

Paper2Speech

2조

김도현, 김은성, 박진성, 송기훈

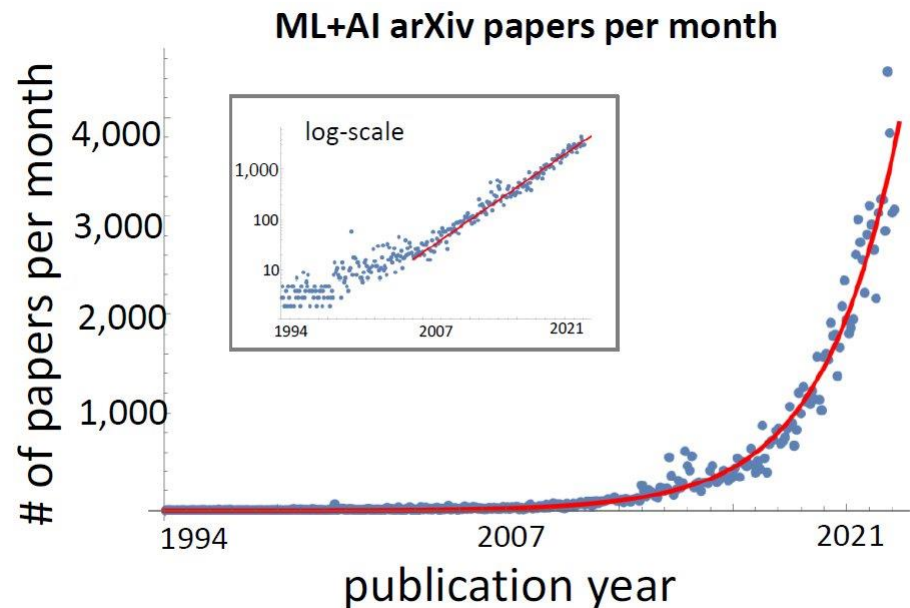
Table of Content

목차

1. Background, Subject
2. Explanation of Data
3. Explanation of Model
4. Analysis
5. Conclusion

배경 및 주제 설명

- 인공지능 학습 특징
 - 매년, 매 분기 다양한 구조의 모델이 발표됨
 - 이 때문에 정해진 하나의 교재가 없음
 - 좋은 결과를 낸 연구들의 논문으로 학습
- 학습 시 느끼는 어려운 점
 - 논문의 수가 급격하게 늘어남
 - 어떤 논문이 어떤 정보를 가는지 알기 힘들
 - 모든 논문이 영어이므로 읽기 힘들



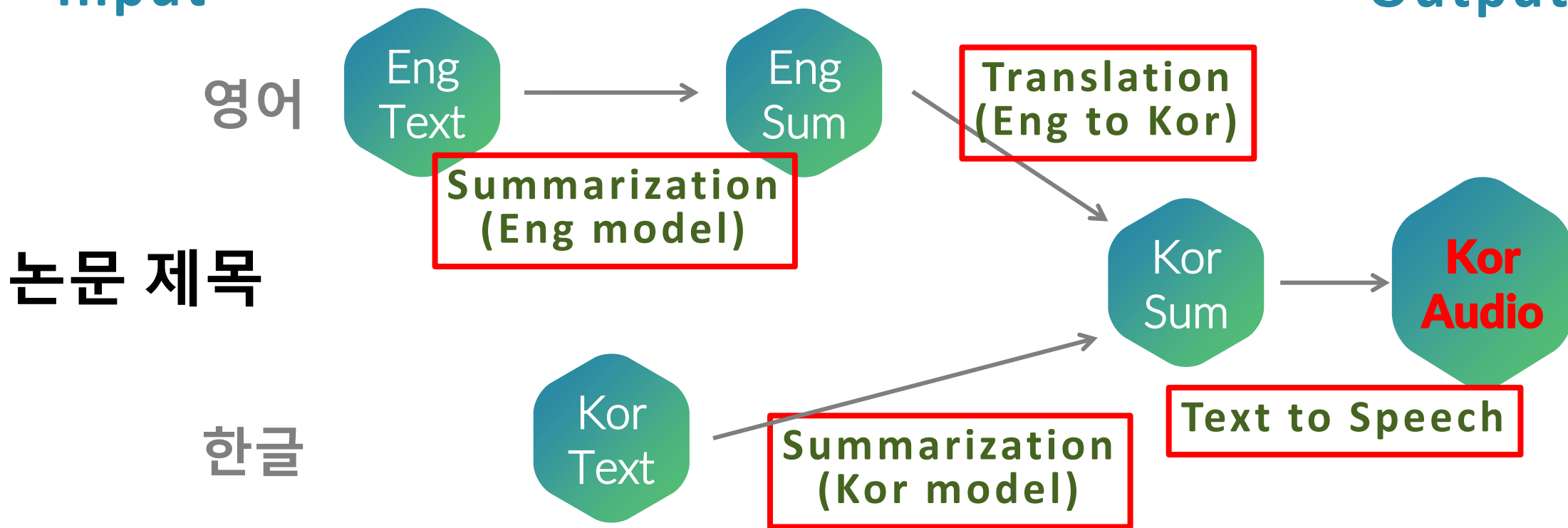
<논문 수집 사이트인 arXiv에서 구한 연도별 ML+AI 논문 수>

Our Subject / Model

논문의 정보를 쉽게 파악할 수 있도록 **한국어 음성데이터**로 변환시켜주자!

Input

Output



Explanation of Data

데이터 설명

- Fine-tuning Data (한글, 영어 논문)

: [Ai-hub](#), [GitHub](#)로부터 수집한 데이터 사용하여 모델 Fine-tuning

- 한글 : 약 3만개의 원문 / Summary

- 영어 : 약 2천개의 원문 / Summary

Papers With Code

- 학습 후 모델에 사용하는 Data

: [Papers With Code](#)(머신러닝 관련 논문과 함께 코드가 제공되는 사이트) -> Crawling

AI  Hub



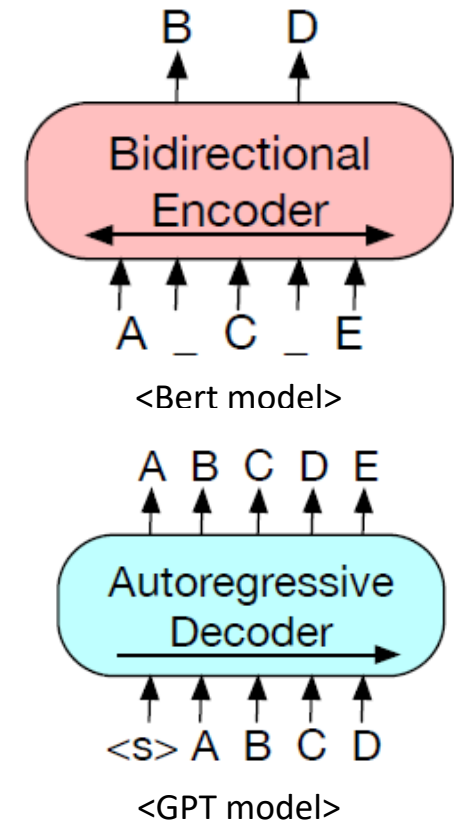
모델 설명

- Bert(Bidirectional Encoder Representation from Transformer)

- 기존 Transformer 구조 중 Encoder로만 구성
- Masked Language Model
- 문서의 의미 추출에 강점

- GPT(Generative Pre-trained Transformer)

- 기존 Transformer 구조 중 Decoder로만 구성
- AutoRegressive Language Model
- 문장 생성에 강점



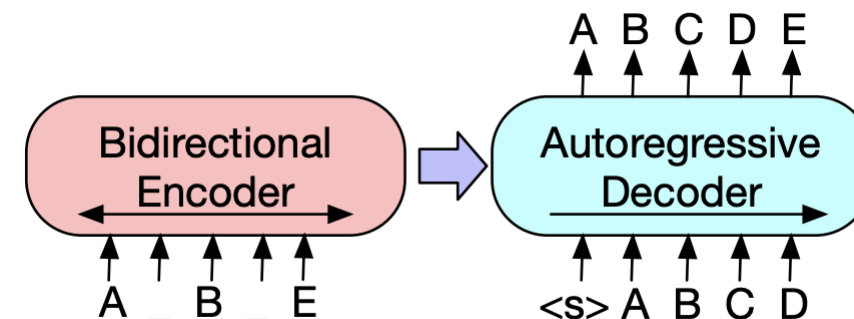


- Summarization Model(English)

Facebook/**BART model**

- Bart(Bidirectional Auto-Regressive Transformers)

- Encoder와 Decoder를 가진 기존의 seq2seq 구조
- Self-Supervised learning을 통한 사전 훈련
- 문장에 다양한 noising(임의 길이 텍스트 mask, 순서 변경 등)을 하여 사전 훈련됨
- 다양한 분야에서 Sota(State of the art) 달성














<BART model>

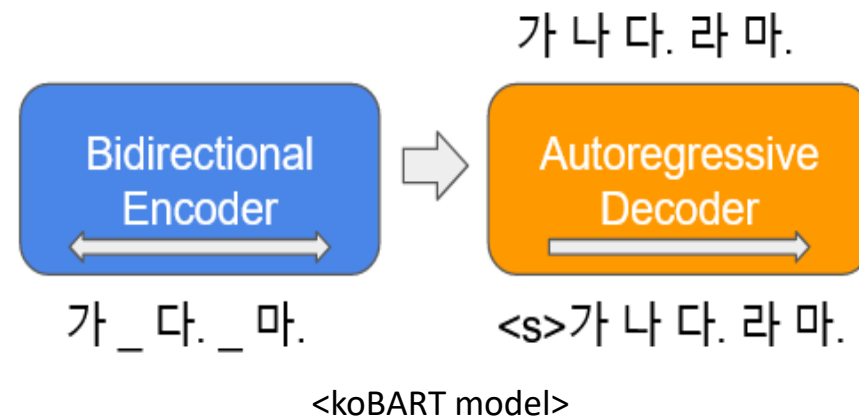
Explanation of Model

모델 설명

Text Summarization on arXiv

Rank	Model	ROUGE- 1	ROUGE- 2	ROUGE- L	Extra Training Data	Paper	Code	Result	Year	Tags
1	Top Down Transformer (AdaPool)	50.95	21.93	45.61	×	Long Document Summarization with Top-down and Bottom-up Inference			2022	
2	BART-LS	50.2			×	Adapting Pretrained Text-to-Text Models for Long Text Sequences			2022	
3	Pegasus-X	50.0	21.8	44.6	×	Investigating Efficiently Extending Transformers for Long Input Summarization			2022	
4	ExtSum + oracle segmentation (extractive)	49.49	21.04	44.34	×	Document Summarization with Text Segmentation			2023	
5	FactorSum	49.32	20.27	44.76	×	Factorizing Content and Budget Decisions in Abstractive Summarization of Long Documents			2022	
6	ExtSum + supervised segmentation (extractive)	49.11	20.68	44.01	×	Document Summarization with Text Segmentation			2023	
7	Lodoss-full-large (extractive)	48.45	20.72	42.55	×	Toward Unifying Text Segmentation and Long Document Summarization			2022	

- Summarization Model(Korean) SKT-AI/KoBART model



Data

Data	# of Sentences
Korean Wiki	5M
Other corpus	0.27B

Model

Model	# of params	Type	# of layers	# of heads	ffn_dim	hidden_dims
KoBART-base	124M	Encoder	6	16	3072	768
		Decoder	6	16	3072	768



- Translation Model(Eng to Kor)

Facebook / **nllb model**(nllb-200-distilled-600M)

Deep learning is being used to detect malware in domain names. A paper shows that there is surprisingly little difference between various convolutional neural network (CNN) and recurrent neuralNetwork (RNN) based architectures in terms of accuracy. The simpler architectures are faster to train and less prone to overfitting

딥러닝은 도메인 이름에서 악성코드를 탐지하는 데 사용되고 있다. 한 논문은 정확도 측면에서 다양한 컨볼루션 신경망 (CNN)과 반복 신경망(RNN) 기반 아키텍처 사이에 놀라운 정도로 거의 차이가 없다는 것을 보여준다. 단순한 아키텍처는 교육 속도가 빠르고 과적합이 적습니다.

<Papago>

'딥러닝은 도메인 이름에서 악성코드를 탐지하는 데 사용됩니다. 이하 논문에서는 다양한 convolutional neural network (CNN) 와 recurrent neuralNetwork (RNN) 기반의 아키텍처의 정확성 차원에서 놀라운 차이가 거의 없다는 것을 보여줍니다. 이보다 간단한 아키텍처는 훈련이 더 빨라지고 과잉 부합이 덜 쉽다.'

<Facebook / nllb model>



Deep learning is being used to detect malware in domain names. A paper shows that there is surprisingly little difference between various convolutional neural network (CNN) and recurrent neuralNetwork (RNN) based architectures in terms of accuracy. The simpler architectures are faster to train and less prone to overfitting

1. 번역하려는 문장의 특성과 관련된 요인들 (ex : 구조적 차이, 어휘, 문화 차이, 전문용어 등)

Translated text :

심층 통신 (心通信, 영어: Deep recommendation) 또는 심층 통신 (心通信, 영어: 心通信, 문화어: 심층 전화) 은 통신망의 한 형태로, 전화 통신망이나 통신망을 구성하는 단위이다.

2. 번역할 문장의 입력과 출력 길이가 모델 설정에 따라 제한되어 있는 경우

Translated text :

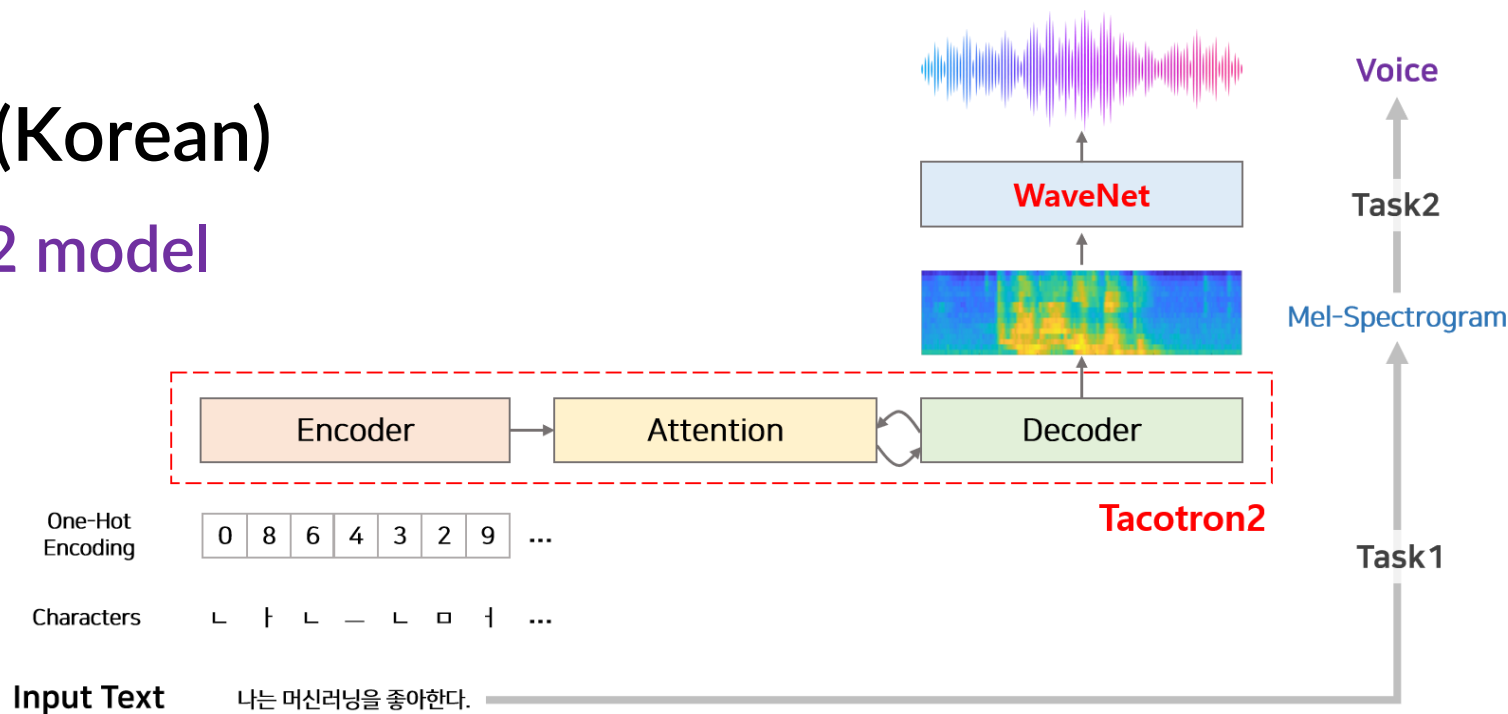
딥 러닝, 인기 있는, 녹음 심장, 예, 값, + 같은 문장 반복



- Text to Speech Model(Korean)

Tensorspeech/**Tacotron2** model

2018년도에 개발된
Attention 기반 TTS 모델



<Tacotron2 model>

Text 데이터를 음성 Spectrogram 데이터로 변환

-> WaveNet 모델을 거쳐 Audio 데이터로 변환

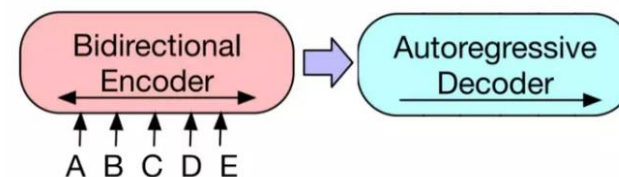
분석 - Fine-tuning

- Fine-tuning of BART

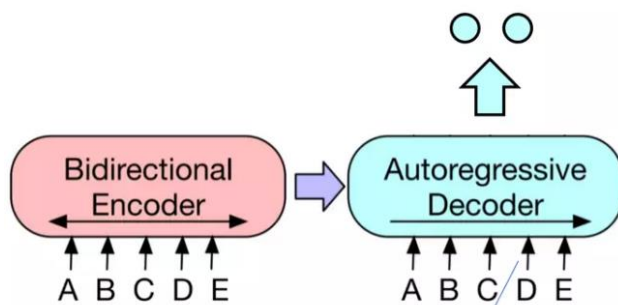
Sequence Generation Task(문장 생성)

: Summarization Task에서 사용되는 파인튜닝 기법

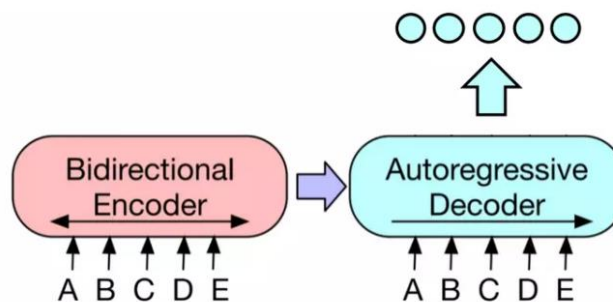
Sequence Generation Tasks



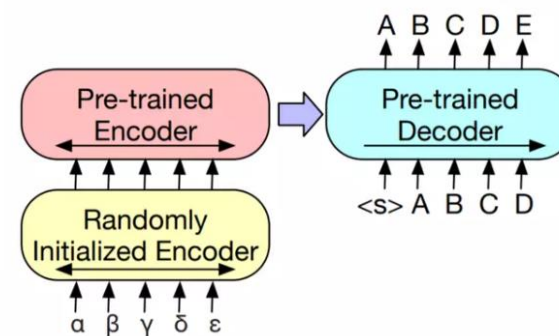
Sentence Classification Tasks



Token Classification Tasks



Machine Translation Tasks





- Fine-tuning : Summarization(Eng / Kor model)
 - Pretrained : Facebook / **bart-large-cnn** , SKT AI / **KoBart-base**

philschmid/bart-large-cnn-samsum

Updated Dec 24, 2022 • ↓ 2.66M • ♥ 117

facebook/bart-large-cnn

Updated Jan 25 • ↓ 833k • ♥ 280

lidiya/bart-large-xsum-samsum

Updated Jul 20, 2022 • ↓ 93.1k • ♥ 22

facebook/bart-large-xsum

Updated Jan 25 • ↓ 76.9k • ♥ 17

knkarthick/MEETING-SUMMARY-BART-LARGE-XSUM-SAMSUM-DIALOG...

Updated Dec 14, 2022 • ↓ 12.4k • ♥ 6

 TensorFlow

 PyTorch

- Fine-tuning : Summarization(Eng model)
 - Rouge Score 비교

```
{ 'rouge1' : 31.701870027485302,  
  'rouge2' : 9.588183543416209,  
  'rougeL' : 19.311453116515825,  
  'rougeLsum' : 19.343632731439115}
```

파인 튜닝 전

```
{ 'rouge1' : 33.87794546055859,  
  'rouge2' : 11.523685278438423,  
  'rougeL' : 22.44646937655922,  
  'rougeLsum' : 22.463109948956216}
```

파인 튜닝 후

- 실제 요약문 비교 (Bart Model)

Pre-trained Bart Model	Fine-tuned Bart Model
<p>The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train.</p>	<p>We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. The Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.</p>

- 실제 요약문 비교 (KoBart Model)

Pre-trained KoBart Model	Fine-tuned KoBart Model
<p>타식무경이라는 교의를 고수했던 중국 법상종은 '타인의 마음'의 의미를 고찰하며 타인의 마음을 아는 지식'이 성립한다는 사실 자체가 또 다른 외경을 긍정하는 논리가 될 수 있기 때문에 '타인의 마음'의 교의를 정당화한다고 말했다.</p>	<p>'유식무경'이라는 교의를 고수했던 중국 법상종의 입장에서 타인의 마음(他心)의 의미를 고찰한 '타심지(他心智)'의 문제를 다루었는데, 타인의 마음이란 그와 분리되지 않는 모든 것[諸法]을 뜻하기 때문에, 타인의 마음도 타인이 향유하는 세계 전체를 포괄하는 개념으로 전환된다.</p>

분석 - 한계점 & 개선할 사항

- 한계점

Fine-tuning

: 모든 모델에 대해 Fine-tuning을 실시하지는 못함

-> 번역 or TTS 모델이 인공지능 논문의 특성에 맞게 적용되지는 않음

ex) Attention -> 주의, Sequence-to-Sequence -> 순서대로

분석 - 한계점 & 개선할 사항

- 한계점

Full-text를 요약하는 모델을 구현해내지 못함

-> 입력 Data : 논문의 Abstract text

- 이미 핵심 내용인 Abstract를 다시 요약하는 것에 대한 효용성 고민

분석 - 한계점 & 개선할 사항

- 개선할 사항

- 정확한 Paper 제목을 입력해야 결과가 출력됨
- 현재 한글 논문 Crawling은 실행되지 않음
- 제목 데이터도 음성으로 변환 모델 추가

Input

논문 제목

“Attention is all you need”

“Bart Denoising
Sequence-to-Sequence”



Model



Output



<Attention Paper>



<Bart Paper>

감사합니다