

# 딥러닝을 이용한 한국어-영어 번역기 만들기

여기에 텍스트 입력

김한울 문서연 박주희 전계동 차정현 최지웅

# 차례

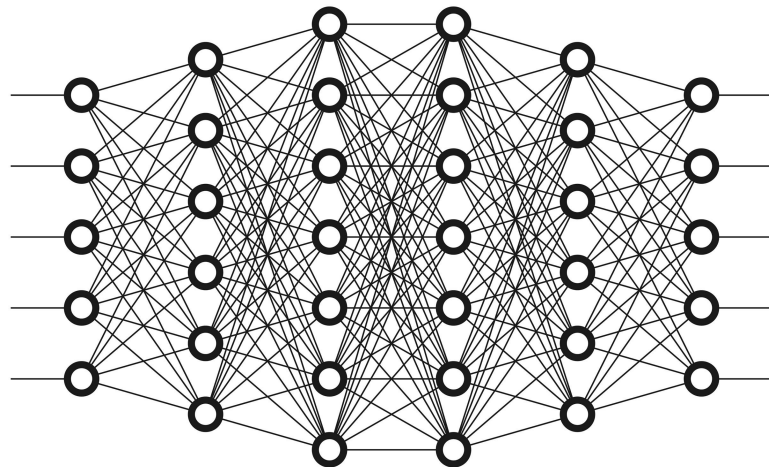
0. 프로젝트 소개

1. 수행 과정

2. 결과

# 한국어-영어 번역

안녕



hello

## AI 데이터

소개

공공/법률

과학기술/정보통신

교육/문화/스포츠

K-Fashion 이미지

감성대화 말뭉치

관광

랜드마크 이미지

멀티모달

멀티모달 영상

시각정보 기반 질의응답

전문분야 한-영 말뭉치

한국어 글자체 이미지

한국어-영어 번역 말뭉치

🏠 > AI데이터 > 한국어-영어 번역(병렬) 말뭉치 AI데이터 다운로드

## 한국어-영어 번역(병렬) 말뭉치 AI데이터 다운로드

소개

다운로드

저작도구

다운로드

문의하기

한국어-영어 말뭉치 다운로드

전체 선택

선택 해제

다운로드

구어체(1) 다운로드

구어체(2) 다운로드

대화체 다운로드

문어체-뉴스(1) 다운로드

문어체-뉴스(2) 다운로드

문어체-뉴스(3) 다운로드

문어체-뉴스(4) 다운로드

문어체-한국문화 다운로드

문어체-조례 다운로드

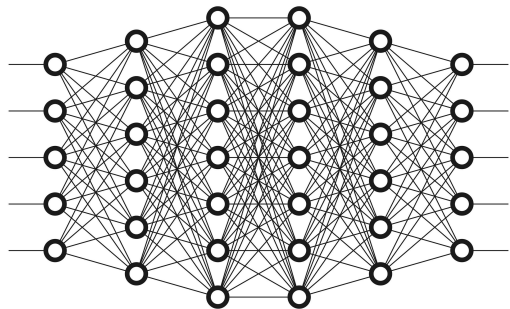
문어체-지자체웹사이트 다운로드

## 구어체 문장만 40만개

399991	당신이 마침내 고백했다니 나는 기쁘군요.	I'm glad that you finally made the confession.
399992	당신이 힘들어하니 마음이 안좋네요.	I feel bad for you having a hard time.
399993	매니저와 문제가 있으면 내게 알려줘요.	If you have any problem with the manager, let me know.
399994	어려움이 있다면 다시 함께 해봅시다.	Let's do it again if you have any difficulties.
399995	당신은 이제 곧 잠자리에 들어야할 것 같아요.	I think you should go to bed soon.
399996	당신은 곧 당신의 조카를 볼 수 있어요.	You can see your nephew soon.
399997	나는 당신이 내게 공유한 음악을 들었어요.	I've listened to the music that you shared with me.
399998	그 서류는 당신이 공유해 준 것과 달라요.	The document is different from what you shared.
399999	당신이 잡지에 광고 낸 것을 봤어요.	I saw your advertisement in the magazine.
400000	당신이 이제 괜찮다니 저는 참 기쁘군요.	I'm really glad you're all right now.

# 프로젝트 순서

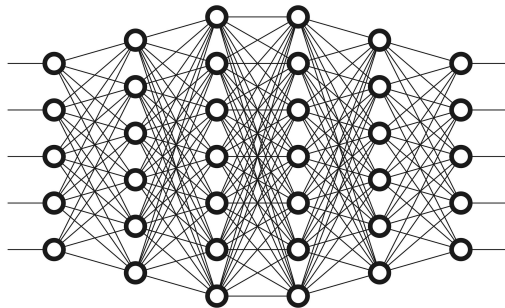
hello



bonjour



안녕



hello

## BLEU Score

Table 2: The Transformer achieves better BLEU scores than previous state-of-the-art models on the English-to-German and English-to-French newstest2014 tests at a fraction of the training cost.

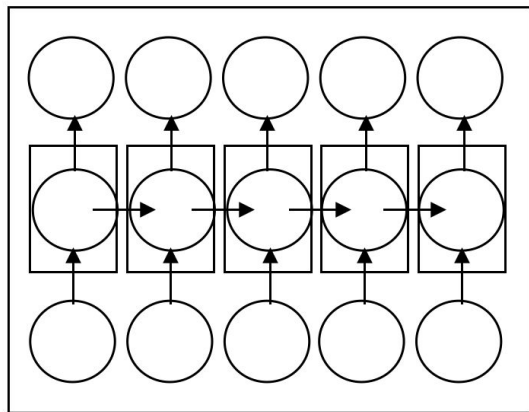
Model	BLEU		Training Cost (in FLOPS * $10^{18}$ )	
	EN-DE	EN-FR	EN-DE	EN-FR
ByteNet (Kalchbrenner et al., 2016)	23.75			
Deep-Att + PosUnk (Zhou et al., 2016)		39.2		100
GNMT + RL (Wu et al., 2016)	24.6	39.92	23	140
ConvS2S (Gehring et al., 2017)	25.16	40.46	9.6	150
MoE (Shazeer et al., 2017)	26.03	40.56	20	120
GNMT + RL Ensemble (Wu et al., 2016)	26.30	41.16	180	1100
ConvS2S Ensemble (Gehring et al., 2017)	26.36	<b>41.29</b>	77	1200
Transformer (base model)	27.3	38.1	<b>3.3</b>	
Transformer (big)	<b>28.4</b>	<b>41.8</b>	23	



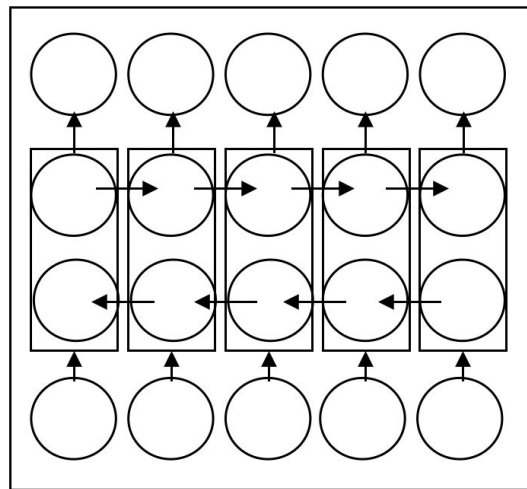
Google Colaboratory



# 모델: RNN과 양방향Bidirectional RNN



(a)



(b)

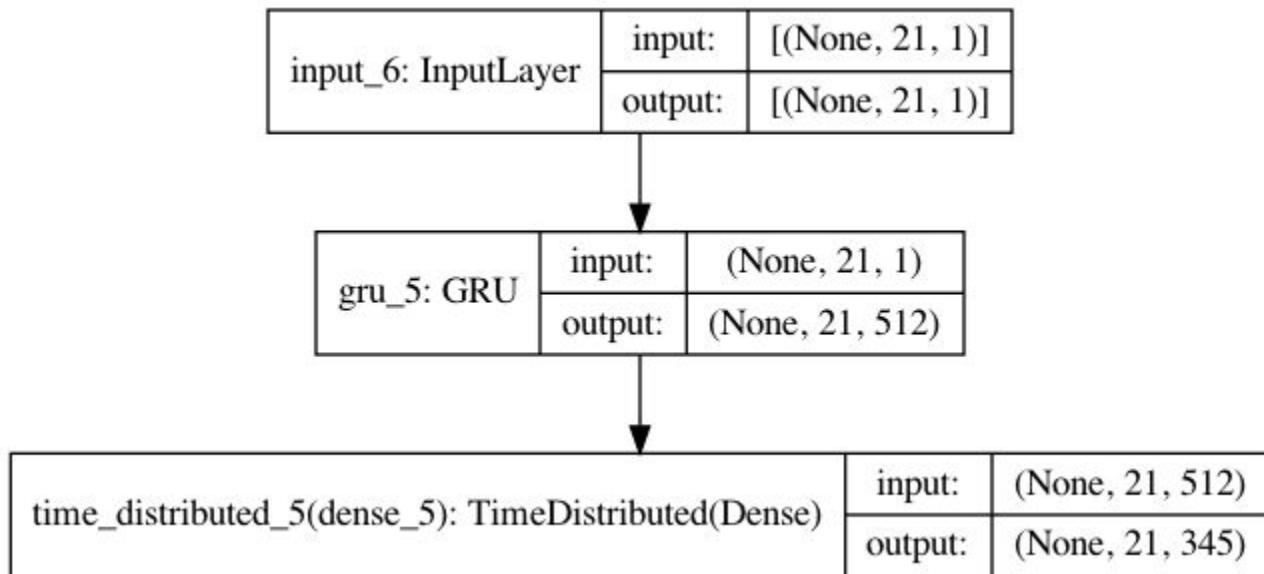
Structure overview

(a) unidirectional RNN

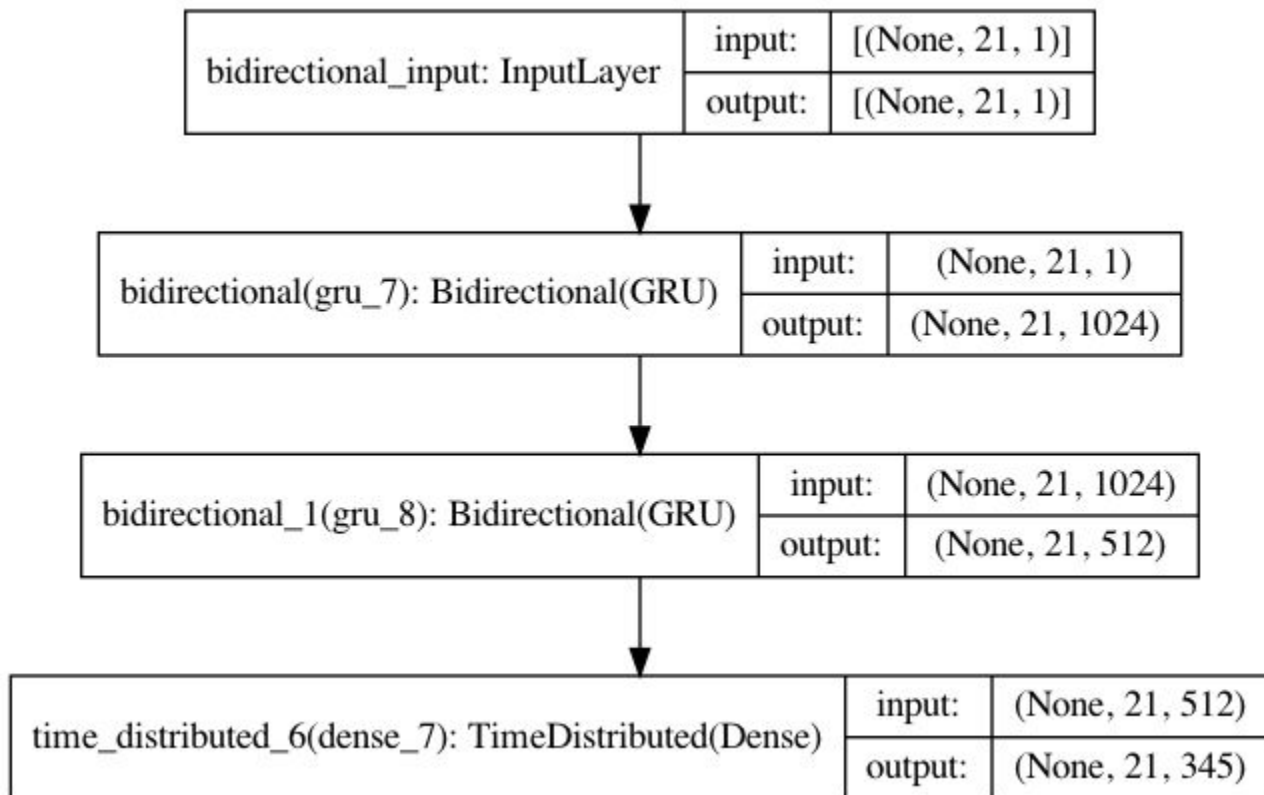
(b) bidirectional RNN

출처: 위키백과 Bidirectional recurrent neural networks

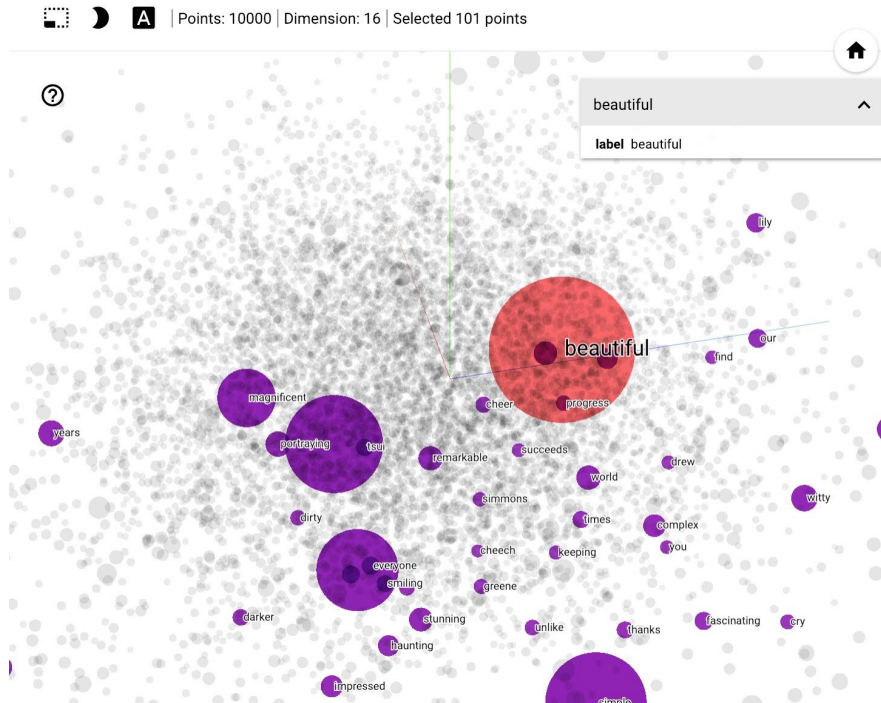
# 모델: RNN



## 모델: 양방향 RNN

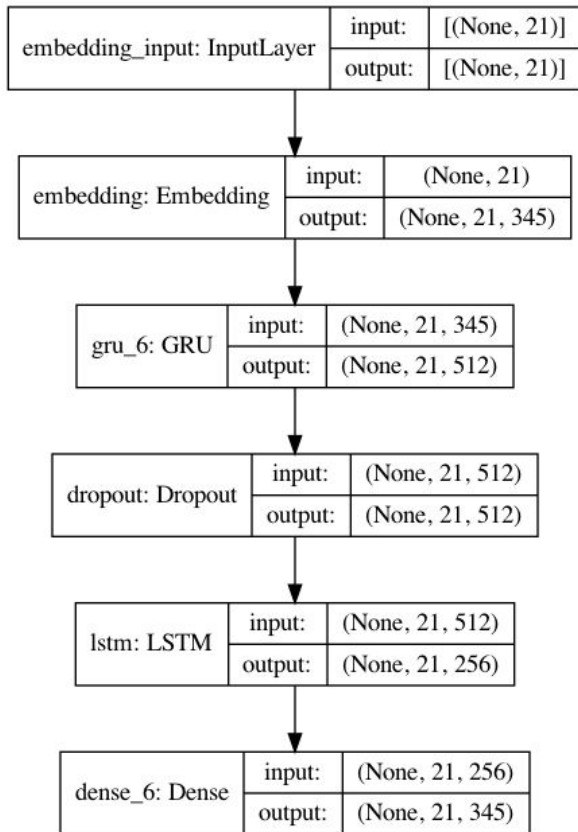


# 모델: 임베딩 Embedding

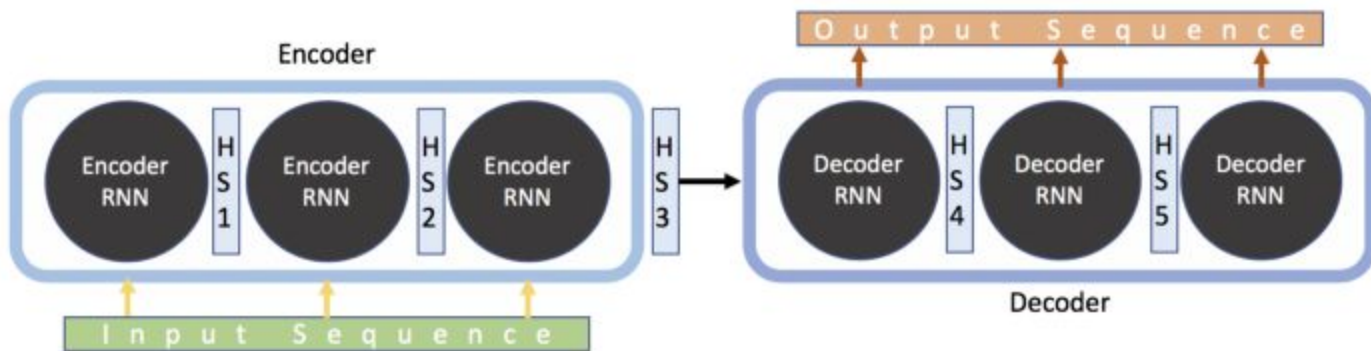


출처: TensorFlow Core 튜토리얼 단어 임베딩

# 모델: 임베딩Embedding



# 모델: 인코더Encoder-디코더Decoder



출처: [towardsdatascience.com](https://towardsdatascience.com/attention-in-seq2seq-models) Attention — Seq2Seq Models

# 모델별 Hyperparameter

Model	Hyperparameter
RNN	GRU units: 512
임베딩Embedding	GRU units: 512, dropout:0.1, LSTM units: 256
양방향 RNN	GRU units: 512, 256
임베딩+양방향 RNN 어텐션attention	GRU units:512, LSTM units: 64, 64

**[공통]** 학습 batch\_size: 1024, epochs: 10, validation split: 0.2, return\_seq:True

# 영어-프랑스어 번역기 제작 결과

Model	validation accuracy	BLEU 점수
RNN	0.699	24.4
임베딩Embedding	0.921	56.6
양방향 RNN	0.824	46.8
인코더-디코더	0.981	
어텐션attention	0.999	



# 영어-프랑스어 문장 번역 결과 비교

원문	he saw a old yellow truck	that cat was my most loved animal
번역문	il a vu un vieux camion jaune	ce chat était mon animal le plus aimé
RNN	elle a conduit une voiture voiture	ils pense était son <b>animal le plus aimé</b>
양방향 RNN	il a conduit une voiture voiture	ce oiseau <b>était</b> votre <b>animal le plus aimé</b>
임베딩Embedding	il a vu une vieille <b>camion jaune</b>	ce chat <b>était</b> mon <b>animal le plus aimé</b>
인코더-디코더	il a vu un vieux camion rouge	ce lapin <b>était</b> mon <b>animal le plus aimé</b>

## 인코더-디코더 모델: 한국어-영어 문장 번역 결과

우리는 언제나 서로를 사랑하려고 노력해요 .

We always try and show our love for each other.

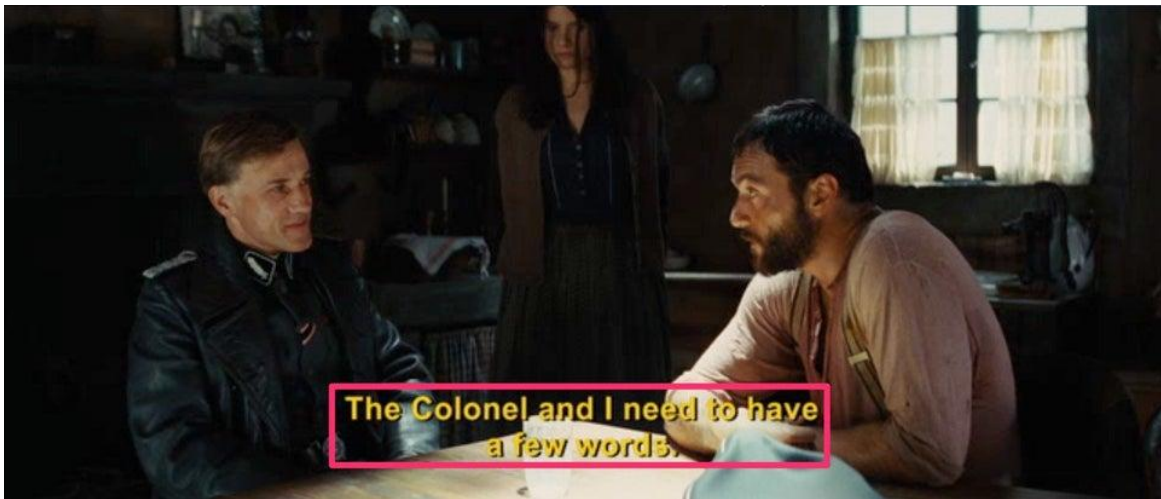
They always try to love each other.

[illegible]

# 향후 활용 방안



hello



여기에 텍스트 입력

김한울 문서연 박주희 전계동 차정현 최지웅