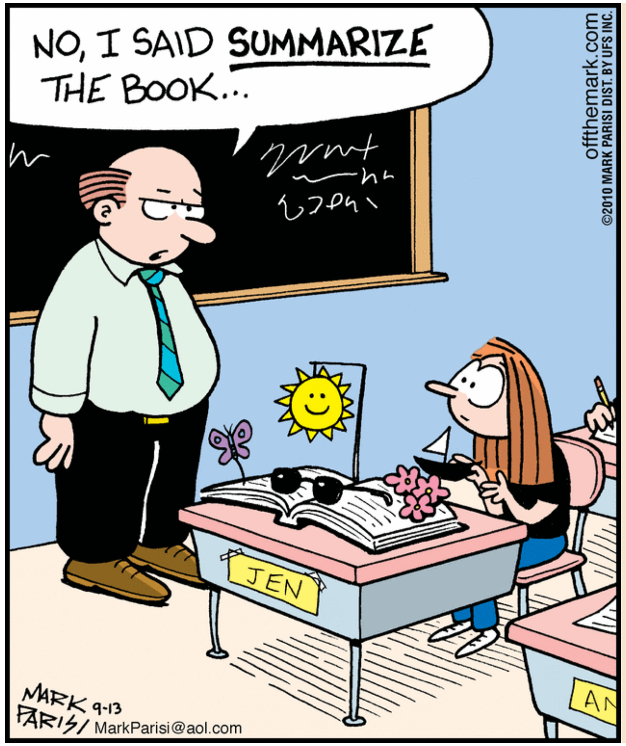


1. 좋은 글이네요
2. 하지만 읽지는 않았습니다
3. 세 줄 요약 좀

이 사이에서 긴 글을 쓴다는 것…….

https://brunch.co.kr/@cowithawrd/72

Text Summarization

Pilsung Kang
School of Industrial Management Engineering
Korea University

TLDR: Too Long, Didn't Read

- Is it really the title of a research paper?

Abstract While many approaches to make neural networks more fathomable have been proposed, they are restricted to interrogating the network with input data. [...] In this work, we propose neural persistence, a complexity measure for neural network architectures based on topological data analysis on weighted stratified graphs. [...]

Intro [...] In this work, we present the following contributions: We introduce neural persistence, a novel measure for characterizing the structural complexity of neural networks that can be efficiently computed. [...]

Conclusion [...] However, this did not yield an early stopping measure because it was never triggered, thereby suggesting that neural persistence captures salient information that would otherwise be hidden among all the weights of a network [...]

TLDR We develop a new topological complexity measure for deep neural networks and demonstrate that it captures their salient properties.

TLDR: Extreme Summarization of Scientific Documents

Isabel Cachola† Kyle Lo† Arman Cohan† Daniel S. Weld†‡

†Allen Institute for AI

‡Paul G. Allen School of Computer Science & Engineering, University of Washington

{isabelc,kylel,armanc,danw}@allenai.org

Peer review The paper proposes variance regularizing adversarial learning (VRAL), a new method for training GANs. The motivation is to ensure that the gradient for the generator does not vanish. [...] The discriminator itself is trained through two additional meta-discriminators. Are the meta-discriminators really necessary? Have you tried matching moments or using other methods [...]

Derived TLDR The paper proposes variance regularizing adversarial learning for training gans to ensure that the gradient for the generator does not vanish.

AI Hub: Document Summarization

- Datasets for AI: https://aihub.or.kr/

AI Hub: Document Summarization

문서요약 텍스트 AI 데이터

소개 다운로드

구축목적

- 다양한 주제의 한국어 원문으로부터 추출요약문과 생성요약문을 도출해낼 수 있도록 인공지능을 훈련하기 위한 데이터셋

활용분야

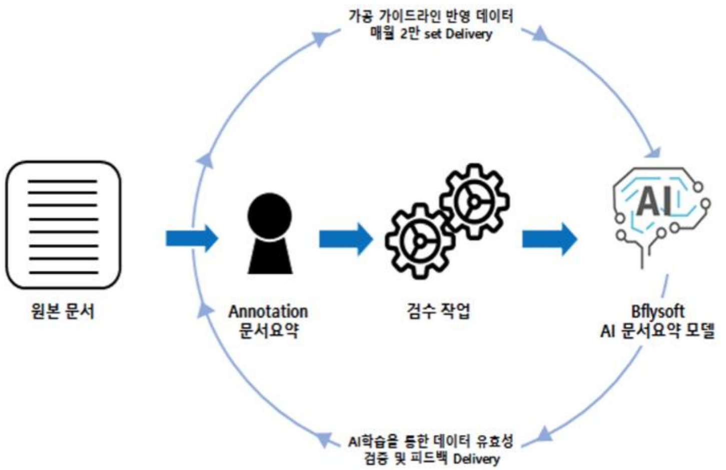
- 뉴스기사 요약, 법률문서 요약, 사업보고서 요약 등 핵심내용을 신속하고 정확하게 파악할 수 있는 AI 요약기술 개발

주요 키워드

- 문서요약, 추출요약, 생성요약, 한국형 문서요약 데이터셋

소개

- 다양한 한국어 원문 데이터로부터 정제된 추출 및 생성 요약문을 도출하고 검증한 한국어 문서요약 AI 데이터셋으로, 기존 영문 문서요약 데이터셋과는 다른 원문 데이터의 다양성을 추구하며, 요약문 재사용에 제한이 없도록 저작권 문제를 완전히 해결한 원천 데이터를 확보



구축 내용 및 제공 데이터량

- 원문데이터 40만 건 (신문기사 30만 건, 기고문 3만 건, 잡지기사 1만 건, 논문 초록 3만 건, 법원 판결문 3만 건)을 활용하여 각각 추출요약 40만 건, 생성요약 40만 건, 총 80만 건의 요약문 도출
- 원문으로부터 변형 없이 그대로 선택된 3개 문장으로 추출요약문 생성
- 원문의 내용을 바탕으로 재작성된 생성요약문 생성

데이터 종류	데이터 형태	목표 수량
신문기사	뉴스 텍스트	원문데이터 30만 건 / 요약문 60만 건
기고문	오피니언 텍스트	원문데이터 3만 건 / 요약문 6만 건
잡지	웹진 기사 텍스트	원문데이터 1만 건 / 요약문 2만 건
법률	법원 판결문 뉴스 텍스트 및 법원 주요 판결문 텍스트	원문데이터 3만 건 / 요약문 6만 건
논문	논문요약 (abstract) 텍스트	원문데이터 3만 건 / 요약문 6만 건
총계		**원문데이터 40만 건 / 요약문 80만 건**

한국정보화진흥원이 추진하는 AI 학습용 데이터 구축 사업(일명 '데이터 댐'사업) 중 문서 요약 텍스트 AI 학습용 데이터 구축사업자로 비플라이소프트가 22일 선정됐다. '데이터 댐' 사업은 AI 개발에 필수적인 학습 데이터를 구축하는 사업이다. 문서 요약 텍스트 AI 데이터 분야 사업 목표는 한국어로 된 자료의 핵심 내용과 의미를 포함하는 요약문을 자동으로 생성하는 AI 기술 개발을 위한 요약 텍스트 데이터 구축이다. 비플라이소프트는 이를 위해 오는 12월까지 신문 기사, 논문, 판결문 등 문서 텍스트 40만 여건의 원문과 핵심내용 요약 데이터를 확보할 예정이다. 사업 규모는 약 26억원이다. 주관사인 비플라이소프트를 비롯 (주)위고, 테스트웍스, 고려대 데이터사이언스 및 비즈니스 애널리틱스 연구실, (주)에이아이닷엠 등이 컨소시엄에 참여했다.

http://biz.khan.co.kr/khan_art_view.html?artid=202006221717001&code=930100

Document Summarization: Demo

- Original Text

Growing demand for electric car batteries will cause prices of the main materials to surge, Goldman Sachs analysts said in a March 18 note. That in turn will drive prices of batteries higher by about 18%, affecting the total profit of electric car makers since the battery accounts for about 20% to 40% of the vehicle cost, the Goldman analysts said. While the report didn't give specific price targets for the commodities, the analysts' model predicted a return to historical peak prices would more than double the cost of lithium for electric battery makers. That of cobalt would also double, while the cost of nickel would rise by 60%. "Prices for the three main natural resources have been rising since the start of 2021," the Goldman report said. "We believe that in order to promote sustainable EV industries, some countries may consider implementing policies to increase national stockpiles." Limited availability of nickel suitable for car batteries could even accelerate a shift to another kind of battery called lithium iron phosphate (LFP), the report said. Tesla and Chinese start-up Xpeng are among automakers already using this type of battery, which does not use nickel or cobalt but stores relatively less energy. If nickel prices hit their historic high of $50,000 per tone, that could add $1,250 to $1,500 per electric vehicle, which could hurt consumer demand for the cars, the analysts said. Ultimately, the growth of the electric car industry and demand for battery materials depends on how many vehicles people buy. The tipping point for consumers broadly to switch from gas-powered vehicles to electric cars is generally expected to come when the battery cost has fallen sufficiently. That shift could happen in the next decade. Goldman predicts battery costs will drop below that of internal combustion engines in 2030.

Document Summarization: Demo

- **Summarization Result:** https://deepai.org/machine-learning-model/summarization

Growing demand for electric car batteries will cause prices of the main materials to surge, Goldman Sachs analysts said in a March 18 note. That in turn will drive prices of batteries higher by about 18%, affecting the total profit of electric car makers since the battery accounts for about 20% to 40% of the vehicle cost, the Goldman analysts said. While the report didn't give specific price targets for the commodities, the analysts' model predicted a return to historical peak prices would more than double the cost of lithium for electric battery makers. That of cobalt would also double, while the cost of nickel would rise by 60%. "Prices for the three main natural resources have been rising since the start of 2021," the Goldman report said. "We believe that in order to promote sustainable EV industries, some countries may consider implementing policies to increase national stockpiles." Limited availability of nickel suitable for car batteries could even accelerate a shift to another kind of battery called lithium iron phosphate (LFP), the report said. Tesla and Chinese start-up Xpeng are among automakers already using this type of battery, which does not use nickel or cobalt but stores relatively less energy. If nickel prices hit their historic high of $50,000 per tone, that could add $1,250 to $1,500 per electric vehicle, which could hurt consumer demand for the cars, the analysts said. Ultimately, the growth of the electric car industry and demand for battery materials depends on how many vehicles people buy. The tipping point for consumers broadly to switch from gas-powered vehicles to electric cars is generally expected to come when the battery cost has fallen sufficiently. That shift could happen in the next decade. Goldman predicts battery costs will drop below that of internal combustion engines in 2030.

Document Summarization: Demo

- Summarization Result: http://document-summarization.labs-projects.findwise.com/demomds/demomds

Document Summarization

Summaries

(Approx. 40 words)

[Submodular optimization]

That in turn will drive prices of batteries higher by about 18%, affecting the total profit of electric car makers since the battery accounts for about 20% to 40% of the vehicle cost, the Goldman analysts said.

[TextRank]

Limited availability of nickel suitable for car batteries could even accelerate a shift to another kind of battery called lithium iron phosphate (LFP), the report said. Ultimately, the growth of the electric car industry and demand for battery materials depends on how many vehicles people buy.

CHALMERS
UNIVERSITY OF TECHNOLOGY

LAB | RESEARCH GROUP

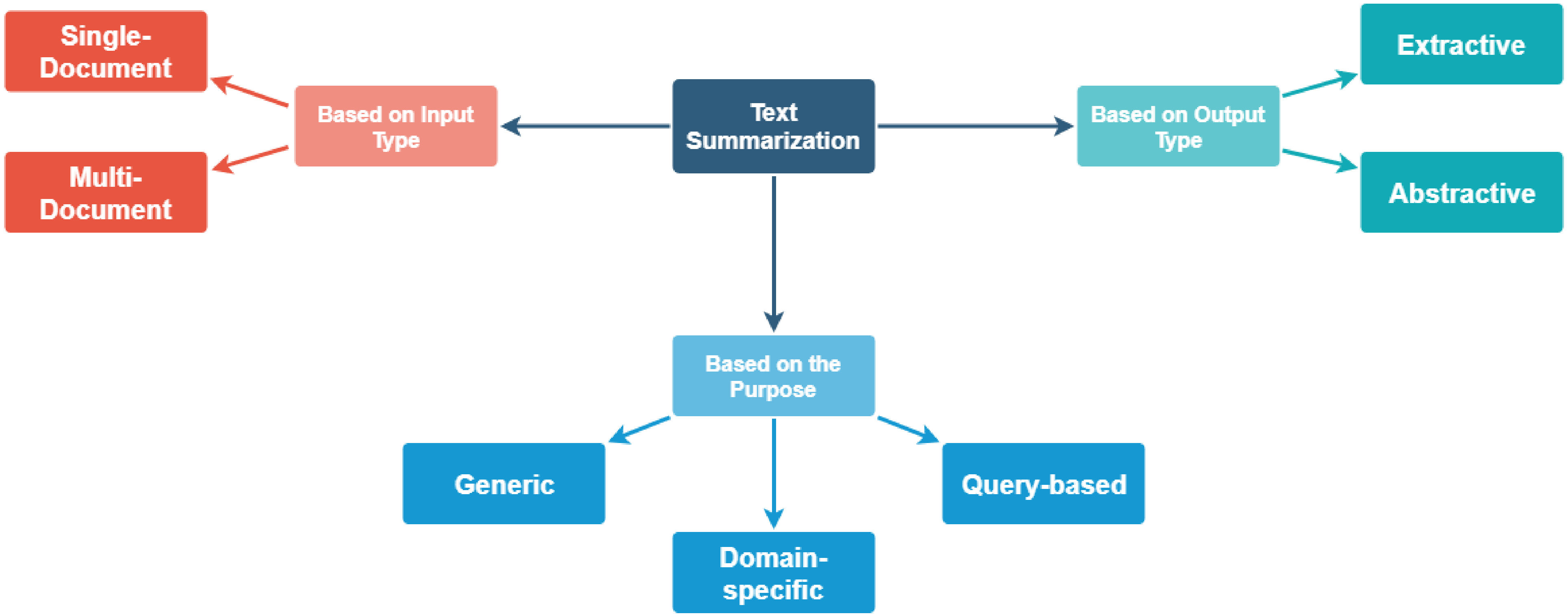
FINDWISE
SEARCH DRIVEN SOLUTIONS

Original Text

Growing demand for electric car batteries will cause prices of the main materials to surge, Goldman Sachs analysts said in a March 18 note. [**That in turn will drive prices of batteries higher by about 18%, affecting the total profit of electric car makers since the battery accounts for about 20% to 40% of the vehicle cost, the Goldman analysts said.**] While the report didn't give specific price targets for the commodities, the analysts' model predicted a return to historical peak prices would more than double the cost of lithium for electric battery makers. That of cobalt would also double, while the cost of nickel would rise by 60%. "Prices for the three main natural resources have been rising since the start of 2021," the Goldman report said. "We believe that in order to promote sustainable EV industries, some countries may consider implementing policies to increase national stockpiles." [**Limited availability of nickel suitable for car batteries could even accelerate a shift to another kind of battery called lithium iron phosphate (LFP), the report said.**] Tesla and Chinese start-up Xpeng are among automakers already using this type of battery, which does not use nickel or cobalt but stores relatively less energy. If nickel prices hit their historic high of $50,000 per tone, that could add $1,250 to $1,500 per electric vehicle, which could hurt consumer demand for the cars, the analysts said. [**Ultimately, the growth of the electric car industry and demand for battery materials depends on how many vehicles people buy.**] The tipping point for consumers broadly to switch from gas-powered vehicles to electric cars is generally expected to come when the battery cost has fallen sufficiently. That shift could happen in the next decade. Goldman predicts battery costs will drop below that of internal combustion engines in 2030.

Document Summarization: Overview

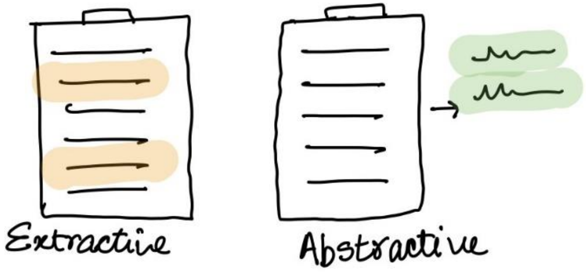
- Definition
  - ✓ A summary is a text that is produced from one or more texts, that contains a significant portion of the information in the original text(s), and that is no longer than half of the original text(s) - Gupta, Joshi, and Dungarwal (MIT)



https://devopedia.org/text-summarization

Document Summarization: Overview

- Type of Summarization
  - ✓ **Extractive** Summarization
    - Extractive summaries are created by reusing portions (words, sentences, etc.) of the input text verbatim.
    - For example, search engines typically generate extractive summaries from webpages.
    - Most of the summarization research today is on extractive summarization.



  - ✓ **Abstractive** Summarization
    - Information from the source text is re-phrased.
    - Human beings generally write abstractive summaries (except when they do their assignments).
    - Abstractive summarization has not reached a mature stage because allied problems such as semantic representation, inference and natural language generation are relatively harder.

Figure source: https://medium.com/@sarthakanand/extractive-text-summary-using-textrank-1032777750a2

Document Summarization: Overview

- Extractive vs. Abstractive

Extractive

Select parts of the original text to form a summary

- Easier
- Restrictive (no paraphrasing)

Abstractive

Generate new text using natural language generation techniques.

- More difficult
- More flexible (more human)

https://www.slideshare.net/ThoPhan27/abstractive-text-summarization

(a) Extractive Summarization

Source Text: Peter and Elizabeth took a taxi to attend the night party in the city.

While in the party, Elizabeth collapsed and was rushed to the hospital.

Summary: Peter and Elizabeth attend party city. Elizabeth rushed hospital.

(b) Abstractive Summarization

Source Text: Peter and Elizabeth took a taxi to attend the night party in the city.

While in the party, Elizabeth collapsed and was rushed to the hospital.

Summary: Elizabeth was hospitalized after attending a party with Peter.

https://devopedia.org/text-summarization

Document Summarization: Supervised vs. Unsupervised

- **Supervised** Summarization
  - ✓ Use a collection of documents and human-generated summaries to train a summarization model

Datasets > Modality > Texts > CNN/Daily Mail

CNN/Daily Mail

Edit

Introduced by Nallapati et al. in Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond

CNN/Daily Mail is a dataset for text summarization. Human generated abstractive summary bullets were generated from news stories in CNN and Daily Mail websites as questions (with one of the entities hidden), and stories as the corresponding passages from which the system is expected to answer the fill-in the-blank question. The authors released the scripts that crawl, extract and generate pairs of passages and questions from these websites.

In all, the corpus has 286,817 training pairs, 13,368 validation pairs and 11,487 test pairs, as defined by their scripts. The source documents in the training set have 766 words spanning 29.74 sentences on an average while the summaries consist of 53 words and 3.72 sentences.

Source: Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond

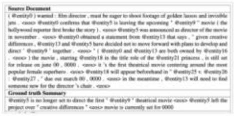
Homepage

Benchmarks

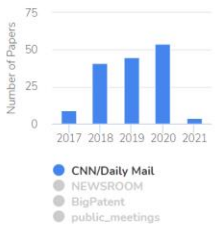
Edit

TREND	TASK	DATASET VARIANT	BEST MODEL	PAPER	CODE
	Abstractive Text Summarization	CNN / Daily Mail	ERNIE-GENLARGE		
	Document Summarization	CNN / Daily Mail	MatchSum		
	Question Answering	CNN / Daily Mail	GA+MAGE		
	Extractive Text Summarization	CNN / Daily Mail	ITS		
	Text Summarization	CNN / Daily Mail (Anonymized)	HSASS		

Samples



Usage



License Edit

MIT

Modalities Edit

Texts

Languages Edit

English

Source Document
(@entity0) wanted : film director , must be eager to shoot footage of golden lassos and invisible jets . <eos> @entity0 confirms that @entity5 is leaving the upcoming " @entity9 " movie (the hollywood reporter first broke the story) . <eos> @entity5 was announced as director of the movie in november . <eos> @entity0 obtained a statement from @entity13 that says , " given creative differences , @entity13 and @entity5 have decided not to move forward with plans to develop and direct ' @entity9 ' together . <eos> " (@entity0 and @entity13 are both owned by @entity16 . <eos>) the movie , starring @entity18 in the title role of the @entity21 princess , is still set for release on june 00 , 0000 . <eos> it 's the first theatrical movie centering around the most popular female superhero . <eos> @entity18 will appear beforehand in " @entity25 v. @entity26 : @entity27 , " due out march 00 , 0000 . <eos> in the meantime , @entity13 will need to find someone new for the director 's chair . <eos>
Ground truth Summary
@entity5 is no longer set to direct the first " @entity9 " theatrical movie <eos> @entity5 left the project over " creative differences " <eos> movie is currently set for 0000

https://paperswithcode.com/dataset/cnn-daily-mail-1

Document Summarization: Supervised vs. Unsupervised

- **Supervised** Summarization
  - ✓ Use a collection of documents and human-generated summaries to train a summarization model

NEWSROOM OVERVIEW

A large, diverse summarization dataset.

- 1.3 million articles and summaries extracted from HTML metadata.
- Written in the newsrooms of 38 publications over the last 20 years.
- Wide range of both abstractive and extractive summaries.

extractive = borrows phrases from article

abstractive = uses more ***novel vocabulary***

ABSTRACTIVE

South African photographer Anton Hammerl, ***missing*** in Libya ***since*** April ***4th***, was ***killed*** in Libya ***more than*** a ***month ago***.

MIXED

A major climate protest in New York on Sunday ***could*** mark a seminal ***shift*** in the ***politics*** of global warming, just ***ahead*** of the ***U.N.*** Climate Summit.

EXTRACTIVE

A person familiar with the search ***tells*** The Associated Press that Texas has ***offered*** its head coaching job to Louisvilles Charlie Strong and he is expected to accept.

COMPARING EXISTING DATASETS

	Type	Authors	Size	References	Sources	Strategy
DUC Datasets	Summary	NIST for DUC	Hundreds	Multiple	3	Abstractive
Gigaword Corpus	Headline	Journalists	9.9 million	Single	7	Abstractive
New York Times	Archival	Library scientists	600 thousand	Single	1	Extractive
CNN / Daily Mail	List of facts	Journalists	300 thousand	Single	2	Extractive
NEWSROOM	Summary	Journalists	1.3 million	Single	38	Both

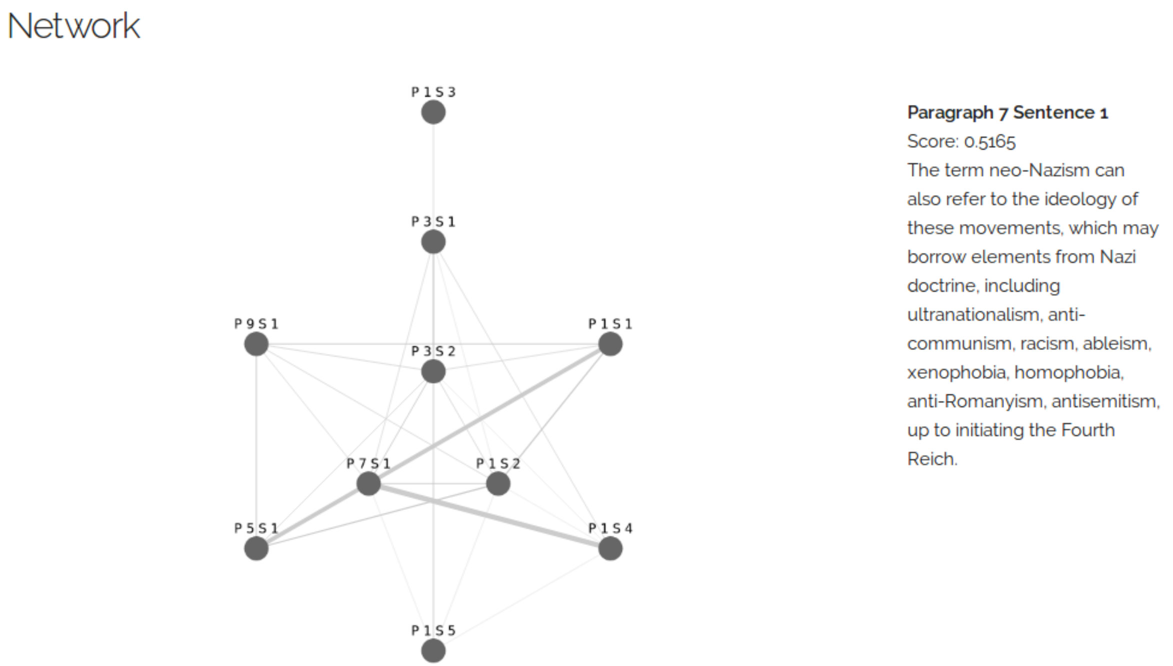
https://yoavartzi.com/pub/gna-naacl.2018.slides.pdf

Document Summarization: Supervised vs. Unsupervised

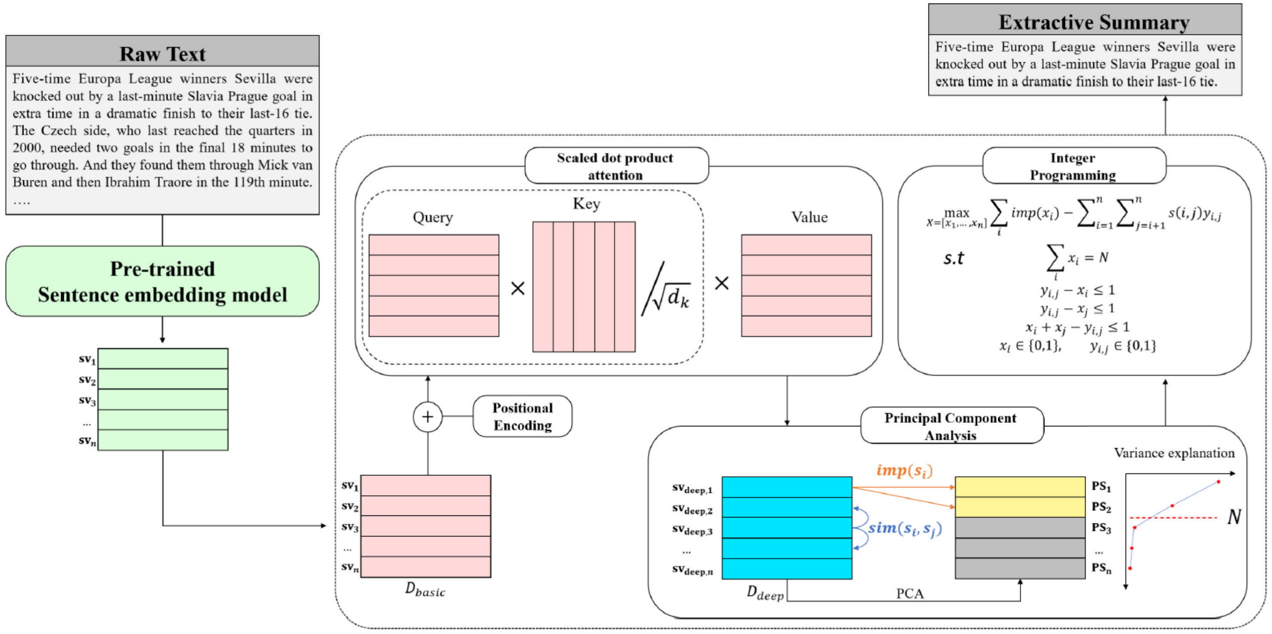
- **Supervised** Summarization
  - ✓ Use a collection of documents and human-generated summaries to train a summarization model.
  - ✓ Features (position of the sentence, number of words in the sentence, etc.) of sentences that make them good candidates for inclusion in the summary are learnt.
  - ✓ Sentences in an original training document can be labelled as "in summary" or "not in summary".
  - ✓ Drawback
    - Training data is expensive to produce and relatively sparse.
    - Most readily available human generated summaries are abstractive in nature.

Document Summarization: Supervised vs. Unsupervised

- **Unsupervised** Summarization
  - ✓ Find important parts of a given sentence based on a significance score that measures the relative importance of the sentence in the document.
  - ✓ A document is represented as a graph in which sentences in the document are represented as nodes and their relationships are expressed by the weight of edges.



https://medium.com/the-artificial-impostor/use-textrank-to-extract-most-important-sentences-in-article-b8efc7e70b4



Jang, M., & Kang, P. (2021). Learning-Free Unsupervised Extractive Summarization Model. *IEEE Access*, *9*, 14358-14368.

Document Summarization: State-of-the-art

- Papers with Code: Text Summarization

Text Summarization

Edit Task

Natural Language Processing

149 papers with code **15** benchmarks **38** datasets

About

Edit

Shortening a set of data computationally, to create a summary that represents the most important or relevant information within the original content (Source: Wikipedia).

Benchmarks

Add a Result

TREND	DATASET	BEST METHOD	PAPER TITLE	PAPER	CODE	COMPARE
	GigaWord	BART-RXF	Better Fine-Tuning by Reducing Representational Collapse			See all
	Pubmed	DANCER PEGASUS	A Divide-and-Conquer Approach to the Summarization of Long Documents			See all
	arXiv	BigBird-Pegasus	Big Bird: Transformers for Longer Sequences			See all
	DUC 2004 Task 1	Transformer+LRPE+PE+Re-ranking+Ensemble	Positional Encoding to Control Output Sequence Length			See all
	CNN / Daily Mail (Anonymized)	HSASS	A Hierarchical Structured Self-Attentive Model for Extractive Document Summarization (HSSAS)			See all
	X-Sum	PEGASUSLARGE	PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization			See all
	GigaWord-10k	ERNIE-GENLARGE (large-scale text corpora)	ERNIE-GEN: An Enhanced Multi-Flow Pre-training and Fine-tuning Framework for Natural Language Generation			See all
	WikiHow	BertSum	Abstractive Summarization of Spoken andWritten Instructions with BERT			See all

Subtasks

Abstractive Text Summarization	Document Summarization	Multi-Document Summarization	Extractive Text Summarization	Sentence Summarization
7 benchmarks	2 benchmarks	2 benchmarks	3 benchmarks	
131 papers with code	94 papers with code	36 papers with code	25 papers with code	15 papers with code

See all 9 subtasks

Datasets

Pubmed

arXiv

WikiHow

Reddit TIFU

How2

https://paperswithcode.com/task/text-summarization

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ ROUGE-N: n-gram recall between a candidate summary and a set of reference summaries.

$$\text{ROUGE} - \text{N} = \frac{\sum_{S \in \{Reference\ Summaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{Reference\ Summaries\}} \sum_{gram_n \in S} Count(gram_n)}$$

  - ✓ ROUGE-N is recall-oriented while BLEU (commonly used in machine translation) is precision-oriented.
    - ROUGE-1: Unigram overlap
    - ROUGE-2: Bigram overlap
    - ROUGE-N: N-gram overlap
    - ROUGE-L: Longest common subsequence overlap

Lin, C.Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ Example

```
[Reference Summary: R]
The capital of Korea, Seoul, is one of the biggest cities of the world.
```

```
[Model Summary: M1]
Seoul is the biggest city of the world.
```

```
[Model Summary: M2]
World is a biggest cities of the Seoul.
```

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ ROUGE-1

```
[Reference Summary: R]
The capital of Korea, Seoul, is one of the biggest cities of the world.
```

```
[Model Summary: M1]
Seoul is the biggest city of the world.
```

$$\text{ROUGE}-1 = \frac{7}{14} = 0.5$$

```
[Model Summary: M1]
World is a biggest cities of the Seoul.
```

$$\text{ROUGE}-1 = \frac{7}{14} = 0.5$$

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ ROUGE-2

```
[Reference Summary: R]
The capital of Korea, Seoul, is one of the biggest cities of the world.
```

```
[Model Summary: M1]
Seoul is the biggest city of the world.
```

$$\mathrm{ROUGE}-2=\frac{4}{13}=0.3077$$

```
[Model Summary: M1]
World is a biggest cities of the Seoul.
```

$$\mathrm{ROUGE}-2=\frac{3}{14}=0.2143$$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ ROUGE-L

```
[Reference Summary: R]
The capital of Korea, Seoul, is one of the biggest cities of the world.
```

```
[Model Summary: M1]
Seoul is the biggest city of the world.
```

$$\text{ROUGE}-\text{L} = \frac{7}{14} = 0.5$$

```
[Model Summary: M1]
World is a biggest cities of the Seoul.
```

$$\text{ROUGE}-\text{L} = \frac{5}{14} = 0.3571$$

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ Correlation with human judgements

Single document summarization

	DUC 2001 100 WORDS SINGLE DOC						DUC 2002 100 WORDS SINGLE DOC					
	1 REF			3 REFS			1 REF			2 REFS		
Method	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP
R-1	0.76	0.76	0.84	0.80	0.78	0.84	0.98	0.98	0.99	0.98	0.98	0.99
R-2	0.84	0.84	0.83	0.87	0.87	0.86	0.99	0.99	0.99	0.99	0.99	0.99
R-3	0.82	0.83	0.80	0.86	0.86	0.85	0.99	0.99	0.99	0.99	0.99	0.99
R-4	0.81	0.81	0.77	0.84	0.84	0.83	0.99	0.99	0.98	0.99	0.99	0.99
R-5	0.79	0.79	0.75	0.83	0.83	0.81	0.99	0.99	0.98	0.99	0.99	0.98
R-6	0.76	0.77	0.71	0.81	0.81	0.79	0.98	0.99	0.97	0.99	0.99	0.98
R-7	0.73	0.74	0.65	0.79	0.80	0.76	0.98	0.98	0.97	0.99	0.99	0.97
R-8	0.69	0.71	0.61	0.78	0.78	0.72	0.98	0.98	0.96	0.99	0.99	0.97
R-9	0.65	0.67	0.59	0.76	0.76	0.69	0.97	0.97	0.95	0.98	0.98	0.96
R-L	0.83	0.83	0.83	0.86	0.86	0.86	0.99	0.99	0.99	0.99	0.99	0.99
R-S*	0.74	0.74	0.80	0.78	0.77	0.82	0.98	0.98	0.98	0.98	0.97	0.98
R-S4	0.84	0.85	0.84	0.87	0.88	0.87	0.99	0.99	0.99	0.99	0.99	0.99
R-S9	0.84	0.85	0.84	0.87	0.88	0.87	0.99	0.99	0.99	0.99	0.99	0.99
R-SU*	0.74	0.74	0.81	0.78	0.77	0.83	0.98	0.98	0.98	0.98	0.98	0.98
R-SU4	0.84	0.84	0.85	0.87	0.87	0.87	0.99	0.99	0.99	0.99	0.99	0.99
R-SU9	0.84	0.84	0.85	0.87	0.87	0.87	0.99	0.99	0.99	0.99	0.99	0.99
R-W-1.2	0.85	0.85	0.85	0.87	0.87	0.87	0.99	0.99	0.99	0.99	0.99	0.99

Multi documents summarization

	(A1) DUC 2001 100 WORDS MULTI						(A2) DUC 2002 100 WORDS MULTI						(A3) DUC 2003 100 WORDS MULTI					
	1 RFF			3 REFS			1 REF			2 REFS			1 REF			4 REFS		
Method	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP
R-1	0.48	0.56	0.86	0.53	0.57	0.87	0.66	0.66	0.77	0.71	0.71	0.78	0.58	0.57	0.71	0.58	0.57	0.71
R-2	0.55	0.57	0.64	0.59	0.61	0.71	0.83	0.83	0.80	0.88	0.87	0.85	0.69	0.67	0.71	0.79	0.79	0.81
R-3	0.46	0.45	0.47	0.53	0.53	0.55	0.85	0.84	0.76	0.89	0.88	0.83	0.54	0.51	0.48	0.76	0.75	0.74
R-4	0.39	0.39	0.43	0.48	0.49	0.47	0.80	0.80	0.63	0.83	0.82	0.75	0.37	0.36	0.36	0.62	0.61	0.52
R-5	0.38	0.39	0.33	0.47	0.48	0.43	0.73	0.73	0.45	0.73	0.73	0.62	0.25	0.25	0.27	0.45	0.44	0.38
R-6	0.39	0.39	0.20	0.45	0.46	0.39	0.71	0.72	0.38	0.66	0.64	0.46	0.21	0.21	0.26	0.34	0.31	0.29
R-7	0.31	0.31	0.17	0.44	0.44	0.36	0.63	0.65	0.33	0.56	0.53	0.44	0.20	0.20	0.23	0.29	0.27	0.25
R-8	0.18	0.19	0.09	0.40	0.40	0.31	0.55	0.55	0.52	0.50	0.46	0.52	0.18	0.18	0.21	0.23	0.22	0.23
R-9	0.11	0.12	0.06	0.38	0.38	0.28	0.54	0.54	0.52	0.45	0.42	0.52	0.16	0.16	0.19	0.21	0.21	0.21
R-L	0.49	0.49	0.49	0.56	0.56	0.56	0.62	0.62	0.62	0.65	0.65	0.65	0.50	0.50	0.50	0.53	0.53	0.53
R-S*	0.45	0.52	0.84	0.51	0.54	0.86	0.69	0.69	0.77	0.73	0.73	0.79	0.60	0.60	0.67	0.61	0.60	0.70
R-S4	0.46	0.50	0.71	0.54	0.57	0.78	0.79	0.80	0.79	0.84	0.85	0.82	0.63	0.64	0.70	0.73	0.73	0.78
R-S9	0.42	0.49	0.77	0.53	0.56	0.81	0.79	0.80	0.78	0.83	0.84	0.81	0.65	0.65	0.70	0.70	0.70	0.76
R-SU*	0.45	0.52	0.84	0.51	0.54	0.87	0.69	0.69	0.77	0.73	0.73	0.79	0.60	0.59	0.67	0.60	0.60	0.70
R-SU4	0.47	0.53	0.80	0.55	0.58	0.83	0.76	0.76	0.79	0.80	0.81	0.81	0.64	0.64	0.74	0.68	0.68	0.76
R-SU9	0.44	0.50	0.80	0.53	0.57	0.84	0.77	0.78	0.78	0.81	0.82	0.81	0.65	0.65	0.72	0.68	0.68	0.75
R-W-1.2	0.52	0.52	0.52	0.60	0.60	0.60	0.67	0.67	0.67	0.69	0.69	0.69	0.53	0.53	0.53	0.58	0.58	0.58

	(C) DUC02 10			(D1) DUC01 50			(D2) DUC02 50			(E1) DUC01 200			(E2) DUC02 200			(F) DUC01 400		
Method	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP	CASE	STEM	STOP
R-1	0.71	0.68	0.49	0.49	0.49	0.73	0.44	0.48	0.80	0.81	0.81	0.90	0.84	0.84	0.91	0.74	0.73	0.90
R-2	0.82	0.85	0.80	0.43	0.45	0.59	0.47	0.49	0.62	0.84	0.85	0.86	0.93	0.93	0.94	0.88	0.88	0.87
R-3	0.59	0.74	0.75	0.32	0.33	0.39	0.36	0.36	0.45	0.80	0.80	0.81	0.90	0.91	0.91	0.84	0.84	0.82
R-4	0.25	0.36	0.16	0.28	0.26	0.36	0.28	0.28	0.39	0.77	0.78	0.78	0.87	0.88	0.88	0.80	0.80	0.75
R-5	-0.25	-0.25	-0.24	0.30	0.29	0.31	0.28	0.30	0.49	0.77	0.76	0.72	0.82	0.83	0.84	0.77	0.77	0.70
R-6	0.00	0.00	0.00	0.22	0.23	0.41	0.18	0.21	-0.17	0.75	0.75	0.67	0.78	0.79	0.77	0.74	0.74	0.63
R-7	0.00	0.00	0.00	0.26	0.23	0.50	0.11	0.16	0.00	0.72	0.72	0.62	0.72	0.73	0.74	0.70	0.70	0.58
R-8	0.00	0.00	0.00	0.32	0.32	0.34	-0.11	-0.11	0.00	0.68	0.68	0.54	0.71	0.71	0.70	0.66	0.66	0.52
R-9	0.00	0.00	0.00	0.30	0.30	0.34	-0.14	-0.14	0.00	0.64	0.64	0.48	0.70	0.69	0.59	0.63	0.62	0.46
R-L	0.78	0.78	0.78	0.56	0.56	0.56	0.50	0.50	0.50	0.81	0.81	0.81	0.88	0.88	0.88	0.82	0.82	0.82
R-S*	0.83	0.82	0.69	0.46	0.45	0.74	0.46	0.49	0.80	0.80	0.80	0.90	0.84	0.85	0.93	0.75	0.74	0.89
R-S4	0.85	0.86	0.76	0.40	0.41	0.69	0.42	0.44	0.73	0.82	0.82	0.87	0.91	0.91	0.93	0.85	0.85	0.85
R-S9	0.82	0.81	0.69	0.42	0.41	0.72	0.40	0.43	0.78	0.81	0.82	0.86	0.90	0.90	0.92	0.83	0.83	0.84
R-SU*	0.75	0.74	0.56	0.46	0.46	0.74	0.46	0.49	0.80	0.80	0.80	0.90	0.84	0.85	0.93	0.75	0.74	0.89
R-SU4	0.76	0.75	0.58	0.45	0.45	0.72	0.44	0.46	0.78	0.82	0.83	0.89	0.90	0.90	0.93	0.84	0.84	0.88
R-SU9	0.74	0.73	0.56	0.44	0.44	0.73	0.41	0.45	0.79	0.82	0.82	0.88	0.89	0.89	0.92	0.83	0.82	0.87
R-W-1.2	0.78	0.78	0.78	0.56	0.56	0.56	0.51	0.51	0.51	0.84	0.84	0.84	0.90	0.90	0.90	0.86	0.86	0.86

Lin, C.Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).

Document Summarization: Evaluation

- **ROUGE**: **R**ecall-**O**riented **U**nderstudy for **G**isting **E**valuation
  - ✓ Limitations
    - Only assess the content selection and do not account for other quality aspects, such as fluency, grammaticality, coherence, etc.
    - Rely mostly on lexical overlap, while abstractive summarization could express the same content as a reference without any lexical overlap.

[Reference]

슬기로운 의사생활은 사람냄새가 나는 메디컬 드라마이다.

[Generated]

슬의생은 병원이라는 공간에서 발생하는 보통 사람들의 이야기야.

$$\mathrm{ROUGE}-1=\mathrm{ROUGE}-2=\mathrm{ROUGE}-\mathrm{L}=0$$

https://www.slideshare.net/ThoPhan27/abstractive-text-summarization

Document Summarization: Evaluation

- ROUGE 2.0
  - ✓ Attempt to capture **synonymous concepts** and **coverage of topics**

ROUGE 2.0 – A Java Package for Automatic Summary Evaluation

You are here: Home » ROUGE 2.0 – A Java Package for Automatic Summary Evaluation

ROUGE 2.0 is an easy to use evaluation toolkit for Automatic Summarization tasks. It uses the ROUGE system of metrics which works by comparing an automatically produced summary or translation against a set of reference summaries (typically human-produced). ROUGE is one of the standard ways to compute effectiveness of auto generated summaries. To understand how ROUGE works you can read this article.

Features

The latest version of ROUGE 2.0 supports the following:

- Evaluation of ROUGE-N (unigram, bigrams, trigrams, etc)
- Evaluation of ROUGE-L (summary level LCS)
- Evaluation of ROUGE-S and ROUGE-SU (skip-gram and skip-gram with unigrams)
- Evaluation of multiple ROUGE metrics at one go
- Stemming for different languages
- Stopword removal with customizable stop words
- Evaluation of unicode texts (e.g. Persian)
- Minimal formatting requirements for system and reference summaries
- Output in CSV – this makes it super easy for score analysis
- Full documentation and support via GitHub Issues and RxNLP Q&A

http://rxnlp.com/rouge-2-0/

Document Summarization: Evaluation

- ROUGE 2.0
  - ✓ Attempt to capture **synonymous concepts** and **coverage of topics**
- Limitations of ROUGE score: Example

Example 1.1.
System Summary 1 (SysSum1):
Lightweight phone.
Bright screen.
Screen is very clear.

System Summary 2 (SysSum2):
I really love this phone it is just superb, it is extremely lightweight.
Hmmm, this was actually a gift to my girlfriend and I do feel that the screen is quite nice and extremely bright.
In terms of screen, the screen is really clear and crisp.

Reference Summary (RefSum):
The phone is very lightweight.
The display is also very bright and clear.

Table 1
ROUGE-1 scores for Example 1.1

System Summary	ROUGE-N	Recall	Precision	F-Score
SysSum1	ROUGE-1	0.462	0.750	0.571
	ROUGE-1+StopWordRemoval	0.800	0.667	0.727
SysSum2	ROUGE-1	0.692	0.196	0.305
	ROUGE-1+StopWordRemoval	0.800	0.174	0.286

Ganesan, K. (2018). Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. *arXiv preprint arXiv:1803.01937*.

Document Summarization: Evaluation

- Limitations of ROUGE
  - ✓ It does not capture **synonymous concepts**.

Example 1.1.
System Summary 1 (SysSum1):
Lightweight phone.
Bright screen.
Screen is very clear.

System Summary 2 (SysSum2):
I really love this phone it is just superb, it is extremely lightweight.
Hmmm, this was actually a gift to my girlfriend and I do feel that the screen is quite nice and extremely bright.
In terms of screen, the screen is really clear and crisp.

Reference Summary (RefSum):
The phone is very lightweight.
The display is also very bright and clear.

  - ✓ Can be reduced by **allowing synonyms to be captured** during ROUGE scoring.

Document Summarization: Evaluation

- Limitations of ROUGE
  - ✓ It expects system summaries to be **identical** to reference summaries.

Example 1.1.
System Summary 1 (SysSum1):
Lightweight phone.
Bright screen.
Screen is very clear.

System Summary 2 (SysSum2):
I really love this phone it is just superb, it is extremely lightweight.
Hmmm, this was actually a gift to my girlfriend and I do feel that the screen is quite nice and extremely bright.
In terms of screen, the screen is really clear and crisp.

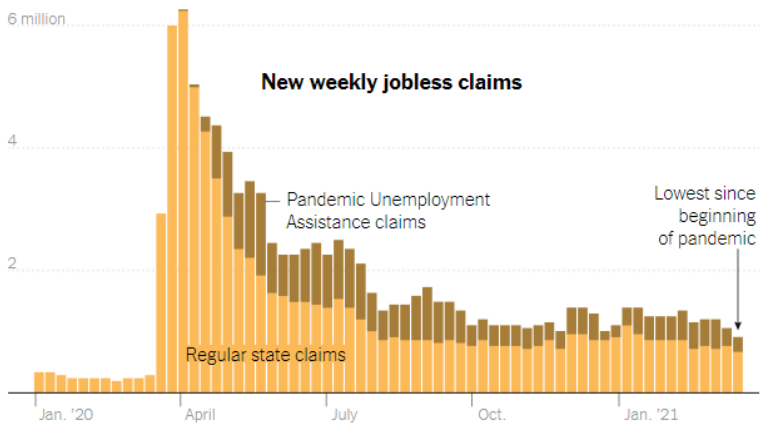
Reference Summary (RefSum):
The phone is very lightweight.
The display is also very bright and clear.

  - ✓ Can be reduced by **allowing synonym capture** as well as **allowing systems to evaluate topic coverage** as opposed to overall content coverage.

Document Summarization: Evaluation

- Limitations of ROUGE
  - ✓ It does not capture **topic or subset coverage**.
    - News: Nouns are more important.
    - Reviews: Adjectives are more important.

Unemployment claims drop to a pandemic low as reopenings offer hope.



Not seasonally adjusted. · Source: Labor Department · By Ella Koeze

While vaccination efforts have gathered speed and restrictions on activities have receded in many states, the job market is showing signs of life.

Initial claims for state unemployment benefits fell last week to 657,000, a decrease of 100,000 from the previous week, the Labor Department reported Thursday. It was the lowest weekly level of initial state claims since the pandemic upended the economy a year ago.

On a seasonally adjusted basis, new state claims totaled 684,000.

In addition, there were 242,000 new claims for Pandemic Unemployment Assistance, a federal program covering freelancers, part-timers and others who do not routinely qualify for state benefits, a decrease of 43,000.

https://www.nytimes.com/live/2021/03/25/business/stock-market-today

Document Summarization: Evaluation

- ROUGE 2.0
  - ✓ ROUGE-{NN | Topic | TopicUniq} + Synonyms
    - Wordnet is used to obtain synonyms for nouns, verbs, and adjectives.
    - Are "screen" and "display" synonymous in the Wordnet?

Example 1.1.
System Summary 1 (SysSum1):
Lightweight phone.
Bright screen.
Screen is very clear.

System Summary 2 (SysSum2):
I really love this phone it is just superb, it is extremely lightweight.
Hmmm, this was actually a gift to my girlfriend and I do feel that the screen is quite nice and extremely bright.
In terms of screen, the screen is really clear and crisp.

Reference Summary (RefSum):
The phone is very lightweight.
The display is also very bright and clear.

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: display [Search WordNet]

Display Options: (Select option to change) [Change]
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss)

Noun

- S: (n) **display**, show (something intended to communicate a particular impression)
- S: (n) **display**, exhibit, showing (something shown to the public)
- S: (n) **display**, presentation (a visual representation of something)
- S: (n) **display** (behavior that makes your feelings public)
- S: (n) **display** (exhibiting openly in public view)
- S: (n) **display**, video display (an electronic device that represents information in visual form)

Verb

- S: (v) expose, exhibit, **display** (to show, make visible or apparent)
- S: (v) **display** (attract attention by displaying some body part or posing; of animals)

http://wordnetweb.princeton.edu/perl/webwn

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: screen [Search WordNet]

Display Options: (Select option to change) [Change]
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss)

Noun

- S: (n) **screen**, silver screen, projection screen (a white or silvered surface where pictures can be projected for viewing)
- S: (n) blind, **screen** (a protective covering that keeps things out or hinders sight)
- S: (n) **screen**, CRT screen (the display that is electronically created on the surface of the large end of a cathode-ray tube)
- S: (n) **screen**, cover, covert, concealment (a covering that serves to conceal or shelter something)
- S: (n) **screen** (a protective covering consisting of netting; can be mounted in a frame)
- S: (n) filmdom, screenland, **screen** (the personnel of the film industry)
- S: (n) sieve, **screen** (a strainer for separating lumps from powdered material or grading particles)
- S: (n) **screen** (partition consisting of a decorative frame or panel that serves to divide a space)

Verb

- S: (v) **screen**, test (test or examine for the presence of disease or infection)
- S: (v) **screen** (examine methodically)
- S: (v) **screen**, screen out, sieve, sort (examine in order to test suitability)
- S: (v) **screen** (project onto a screen for viewing)
- S: (v) **screen**, block out (prevent from entering)
- S: (v) riddle, **screen** (separate with a riddle, as grain from chaff)
- S: (v) shield, **screen** (protect, hide, or conceal from danger or harm)

Table 2
ROUGE-1 scores for Example 1.1 with the use Synonyms in ROUGE scoring.

		ROUGE Scoring Type	Recall	Precision	F-Score
1	SysSum1	ROUGE-1	0.462	0.750	0.571
2		ROUGE-1 + Synonyms	0.538	0.875	**0.667**
3		ROUGE-1 + StopWordRemoval	0.800	0.667	0.727
4		ROUGE-1 + StopWordRemoval + Synonyms	1.000	0.833	**0.909**
		ROUGE Scoring Type	**Recall**	**Precision**	**F-Score**
5	SysSum2	ROUGE-1	0.692	0.196	0.305
6		ROUGE-1 + Synonyms	0.769	0.217	**0.339**
7		ROUGE-1 + StopWordRemoval	0.800	0.174	0.286
8		ROUGE-1 + StopWordRemoval + Synonyms	1.000	0.217	**0.357**

Document Summarization: Evaluation

- ROUGE 2.0
  - ✓ ROUGE Topic provides the ability to evaluate different dimensions (i.e. topics) of a summary.
    - News summarization: all the nouns in the reference summaries
    - Opinion: all the nouns and the adjectives.
    - Allow users to specify which Part of Speech (POS) combinations should be used for evaluation
    - POS tags are based on the Standord's POS Tagger

Table 3
Part-of-speech options for ROUGE-Topic and ROUGE-TopicUniq scoring. Note that multiple POS options can be used concurrently and this is only a subset of POS tags that can be used. Any POS tag supported by the Stanford's POS tagger may be specified.

POS Options	Description
JJ	All types of adjectives
VB	All verbs
NN	All nouns including proper nouns
VBD	Verbs in past tense form
RB	Adverbs
NNP	Proper Nouns

$$ROUGE - Topic_{recall} = \frac{\sum Overlap(REF_{i_{pos}}, SYS_{j_{pos}})}{|REF_{i_{pos}}|}$$

$$ROUGE - Topic_{precision} = \frac{\sum Overlap(REF_{i_{pos}}, SYS_{j_{pos}})}{|SYS_{j_{pos}}|}$$

$$ROUGE - TopicUniq_{recall} = \frac{REFUniq_{i_{pos}} \cap SYSUniq_{j_{pos}}}{|REFUniq_{i_{pos}}|}$$

$$ROUGE - TopicUniq_{precision} = \frac{REFUniq_{i_{pos}} \cap SYSUniq_{j_{pos}}}{|SYSUniq_{j_{pos}}|}$$

Document Summarization: Evaluation

- ROUGE 2.0
  - ✓ ROUGE-Topic Example

Example 1.1.
System Summary 1 (SysSum1):
Lightweight phone.
Bright screen.
Screen is very clear.

System Summary 2 (SysSum2):
I really love this phone it is just superb, it is extremely lightweight.
Hmmm, this was actually a gift to my girlfriend and I do feel that the screen is quite nice and extremely bright.
In terms of screen, the screen is really clear and crisp.

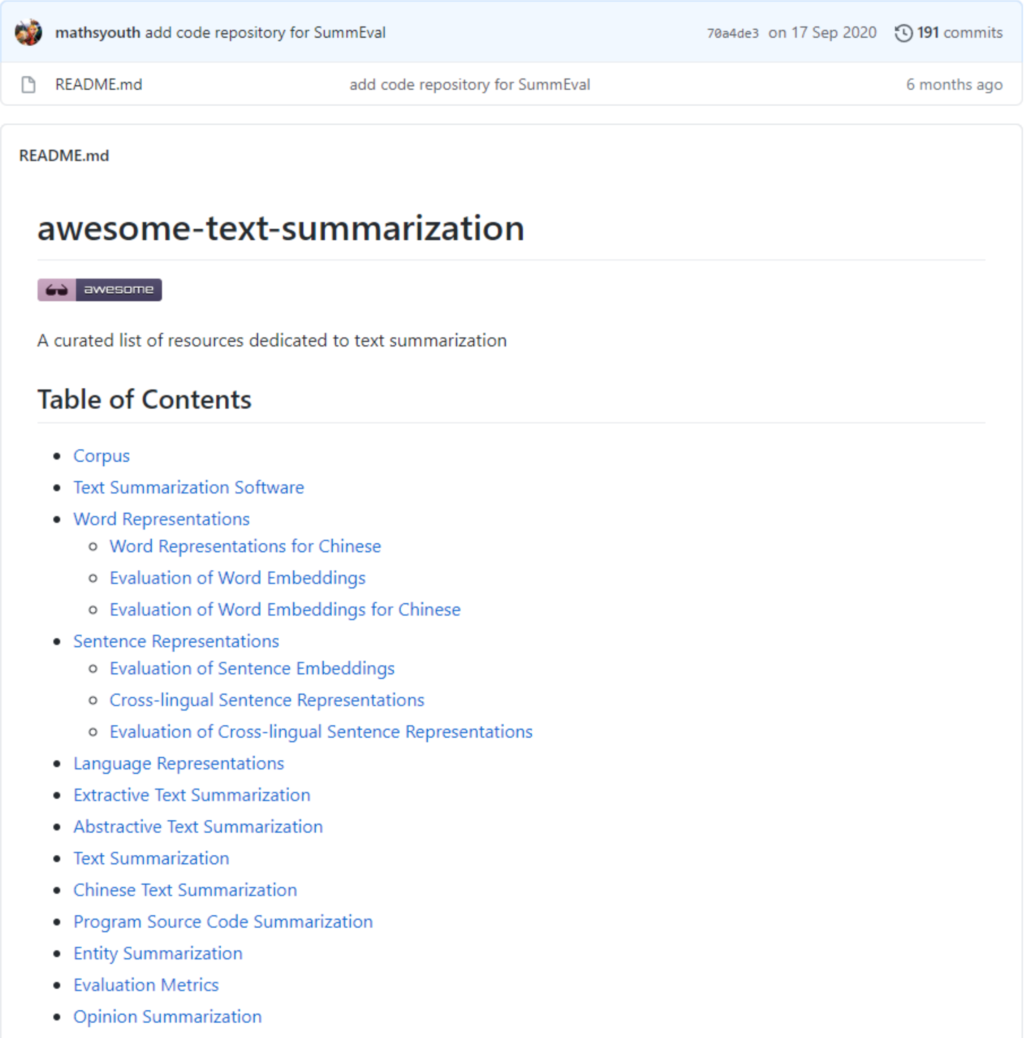
Reference Summary (RefSum):
The phone is very lightweight.
The display is also very bright and clear.

Table 4
ROUGE-TopicNN | JJ and ROUGE-TopicUniqNN | JJ scores based on Example 1.1

		RougeTopic	**Recall**	**Precision**	**F-Score**
1	SysSum1	ROUGE-TopicNN | JJ	0.800	0.667	0.727
2		ROUGE-TopicNN | JJ + Synonyms	**1.000**	**0.833**	**0.909**
3		ROUGE-TopicUniqNN | JJ	0.800	0.800	0.800
4		ROUGE-TopicUniqNN | JJ + Synonyms	**1.000**	**1.000**	**1.000**
		RougeTopic	**Recall**	**Precision**	**F-Score**
5	SysSum2	ROUGE-TopicNN | JJ	0.800	0.308	0.444
6		ROUGE-TopicNN | JJ + Synonyms	**1.000**	**0.385**	**0.556**
7		ROUGE-TopicUniqNN | JJ	0.800	0.364	0.500
8		ROUGE-TopicUniqNN | JJ + Synonyms	**1.000**	**0.455**	**0.625**

Document Summarization: Study Materials

- Curated Paper List: https://github.com/mathsyouth/awesome-text-summarization


mathsyouth add code repository for SummEval 70a4de3 on 17 Sep 2020 191 commits

README.md add code repository for SummEval 6 months ago

README.md

awesome-text-summarization

awesome

A curated list of resources dedicated to text summarization

Table of Contents

- Corpus
- Text Summarization Software
- Word Representations
  - Word Representations for Chinese
  - Evaluation of Word Embeddings
  - Evaluation of Word Embeddings for Chinese
- Sentence Representations
  - Evaluation of Sentence Embeddings
  - Cross-lingual Sentence Representations
  - Evaluation of Cross-lingual Sentence Representations
- Language Representations
- Extractive Text Summarization
- Abstractive Text Summarization
- Text Summarization
- Chinese Text Summarization
- Program Source Code Summarization
- Entity Summarization
- Evaluation Metrics
- Opinion Summarization


ANY
questions?