Natural Language Processing: Temporal Tagging, Document Summarization

22 February 2022

Temporal & Event Information Extraction

Is time important?

News articles

Narrative documents

Biographies

Temporal information plays an important role in many types of text documents

Two tasks of temporal taggers

1. Extraction of temporal expressions

Greece Makes 'Good Progress' in Payment Talks

By Maria Petrakis and Natalie   Sep 20, 2011 10:39 PM GMT+0200   ➕ ADD TO QUEUE

👍 Recommend 10
🐦 Tweet 33
in Share 4
+1 2
🔍 Enlarge image

Evangelos Venizelos, Greece's finance minister, listens to an aide during a

Sept. 20, 2011

Tsironis/Bloomberg

Greek Finance Minister Evangelos Venizelos made "good progress" in a second round of talks with the European Union and International Monetary Fund aimed at staving off default, the EU said.

The telephone me **tonight,** ch follow **yesterday,** were intended to damp concerns that Gree reduction targets and to clear the way for a sixth installment of rescue funds. The EU statement said a "full mission" will return to At **next week** liscussions in coming days at the IMF's annual meeting in Washington.

Staying in the euro area is an "irreversible and fundamental national choice," Venizelos said in a sta **earlier today,** wledge that our fiscal data and economic structures are a problem for the euro determined to tackle once and for all."

The EU **next month** t payment for Greece is on track to be relea inister George Papandreou fights

Different types of temporal expressions

Temporal markup language TimeML defines 4 types:
[Pustejovsky et al. 2005] (http://timeml.org/)

• Dates
→ June 24, 2013
→ September 2000
→ two weeks ago

Durations
→ two weeks
→ 12.5 hours
→ several months

• Times
→ 3 p.m.
→ yesterday morning
→ 2012-06-28T16:25

Sets
→ every day
→ annually
→ twice a month

Two tasks of temporal taggers

2. Normalization of extracted temporal expressions

Sept. 20, 2011 → **2011-09-20**

next month → **2011-10**

normalization of relative and underspecified expressions

Different realizations of temporal expressions

- explicit
  - → June 24, 2013
  - → the 20th century
  - → easy to normalize
- implicit
  - → Christmas 2012
  - → Columbus Day 2006
  - → additional knowledge

- relative
  - → two weeks ago
  - → yesterday
  - → reference time
- underspecified
  - → Monday
  - → June 24
  - → reference time and etc.

Normalization

- Temporal information can be normalized: expressions with same semantics → same value
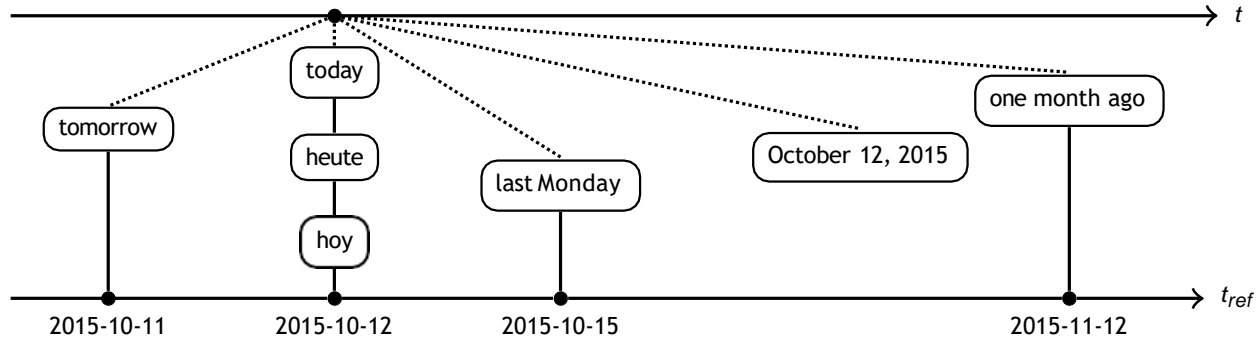
Hierarchical organization

- Temporal information can be organized hierarchically: expressions of different granularities

News articles: Doc timestamp used as reference time

Document Creation Time: 2000-12-26
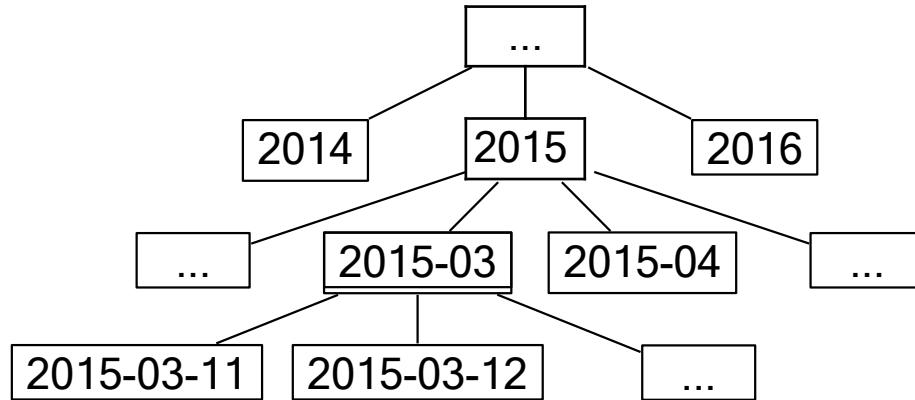. . . On Thursday , the Census Bureau will publish the official population count for the United States, including the state-by-state totals required under the Constitution to determine how many seats each state is allocated in the House. The figures, eagerly awaited by many state government officials, are the first in a wave of releases of demographic data based on the 2000 census. . . .
Population estimates issued periodically by the Census Bureau indicate that as of October , 275,843,000 people were living in . . .
Additional seats are then assigned to each state based on a person-to-House-member ratio that changes every 10 years because the country's population keeps growing . . .

Narrative documents: reference time needs to be found in text

Document Creation Time: 2009-12-19

1979 : Soviet invasion

On December 7, 1979 , Soviet informants to the Afghan Armed Forces …and began to land in Kabul on December 25 .

On December 27, 1979 , 700 Soviet troops dressed in Afghan uniforms, …That operation began at 19:00 hr. , …The operation was fully complete by the morning of December 28, 1979 . … According to the Soviet Politburo they were complying with the 1978 Treaty of Friendship, …........ Soviet ground forces, under the command of Marshal Sergei Sokolov, entered Afghanistan from the north on December 27 . In the morning , the 103rd Guards 'Vitebsk' …

cc Jannik Strötgen - ATIR-07

Colloquial Texts: timestamp as reference time, missing context and non-standard language

SMS 2010-01-10T05:19
Whats it u wanted 2 say last nite ?

SMS 2010-09-23T09:50
Yo! Rem to come for lab tmr :-) ...

SMS 2011-02-16T12:42
... andy is availableat 10 am in his office

Extracting Events and their Times

- Event extraction: identify mentions of events in text
  - An event mention is any expression denoting an event or state that can be assigned to a particular point, or interval, in time
  - Events are to be classified as actions, states, reporting events (say, report, tell, explain), perception events, and so on

[EVENT Citing] high fuel prices, United Airlines [EVENT said] Friday it has [EVENT increased] fares by $6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR Corp., immediately [EVENT matched] [EVENT the move], spokesman Tim Wagner [EVENT said]. United, a unit of UAL Corp., [EVENT said] [EVENT the increase] took effect Thursday and [EVENT applies] to most routes where it [EVENT competes] against discount carriers, such as Chicago to Dallas and Denver to San Francisco.

Extracting Events and their Times

- Temporal Ordering of Events
  - impose a partial ordering on the events and temporal expressions mentioned in a text (Allen relations)
- TimeBank
  - corpus consists of text annotated with temporal information (Pustejovsky et al., 2003b)
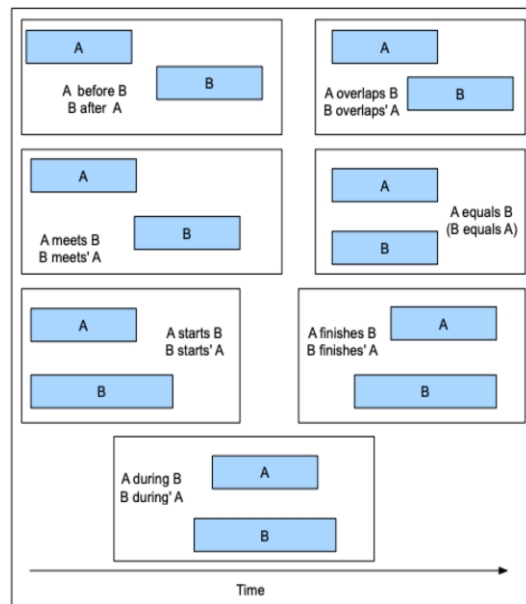
Most research focused on English

- Manually annotated corpora, e.g.,
  - TimeBank [Pustejovsky et al. 2003]
- Research competitions, e.g.,
  - TempEval series e.g., [UzZaman et al. 2013]

- Temporal taggers, e.g.,

  GUTime [Verhagen et al. 2005]

  SUTime [Chang & Manning 2012, 2013]

  HeidelTime [Strötgen & Gertz 2010, 2013; Strötgen et al. 2013]

  ClearTK-TimeML with Timenorm [Bethard 2013]

Ex: Temporal Expressions Analyzed en masse

Future

Past

(b) Snapshot of US dataset available online [44]
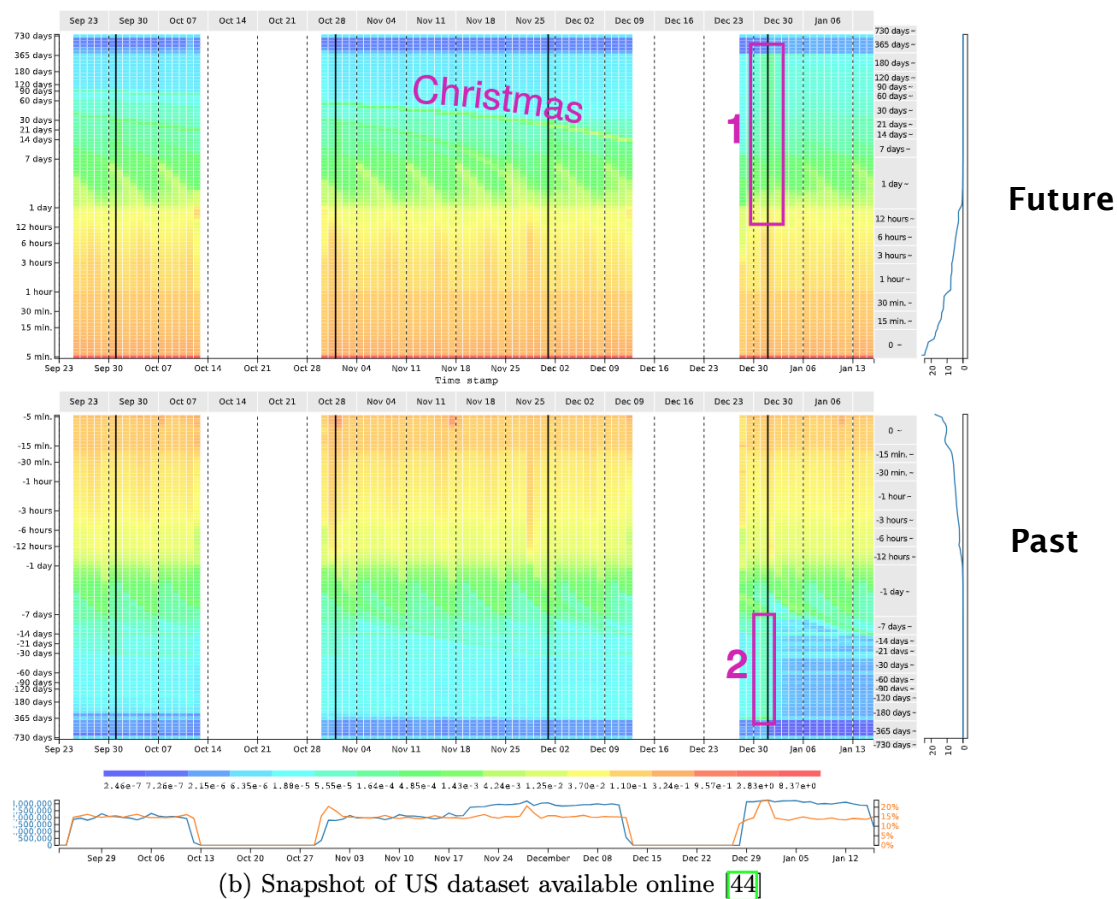
Distribution of temporal expressions in large amount of tweets from US mapped on axis y in relation to tweet timestamp (axis x)

A. Jatowt, É. Antoine, Y. Kawai, T. Akiyama: Mapping Temporal Horizons: Analysis of Collective Future and Past related Attention in Twitter. WWW 2015: 484-494

Text Summarization
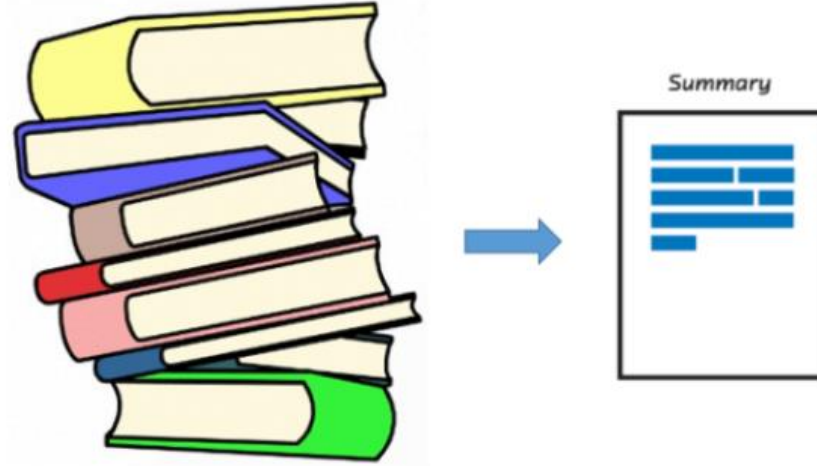
What is summarization?

Summarization

- <u>Summary</u>: a comprehensive and usually brief abstract, recapitulation, or compendium of previously stated facts or statements.

- <u>Automatic summarization</u>: It is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content.

Why is summarization important?

Too much data!

- Every 2 days we create as much data as we did from the beginning of time until 2003
- On Google alone, we submit 40,000 search queries per second. That amounts to 1.2 trillion searches yearly!
- Each minute, 300 new hours of video show up on YouTube.
- Facebook users send on average 31.25 million messages and view 2.77 million videos every minute

- This makes it hard for users to keep up with major happenings across the globe.

Classification of Text Summarization

Classification of Text Summarization
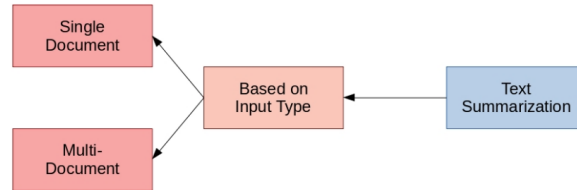
Classification of Text Summarization
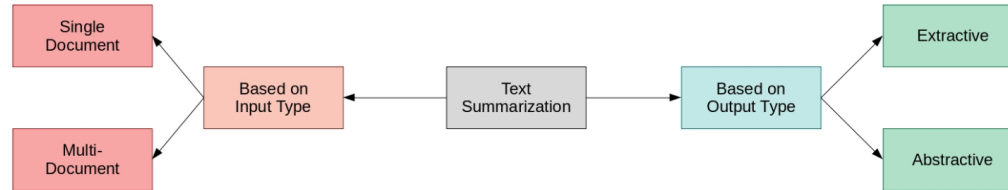
Extractive VS Abstractive Summarization

Two broad approaches to summarization:

- Extractive summarization
  - Easy to perform
  - Assemble summaries exclusively from passages taken directly from source text
  - Ensures baseline levels of grammaticality and accuracy
- Abstractive summarization
  - Generate novel words and phrases not featured in the source text - as human-written abstract usually does.
  - Sophisticated abilities that are crucial to high quality summarization such as paraphrasing, generalization or the incorporation of real-world knowledge are possible only in an abstractive framework.

Due to the difficulty of abstractive summarization, the great majority of past work has been extractive.

Recent success of sequence-to-sequence, in which recurrent neural networks (RNNs) both read and freely generate text has made abstractive summarization viable.

Classification of Text Summarization
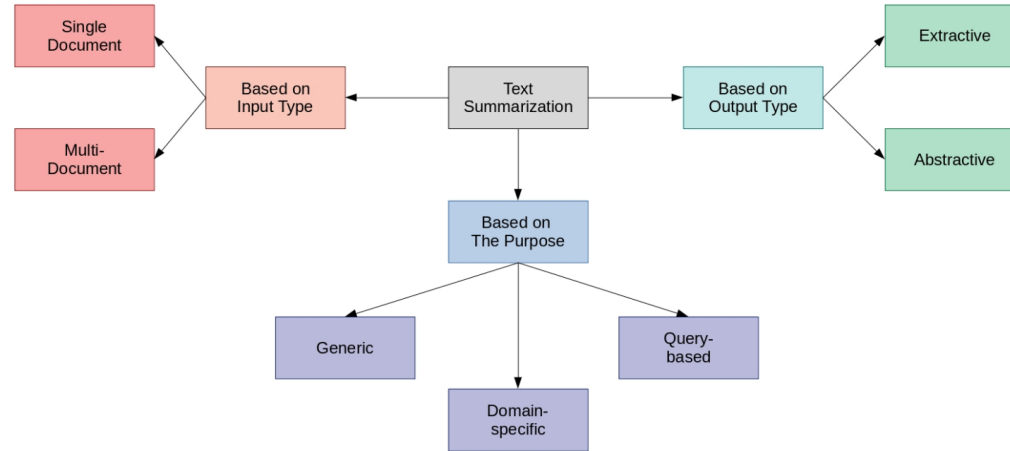
Classification of Text Summarization

- Text summarization can also be classified in the following manner:
  - On the basis of <u>content intensity</u> in output
    - *Indicative*
    - *Informative*
  - On the basis of learning process
    - *Supervised*
    - *Unsupervised*

Extractive Summarization

Document Example

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency. [Wikipedia]

Extractive Summarization

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency. [Wikipedia]

Extractive Summarization - Lack of Balance

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency.

Extractive Summarization - Lack of Balance

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency.

Extractive Summarization - Lack of cohesion

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency.

Extractive Summarization - Lack of cohesion

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency.

Abstractive Summarization

Abstractive Summarization

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

The disease was first identified in 1976, in two simultaneous outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku (Democratic Republic of the Congo), a village near the Ebola River from which the disease takes its name. EVD outbreaks occur intermittently in tropical regions of sub-Saharan Africa. From 1976 to 2012, the World Health Organization reports 24 outbreaks involving 2,387 cases with 1,590 deaths. The largest outbreak to date was the epidemic in West Africa, which occurred from December 2013 to January 2016, with 28,646 cases and 11,323 deaths. It was declared no longer an emergency on 29 March 2016. Other outbreaks in Africa began in the Democratic Republic of the Congo in May 2017, and 2018. In July 2019, the World Health Organization declared the Congo Ebola outbreak a world health emergency.

Ebola virus is a viral disease that was first identified in African regions of Nazara and Yambuku in 1976. It is a deadly disease with an average mortality rate of 50%. Typical symptoms for someone infected by Ebola virus include fever, sore throat, headaches, vomiting and diarrhoea.

Some Popular Extractive Summarization Systems

Positional method

Positional Method

- Introduced by P. Baxendale in 1958.
  - *Man-made index for technical literature - An experiment*
- 200 paragraphs
- First sentence was topic sentence in 85% of the cases
- Last sentence was topic sentence in 7% of the cases

Ebola, also known as Ebola virus disease (EVD) or Ebola hemorrhagic fever (EHF), is a viral hemorrhagic fever of humans and other primates caused by ebolaviruses. Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Vomiting, diarrhoea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally. The disease has a high risk of death, killing 25% to 90% of those infected, with an average of about 50%. This is often due to shock from fluid loss, and typically follows six to 16 days after symptoms appear.

Some Popular Extractive Summarization Systems

Luhn's method

Luhn's Method

- Introduced by H. P. Luhn in 1958.
  - *The Automatic creation of literature abstracts*
- Frequency of content terms
- Data Pre-processing
  - Stop word removal
  - Stemming (dogs -> dog)
- Select sentences with highest concentration of salient content terms
  - score = (# content terms in sentence)^2 /

    (span of content terms)

Figure 1 **Word-frequency diagram.**
Abscissa represents individual words arranged in order of frequency.

RESOLVING POWER OF SIGNIFICANT WORDS
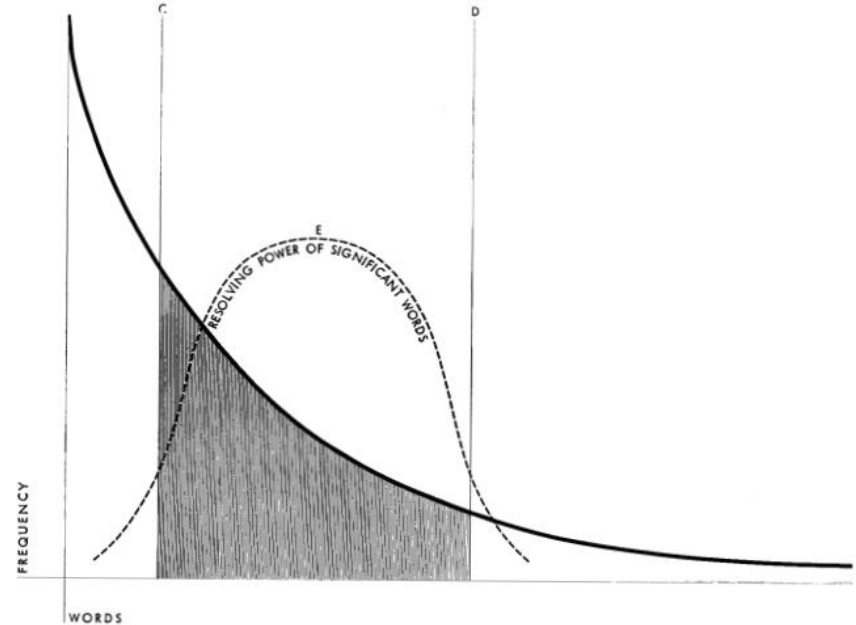
FREQUENCY

WORDS

Some Popular Extractive Summarization Systems

Edmundson's method

Edmundson's Method

- Introduced by H. P. Edmundson in 1968.
  - *New Methods in Automatic Extracting*
- Position (P)
- Word frequency (F)
- Cue words (C)
  - **Bonus words** — pointing to the importance of sentences
  - **Stigma words** — negative effect on the sentence importance
  - **Null words** — neutral or irrelevant to the importance of the sentences
- Document Structure (S)
- Linear combination of these 4 features: $a_1P + a_2F + a_3C + a_4S$

Some Popular Extractive Summarization Systems

FRUMP

FRUMP

- Fast Reading Understanding and Memory Program
- Introduced by G. Dejong in 1979
  - *Prediction and Substantiation: A new approach to natural language processing*
- Knowledge based summarization system
- Template filling approach based on UPI news stories
- **50 sketchy scripts**
  - Contain important events that are expected to occur in a specific situation
  - Summarizer looks for instances of salient events, filling in as many as possible

FRUMP - Demonstration Script

- The demonstrators arrived at the demonstration location
- The demonstrators march
- Police arrives at the scene
- The demonstrators communicate with the target of the demonstration
- The demonstrators attack the target of the demonstration
- The demonstrators attack the police
- The police arrests the demonstrators

Some Popular Extractive Summarization Systems

Classification

Classification

- Introduced by Kupiec et al. in 1995
  - *A trainable document summarization*
- First trainable method
  - Training set: original documents and manually created extracts
- Uses Naive Bayes classifier:

$$P(s \in \mathcal{S} | F_1, F_2, \ldots F_k) = \frac{P(F_1, F_2, \ldots F_k | s \in \mathcal{S}) P(s \in \mathcal{S})}{P(F_1, F_2, \ldots F_k)}$$

- Assuming the statistical independence of the features:

$$P(s \in \mathcal{S} | F_1, F_2, \ldots F_k) = \frac{\prod_{j=1}^{k} P(F_j | s \in \mathcal{S}) \, P(s \in \mathcal{S})}{\prod_{j=1}^{k} P(F_j)}$$

Classification

- Performance:
  - For 25% extracts - 84% precision
  - For smaller summaries - 74% improvement over lead summaries

- In 2002, Miles Osborne published a paper *Using Maximum Entropy for Sentence Extraction*
  - Maximum entropy models are performing better than Naive Bayes

Some Popular Extractive Summarization Systems

Maximal Marginal Reference

Maximal Marginal Reference

- Introduced by Carbonell and Goldstien in 1998
  - *The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries*
- Query based summaries

$$MMR \overset{\text{def}}{=} Arg \max_{D_i \in R \setminus S} \left[\lambda(Sim_1(D_i, Q) - (1-\lambda) \max_{D_j \in S} Sim_2(D_i, D_j)) \right]$$

  - Q = user query
  - R = ranked list of sentences
  - S = already retrieved sentences
  - Sim = similarity metrics

Some Popular Extractive Summarization Systems

Mead

Mead

- Introduced by Radev et al. in 2000
- Centroid based method
- Single and multi-document
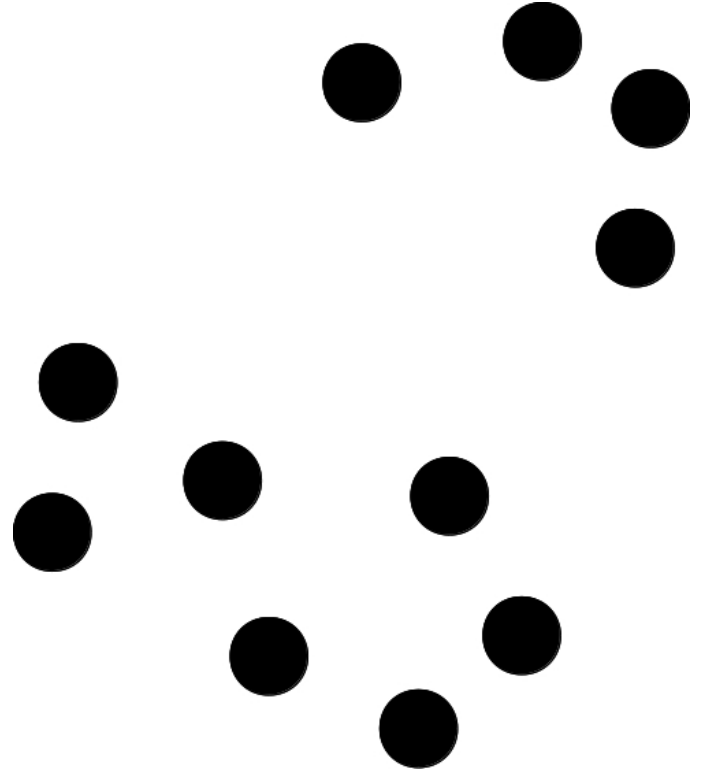
Mead

- Introduced by Radev et al. in 2000
- Centroid based method
- Single and multi-document

Mead

- Introduced by Radev et al. in 2000
- Centroid based method
- Single and multi-document
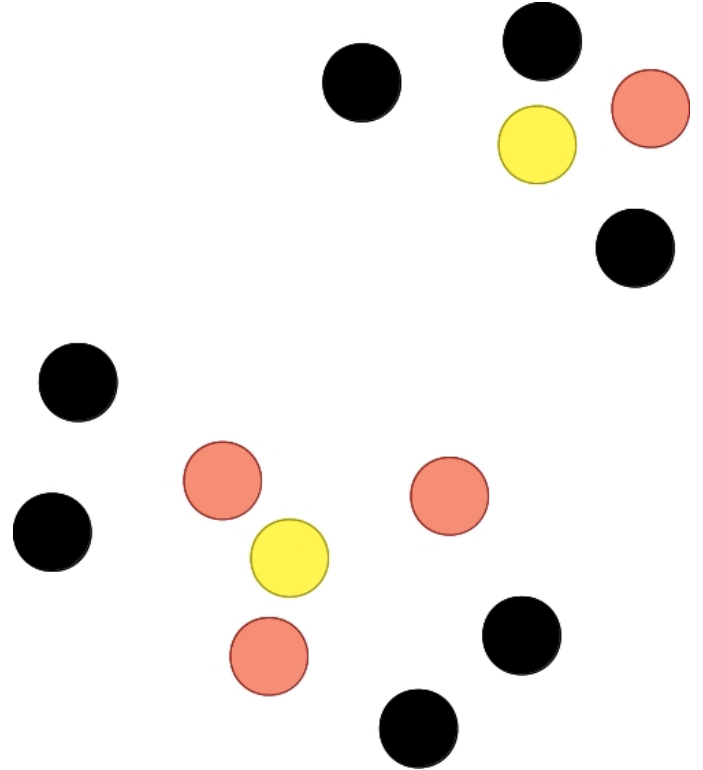
Some Popular Extractive Summarization Systems

LexRank

LexRank (similar to TextRank)

- Introduced by Erkan and Radev in 2004
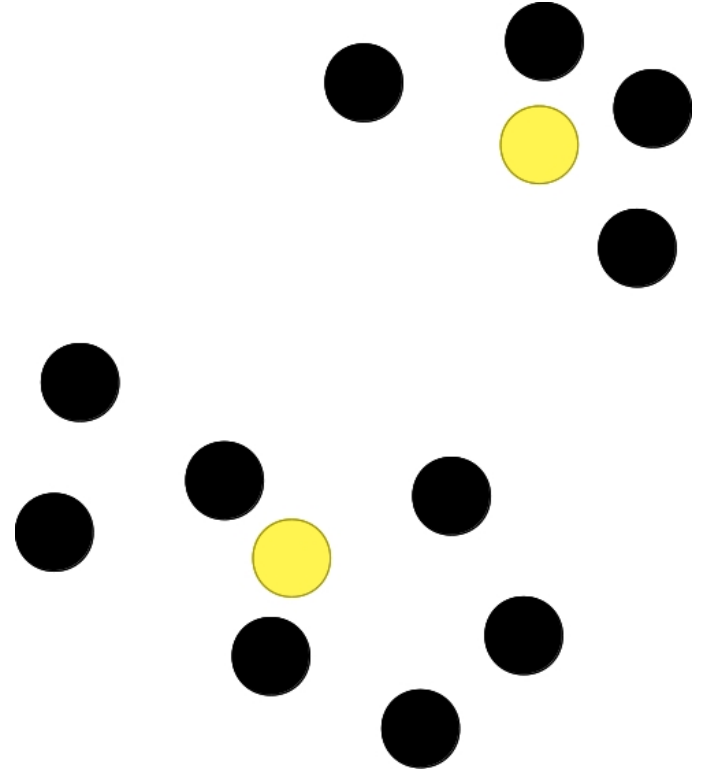 - *LexRank: Graph-based Lexical Centrality as Salience in Text Summarization*
- Graph based method
- **Lexical centrality**
- Idea:
  - Create a similarity matrix
  - Represent sentences as nodes in a graph
  - Connecting nodes based on similarity matrix
  - Random walk on the graph to find important sentences

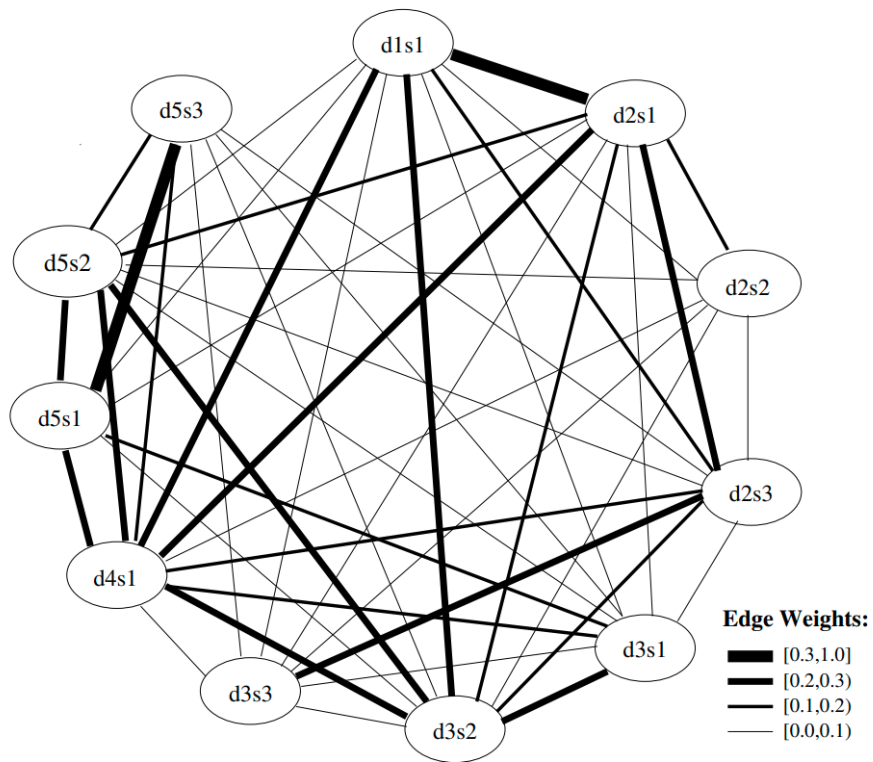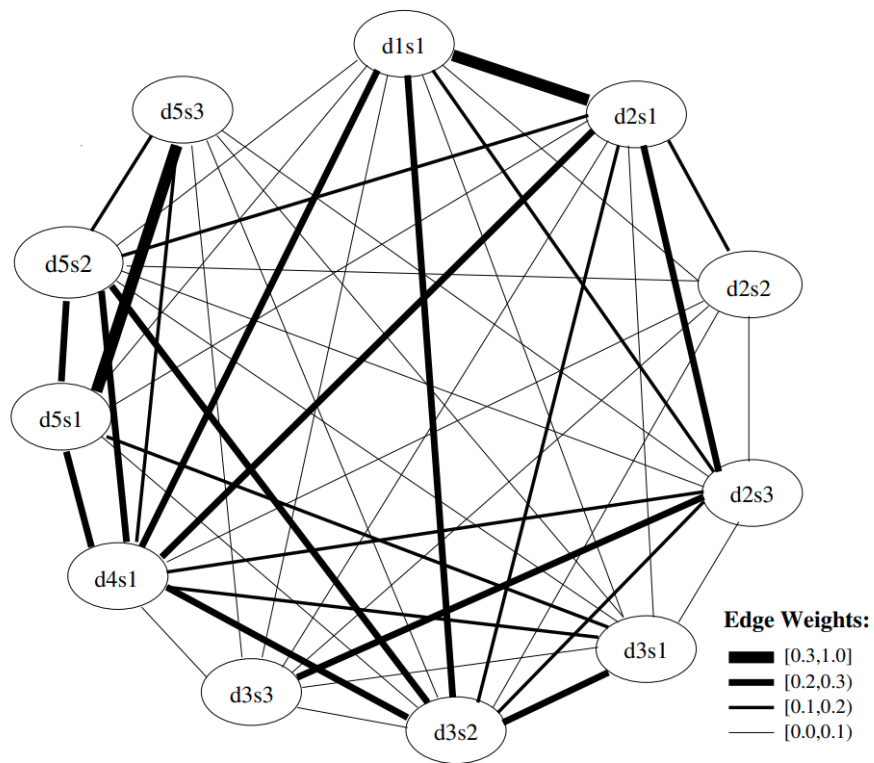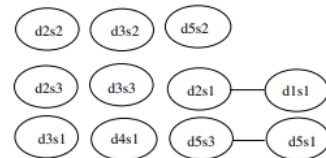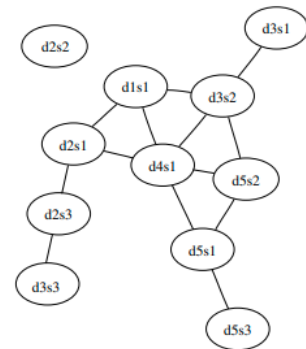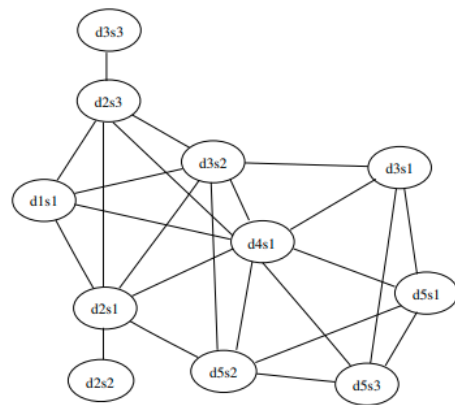| SNo	ID	Text
1	d1s1	Iraqi Vice President Taha Yassin Ramadan announced today, Sunday, that Iraq refuses to back down from its decision to stop cooperating with disarmament inspectors before its demands are met.
2	d2s1	Iraqi Vice president Taha Yassin Ramadan announced today, Thursday, that Iraq rejects cooperating with the United Nations except on the issue of lifting the blockade imposed upon it since the year 1990.
3	d2s2	Ramadan told reporters in Baghdad that "Iraq cannot deal positively with whoever represents the Security Council unless there was a clear stance on the issue of lifting the blockade off of it.
4	d2s3	Baghdad had decided late last October to completely cease cooperating with the inspectors of the United Nations Special Commission (UNSCOM), in charge of disarming Iraq's weapons, and whose work became very limited since the fifth of August, and announced it will not resume its cooperation with the Commission even if it were subjected to a military operation.
5	d3s1	The Russian Foreign Minister, Igor Ivanov, warned today, Wednesday against using force against Iraq, which will destroy, according to him, seven years of difficult diplomatic work and will complicate the regional situation in the area.
6	d3s2	Ivanov contended that carrying out air strikes against Iraq, who refuses to cooperate with the United Nations inspectors, "will end the tremendous work achieved by the international group during the past seven years and will complicate the situation in the region."
7	d3s3	Nevertheless, Ivanov stressed that Baghdad must resume working with the Special Commission in charge of disarming the Iraqi weapons of mass destruction (UNSCOM).
8	d4s1	The Special Representative of the United Nations Secretary-General in Baghdad, Prakash Shah, announced today, Wednesday, after meeting with the Iraqi Deputy Prime Minister Tariq Aziz, that Iraq refuses to back down from its decision to cut off cooperation with the disarmament inspectors.
9	d5s1	British Prime Minister Tony Blair said today, Sunday, that the crisis between the international community and Iraq "did not end" and that Britain is still "ready, prepared, and able to strike Iraq."
10	d5s2	In a gathering with the press held at the Prime Minister's office, Blair contended that the crisis with Iraq "will not end until Iraq has absolutely and unconditionally respected its commitments" towards the United Nations.
11	d5s3	A spokesman for Tony Blair had indicated that the British Prime Minister gave permission to British Air Force Tornado planes stationed in Kuwait to join the aerial bombardment against Iraq.

	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00

|	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00

Edge Weights:
- [0.3,1.0]
- [0.2,0.3)
- [0.1,0.2)
- [0.0,0.1)

$$p(u) = \frac{d}{N} + (1-d) \sum_{v \in adj[u]} \frac{p(v)}{deg(v)}$$

Some Popular Abstractive Summarization Systems

Sequence-to-Sequence frameworks

Text Summarization using Encoder Decoder network
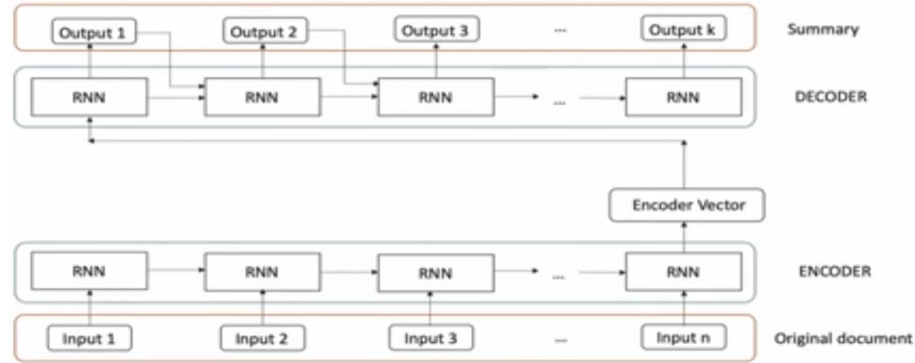
Text Summarization using Encoder Decoder networks

- Major shortcomings of sequence-to-sequence networks:
  - The model is liable to produce <u>factual inaccuracies</u>.
  - The output generated tends to <u>repeat</u> itself.
  - The model doesn't propose an alternative for out-of-vocabulary scenario.

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, *muhammadu buhari* told cnn's christiane amanpour that he plans to aggressively fight corruption that has long plagued nigeria and go after the root of the nation's unrest. *buhari* said he'll "rapidly give attention" to curbing violence in the northeast part of nigeria, where the terrorist group boko haram operates. by cooperating with neighboring nations chad, cameroon and niger, he said his administration is confident it will be able to thwart criminals and others contributing to nigeria's instability. for the first time in nigeria's history, the opposition defeated the ruling party in democratic elections. *buhari* defeated incumbent goodluck jonathan by about 2 million votes, according to nigeria's independent national electoral commission. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

Baseline Seq2Seq + Attention: UNK UNK says his administration is confident it will be able to destabilize nigeria's economy. UNK says his administration is confident it will be able to thwart criminals and other nigerians. he says the country has long nigeria and nigeria's economy.