NATURAL LANGUAGE PROCESSING

LECTURE 11: Applications

goorm

KAIST AI
Graduate School of AI

DAVIAN

INDEX

How to exploit model?

- Text Classification
  - Sentiment Analysis
- Token Classification
  - NER
  - QA
- Similarity Measure
  - Retrieval

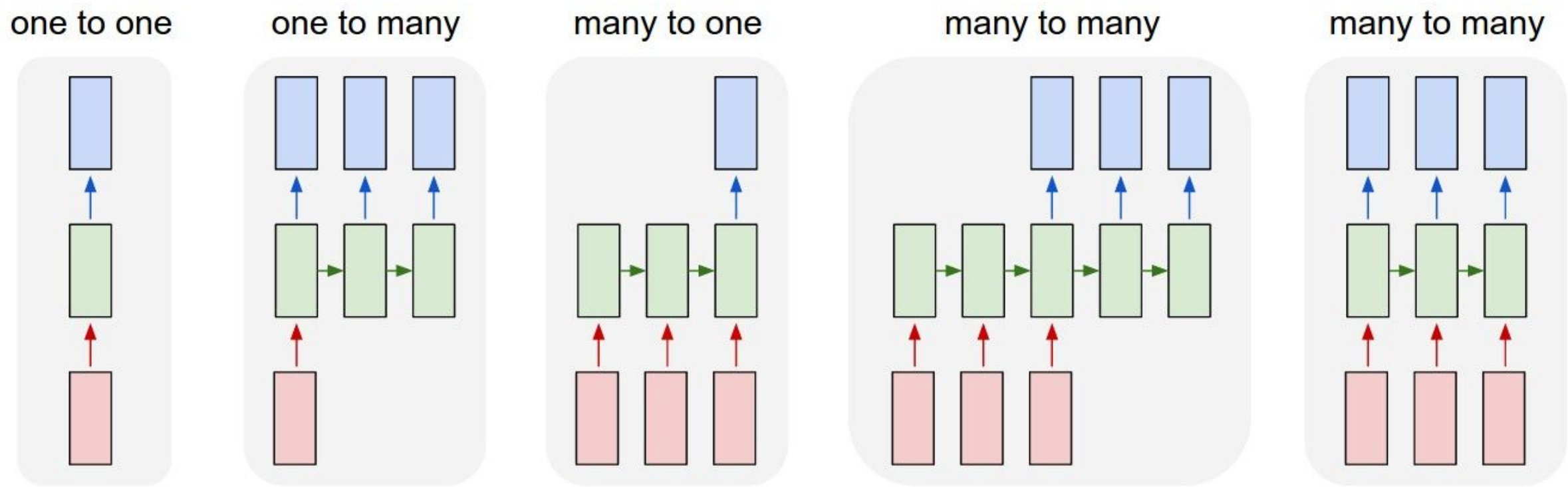
NLP Roadmap

Task	Formulation	Model	Learning
Sentiment Classification	Text Classification	RNNs	Vanilla
Machine Reading	Token Classification	Encoder-Decoder	Pretraining & Finetuning
Machine Translation	Retrieval	Attention	In-context Learning
Language Model	Text Generation	Transformer	
...			

https://seominjoon.github.io/kaist-ai605/schedule.html

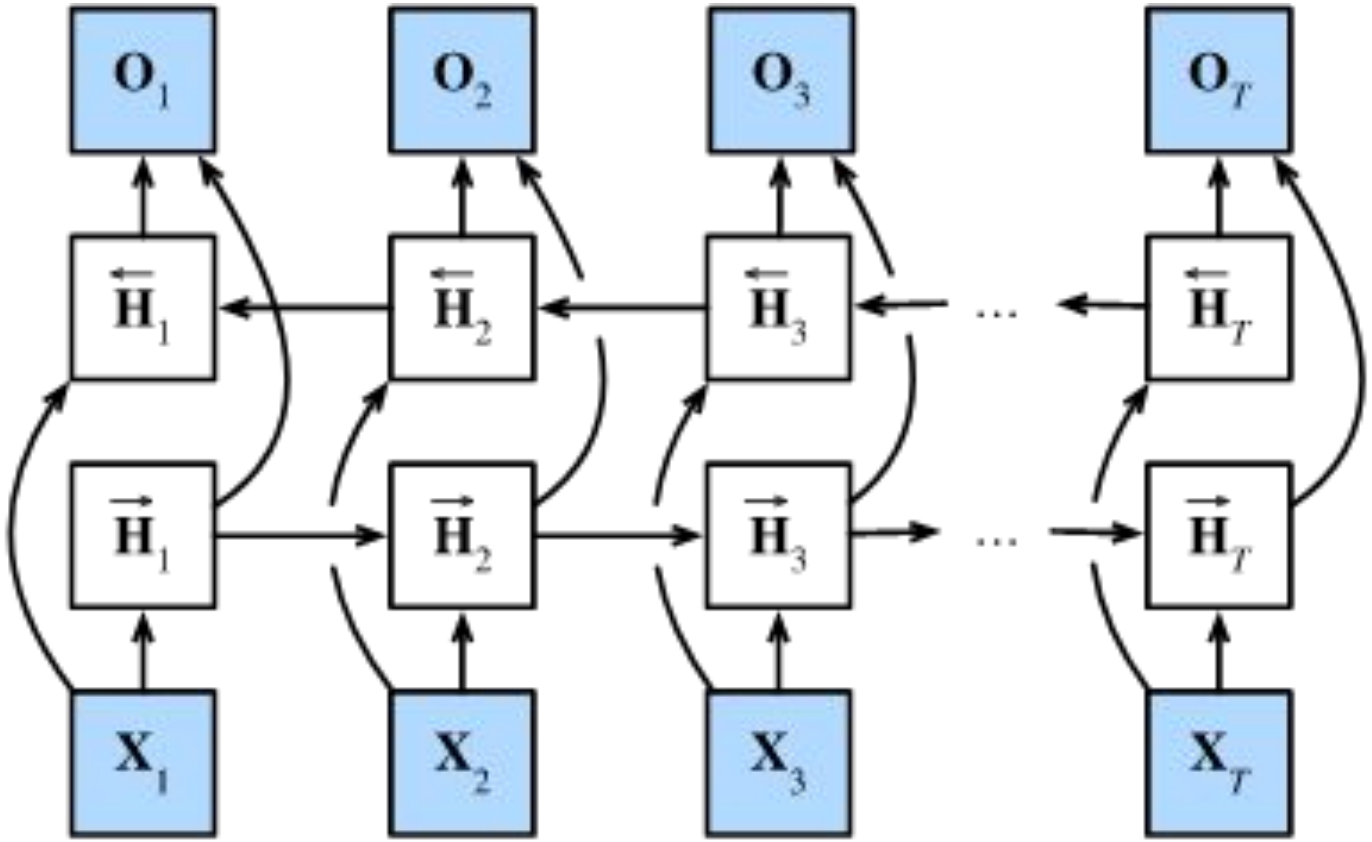
Recap

Recap of Various Model Architecture



Recap

Bi-directional RNNs



Token Classification

Text Classification is also known as sequence classification.

- In text classification, classify the entire text into categories
- extract “prototype” representation from entire token representation.

Token Classification

Token Classification is also known as sequence tagging.

- In token classification, classify each token of the text.
- Why?
- How?

Token Classification

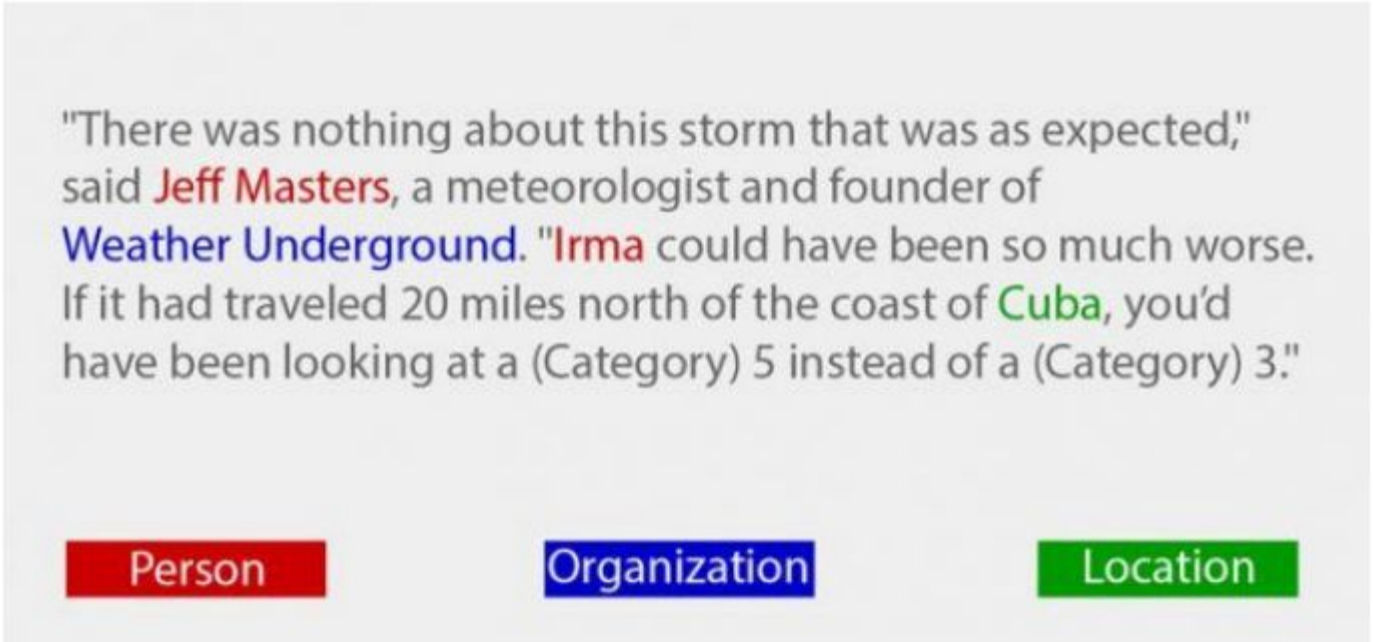
Token classification vs Text Classification

Token Classification

The_DT first_JJ time_NN he_PRP was_VBD shot_VBN in_IN the_DT hand_NN as_IN he_PRP chased_VBD the_DT robbers_NNS outside_RB ...

first	time	shot	in	hand	as	chased	outside
JJ	NN	NN	IN	NN	IN	JJ	IN
RB	VB	VBD	RB	VB	RB	VBD	JJ
		VBN	RP			VBN	NN
							RB

Named Entity recognition

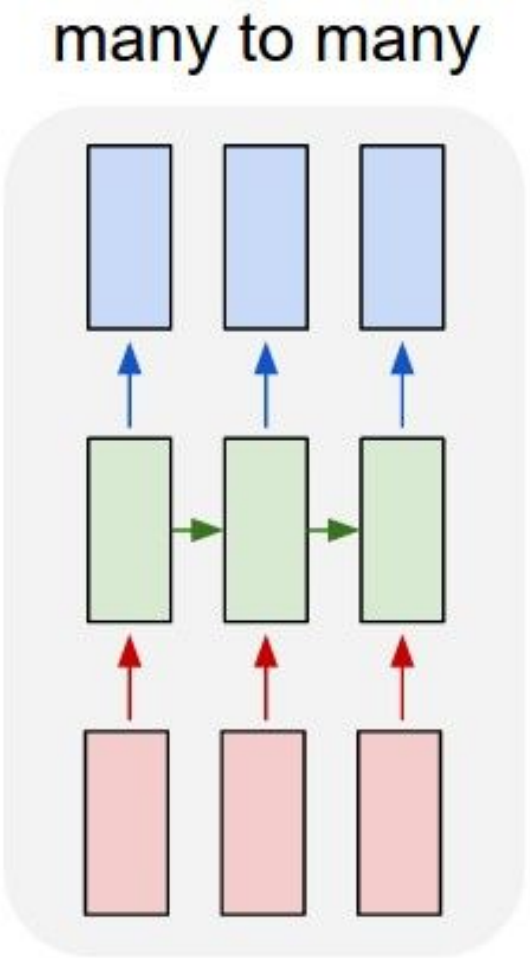


Named Entity Recognition

NER as BIO tagging (Token-level prediction)

B - Beginning / I - Intermediate / O-others

Example : Barack Obama was the president of the United States



Named Entity Recognition

깜짝 Quiz (5 min break)

1. True / False : LSTM's forget and input gates are defined such that their dimension-wise summation is always to 1.0

2. Suppose if we want to train BIO tagging model to classify **n** categories, how many classes are?

3. Suppose the output from the first layer is [4,2,-1]. we apply dropout with probability of 0.2. if the second dimension is dropped, what is the input to the second layer during training time?

Machine Reading Comprehension (MRC)

Question Answering (Extractive) :

- Input is Context and question
- Expected Output is a span in the context
- Classifying start, end and others

Hypothesis :

Ground truth answer always in the paragraph

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

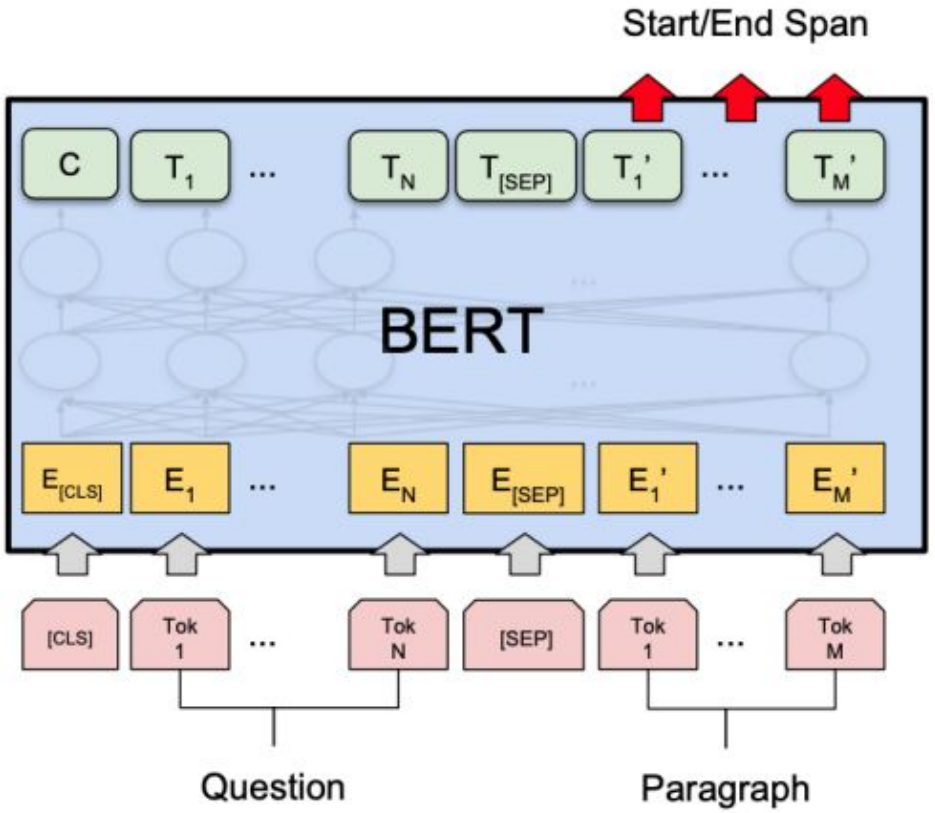
What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

QA

Question Answering (Extractive) :



$$\mathcal{L} = -\log p_{\text{start}}(s^*) - \log p_{\text{end}}(e^*)$$

$$p_{\text{start}}(i) = \text{softmax}_i(\mathbf{w}_{\text{start}}^\intercal \mathbf{H})$$

$$p_{\text{end}}(i) = \text{softmax}_i(\mathbf{w}_{\text{end}}^\intercal \mathbf{H})$$

QA : Long Term dependency in QA

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

- A model needs to be sufficiently aware of distant tokens

- When dealing with long text and paragraphs, LSTM is not good enough.

QA : Long Term dependency in QA

To enhance MRC model …

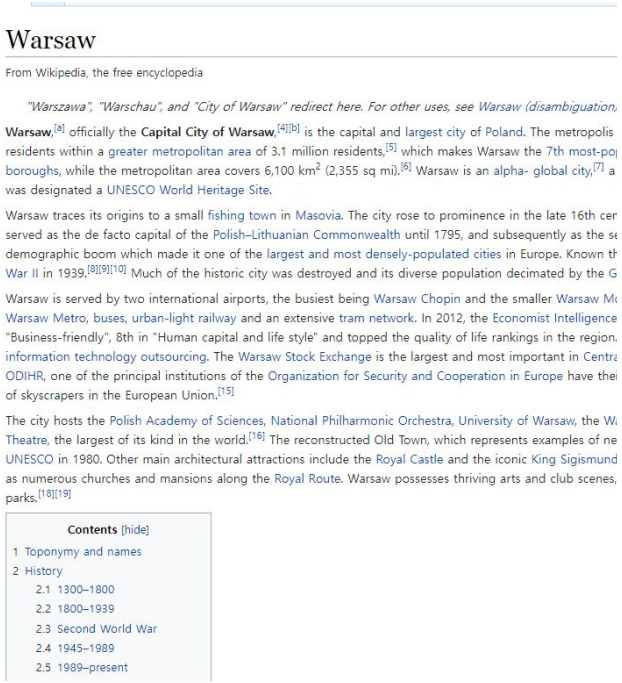
- **Attention Mechanism..?**
- Also motivates the development of Transformer, which becom general-purpose architecture for modernn AI.
-

Retrieval

Open Domain QA / Entity Retrieval

Question : How many of warsaw's inhabitants spoke polish in 1933?

Retriever



Warsaw

From Wikipedia, the free encyclopedia

"Warszawa", "Warschau", and "City of Warsaw" redirect here. For other uses, see Warsaw (disambiguation),

Warsaw,[6] officially the **Capital City of Warsaw**,[4][5] is the capital and largest city of Poland. The metropolis residents within a greater metropolitan area of 3.1 million residents,[5] which makes Warsaw the 7th most-po boroughs, while the metropolitan area covers 6,100 km² (2,355 sq mi).[6] Warsaw is an alpha- global city,[7] a was designated a UNESCO World Heritage Site.

Warsaw traces its origins to a small fishing town in Masovia. The city rose to prominence in the late 16th cer served as the de facto capital of the Polish–Lithuanian Commonwealth until 1795, and subsequently as the se demographic boom which made it one of the largest and most densely-populated cities in Europe. Known th War II in 1939.[8][9][10] Much of the historic city was destroyed and its diverse population decimated by the G

Warsaw is served by two international airports, the busiest being Warsaw Chopin and the smaller Warsaw M Warsaw Metro, buses, urban-light railway and an extensive tram network. In 2012, the Economist Intelligence "Business-friendly", 8th in "Human capital and life style" and topped the quality of life rankings in the region. information technology outsourcing. The Warsaw Stock Exchange is the largest and most important in Centra ODIHR, one of the principal institutions of the Organization for Security and Cooperation in Europe have thei of skyscrapers in the European Union.[15]

The city hosts the Polish Academy of Sciences, National Philharmonic Orchestra, University of Warsaw, the W Theatre, the largest of its kind in the world.[16] The reconstructed Old Town, which represents examples of ne UNESCO in 1980. Other main architectural attractions include the Royal Castle and the iconic King Sigismund as numerous churches and mansions along the Royal Route. Warsaw possesses thriving arts and club scenes, parks.[18][19]

Contents [hide]

1 Toponymy and names
2 History
  2.1 1300–1800
  2.2 1800–1939
  2.3 Second World War
  2.4 1945–1989
  2.5 1989–present

Reader

Retrieval

Retrieval

- Minimize candidates of possible documents from millons of passages.

- Question and Passage similarity measure

Retrieval

Question and Passage similarity measure (sparse)

- Bag Of Words (BOW)
- TF-IDF

Retrieval

Sparse Retriever

- Bag Of Words (BOW)

- TF-IDF
  - Vanilla BoW is limited.
  - Different wrods should have different weights; we don’t care common words as “is” “a”..

Retrieval

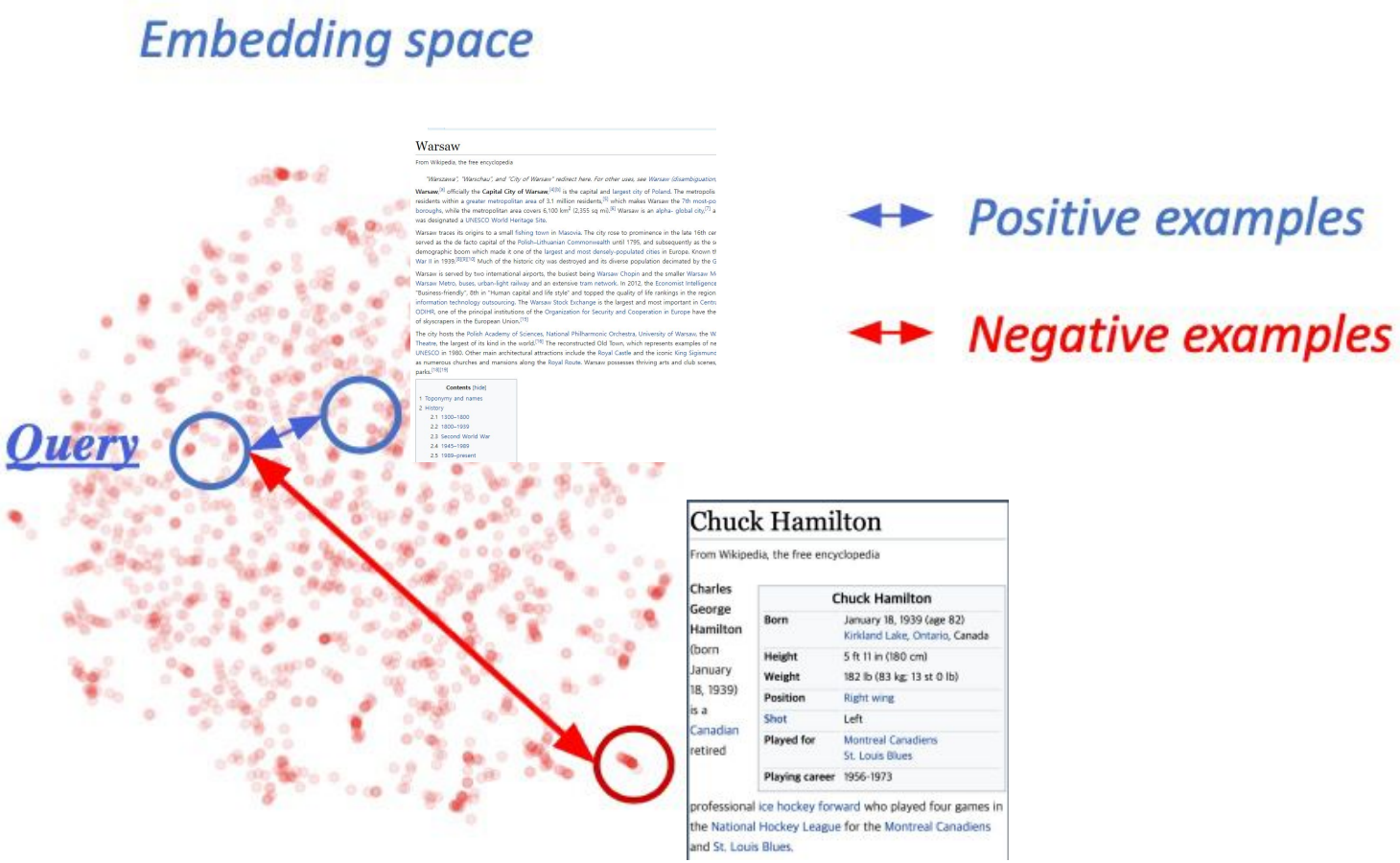
Dense Retriever

Query →

Passage (Document) →

Retrieval

Dense Retriever



Text Generation : Summarization

References:

- CS224n(http://web.stanford.edu/class/cs224n/slides/cs224n-2021-lecture11-qa-v2.pdf)
- https://d2l.ai/chapter_natural-language-processing-applications/finetuning-bert.html
-