by falling oil prices and by Western sanctions imposed in response to its interventions in the crisis in neighbouring Ukraine.
- Speaking to both chambers in the Kremlin, Mr Putin also accused Western governments of seeking to raise a new "iron curtain" around Russia.
- So it's telling that Mr Putin chose to stress his unwavering hard line on the crisis in Ukraine: what happened in Kiev was an "illegal coup" and Crimea, which Russia annexed, is like "holy land" for Russia and will always be treated that way.
- Vladimir Putin again accused the West of meddling in Russia's internal affairs and using sanctions to "contain" the country as it grew stronger and more independent.
- Russians are believed to have taken more than 100bn out of the country this year and Mr Putin promised an amnesty for anyone choosing to bring their money back.

### Summarizer 3

- He insisted the tragedy in Ukraines south-east had proved that Russian policy had been right but said Russia would respect its neighbour as a brotherly country.
- So its telling that Mr Putin chose to stress his unwavering hard line on the crisis in Ukraine: what happened in Kiev was an illegal coup and Crimea, which Russia annexed, is like holy land for Russia and will always be treated that way.
- Vladimir Putin again accused the West of meddling in Russias internal affairs and using sanctions to contain the country as it grew stronger and more independent.
- Russians are believed to have taken more than 100bn out of the country this year and Mr Putin promised an amnesty for anyone choosing to bring their money back.
- Stressing that Russia remained open for the world , Mr Putin suggested Western sanctions should be seen as a stimulus.

#### Summarizer 4

- Russia has been hit hard by falling oil prices and by Western sanctions imposed in response to its interventions in the crisis in neighbouring Ukraine.
- The rouble, once a symbol of stability under Mr Putin, suffered its biggest one-day decline since 1998 on Monday.
- The government has warned that Russia will fall into recession next year.
- Speaking to both chambers in the Kremlin, Mr Putin also accused Western governments of seeking to raise a new “iron curtain” around Russia.
- He expressed no regrets for annexing Ukraine’s Crimea peninsula, saying the territory had a “sacred meaning” for Russia.

#### Overall Results

TextRank matches one sentence with summarizer 2, three sentences with summarizer 3 and no sentences with summarizer 4.

#### Article 6

US-led air raids have struck Syrian rebels not linked to the Islamic State of Iraq and the Levant (ISIL), expanding the coalitions raids for the second time to other groups fighting in Syria.

Activists in Syria told Al Jazeera the raids struck the Ahrar al-Sham rebel group in the northern province of Idlib, and the Syrian branch of al-Qaeda, the Nusra Front, in western Aleppo on Thursday.

At least six coalition air raids struck Idlib overnight, with one targeting an Ahrar al-Sham position in the village of Babska.

#### Summarizer 1

- US-led air raids have struck Syrian rebels not linked to the Islamic State of Iraq and the Levant (ISIL), expanding the coalitions raids for the second time to other groups fighting in Syria.
- Activists in Syria told Al Jazeera the raids struck the Ahrar al-Sham rebel group in the northern province of Idlib, and the Syrian branch of al-Qaeda, the Nusra Front, in western Aleppo on Thursday.
- Air raids also targeted the Nusra Front in the Reef al-Muhandisin area of western Aleppo, reportedly killing several of the groups’ fighters as well as six civilians.
- The US and Arab states have been carrying out air strikes against ISIL in Syria since September 23, in a bid to prevent the armed group from seizing more territory in the conflict-hit nation.
- There have been no reported attacks on groups outside ISIL since, but the latest air raids come after the group made gains against Western-backed rebel fighters in the Idlib region.

#### Summarizer 2

- US-led air raids have struck Syrian rebels not linked to the Islamic State of Iraq and the Levant (ISIL), expanding the coalitions raids for the second time to other groups fighting in Syria.
- Activists in Syria told Al Jazeera the raids struck the Ahrar al-Sham rebel group in the northern province of Idlib, and the Syrian branch of al-Qaeda, the Nusra Front, in western Aleppo on Thursday.
- One of Syria’s largest armed opposition groups, Ahrar al-Sham is a key member of the Islamic Front coalition, which seeks a state run on Islamic principles as its fights to topple Syrian President Bashar al-Assad.
- There have been no reported attacks on groups outside ISIL since, but the latest air raids come after the group made gains against Western-backed rebel fighters in the Idlib region.

### Summarizer 3

- US-led air raids have struck Syrian rebels not linked to the Islamic State of Iraq and the Levant, expanding the coalitions raids for the second time to other groups fighting in Syria.
- Activists in Syria told Al Jazeera the raids struck the Ahrar al-Sham rebel group in the northern province of Idlib, and the Syrian branch of al-Qaeda, the Nusra Front, in western Aleppo on Thursday.
- Air raids also targeted the Nusra Front in the Reef al-Muhandisin area of western Aleppo, reportedly killing several of the groups fighters as well as six civilians.
- There have been no reported attacks on groups outside ISIL since, but the latest air raids come after the group made gains against Western-backed rebel fighters in the Idlib region.

### Summarizer 4

- US-led air raids have struck Syrian rebels not linked to the Islamic State of Iraq and the Levant (ISIL), expanding the coalitions raids for the second time to other groups fighting in Syria.
- Activists in Syria told Al Jazeera the raids struck the Ahrar al-Sham rebel group in the northern province of Idlib, and the Syrian branch of al-Qaeda, the Nusra Front, in western Aleppo on Thursday.
- At least six coalition air raids struck Idlib overnight, with one targeting an Ahrar al-Sham position in the village of Babska.
- One of Syria's largest armed opposition groups, Ahrar al-Sham is a key member of the Islamic Front coalition, which seeks a state run on Islamic principles as its fights to topple Syrian President Bashar al-Assad.
- Air raids also targeted the Nusra Front in the Reef al-Muhandisin area of western Aleppo, reportedly killing several of the groups' fighters as well as six civilians.

### Overall Results

TextRank matches three sentences with summarizer 2 and 4 and four sentences with summarizer 3. All the summarizers did a fair job of summarizing this particular article.

## Conclusion

Overall TextRank's results were comparable to the other summarizers most of the time. TextRank was able to include some context that the other summarizers missed. For some of the more difficult results TextRank had comparable results to the other online summarizers. In some cases where TextRank missed important information in the summary it was still able to match a large number of sentences that the other summarizers identified. These are promising results as our implementation of TextRank uses very few heuristics. The only heuristic we employed was the ordering of the returned sentences; we did not employ heuristics for sentence position or paragraph position as we suspect the online summarizers used. Possible future work would involve comparing the results of difference weighting functions and sentence position heuristics. Additionally news article titles are important overviews of the article and a possible extension would be to give some weight or prior to text contained in an article's title.

## Sources

1. Text Summarization: <http://pages.cs.wisc.edu/~goldberg/publications/summarization.pdf>
2. TextRank: <http://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>
3. Supervised text summarization: <http://www.extractor.com/IR2000.pdf>
4. Latent semantic analysis for text summarization: <http://www.kiv.zcu.cz/~jstein/publikace/isim2004.pdf>
5. LexRank: <http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume22/erkan04a-html/erkan04a.html>
6. PageRank: <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>

7. Lucene Java Library: <http://lucene.apache.org/>
8. OpenNLP Java Library: <https://opennlp.apache.org/>