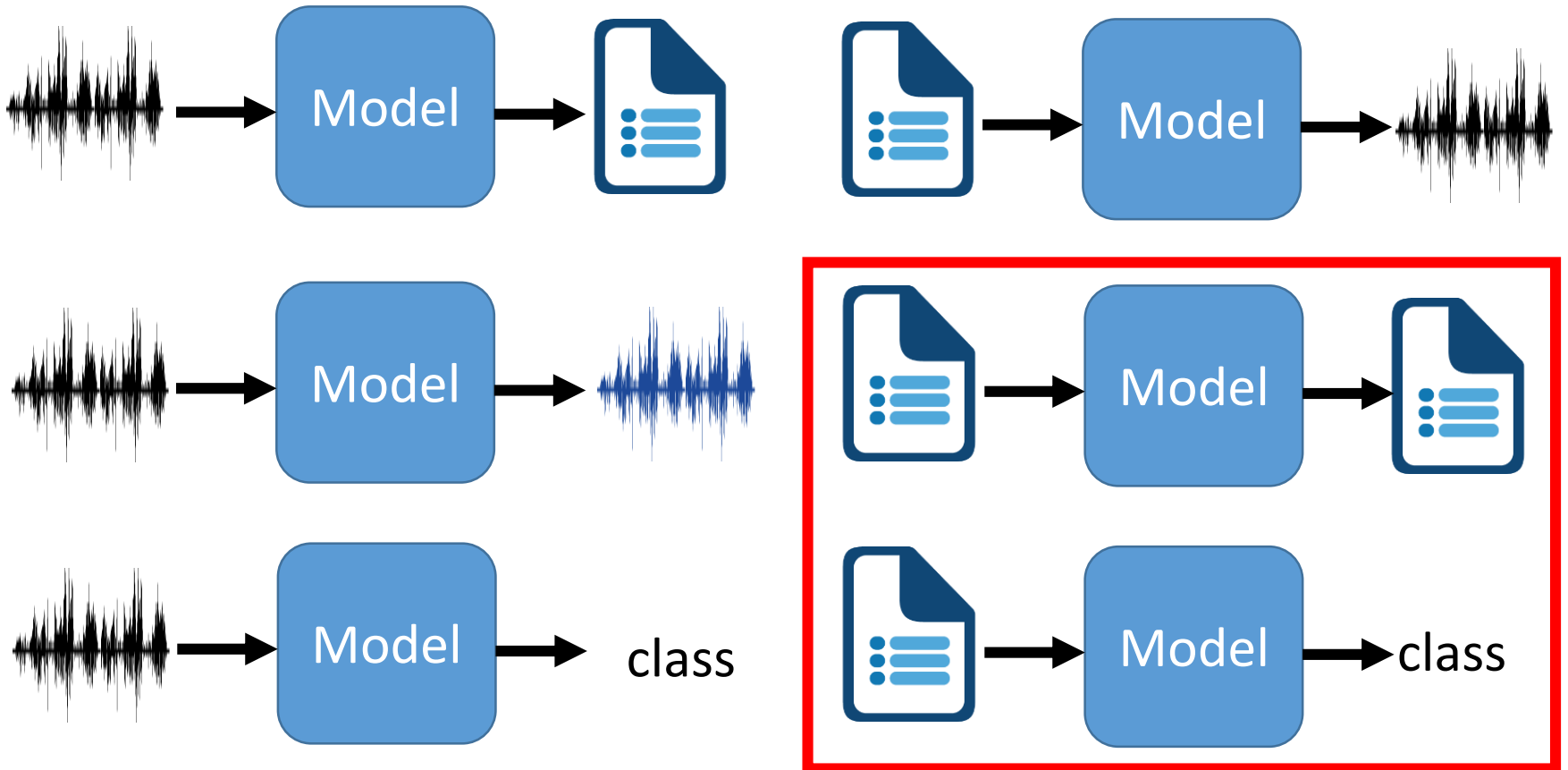


# NLP Tasks

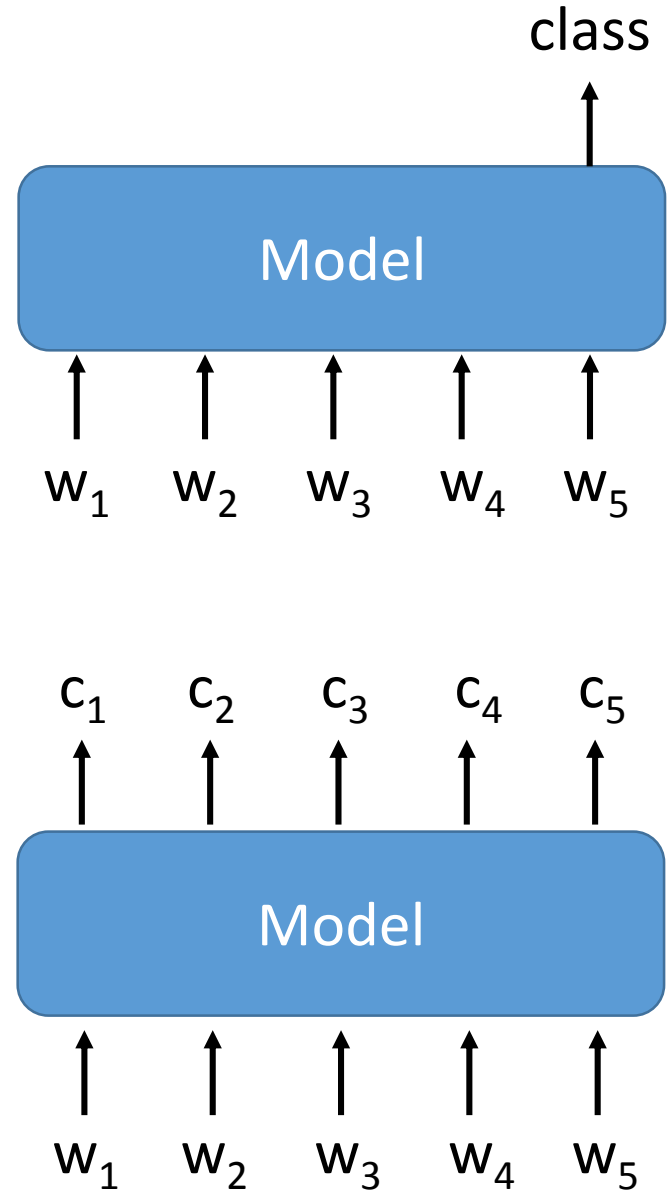
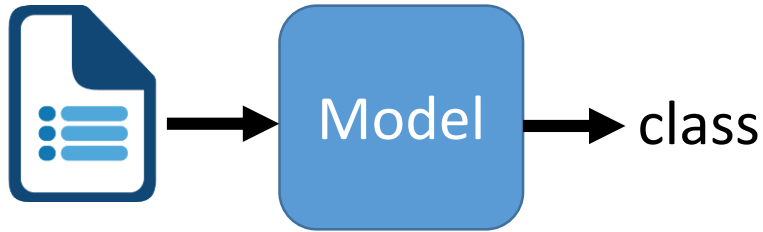
Hung-yi Lee

# One slide for this course

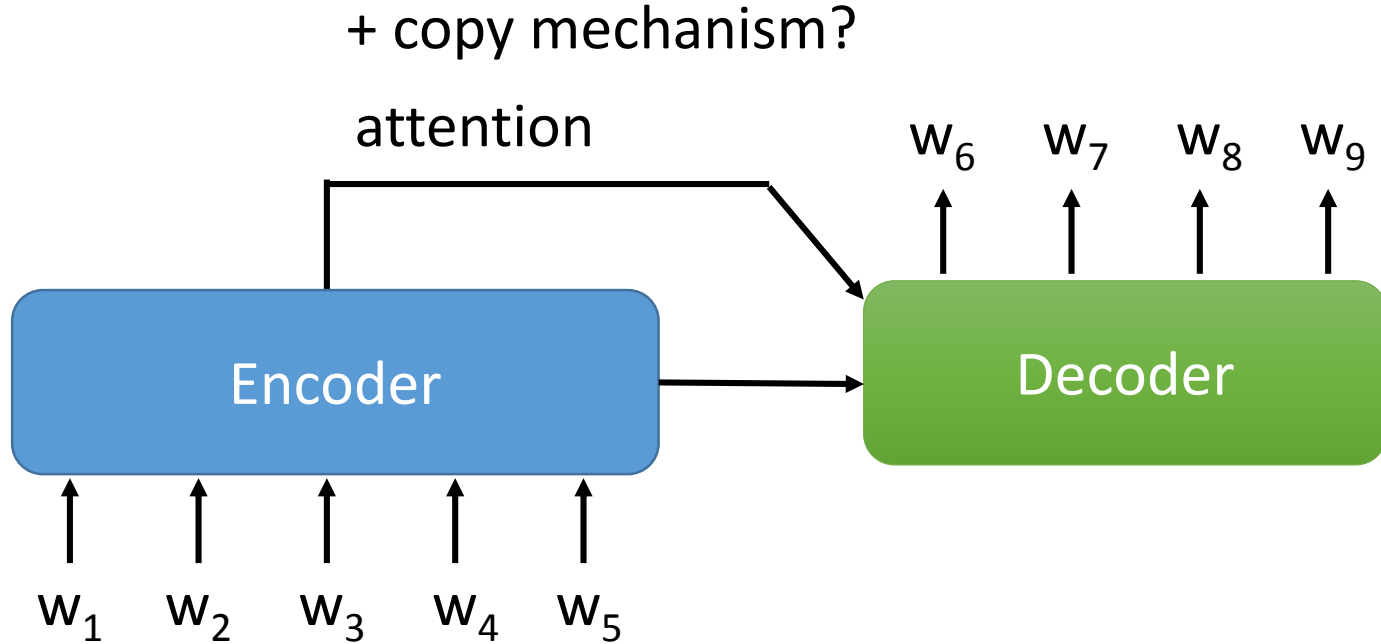
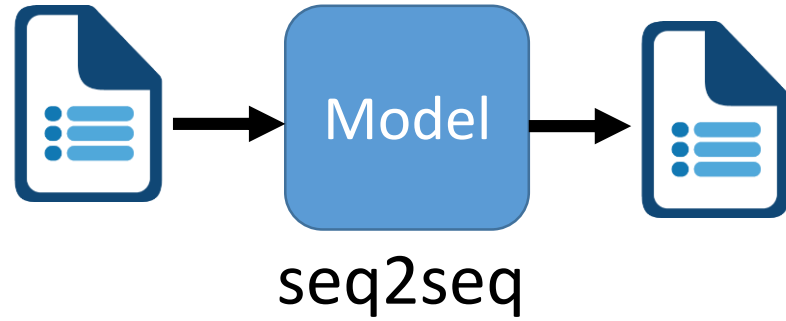


Usually people call them “NLP” tasks.

# Category

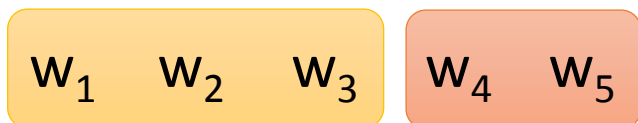


# Category

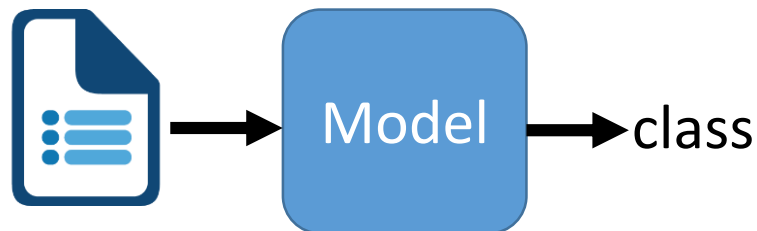
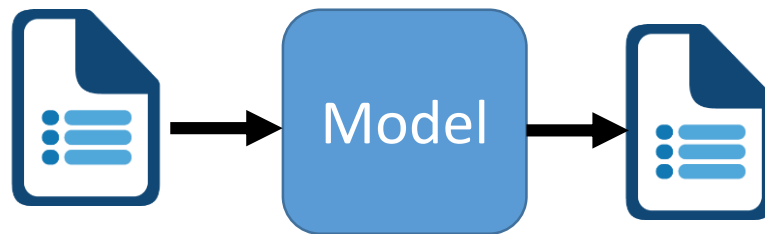
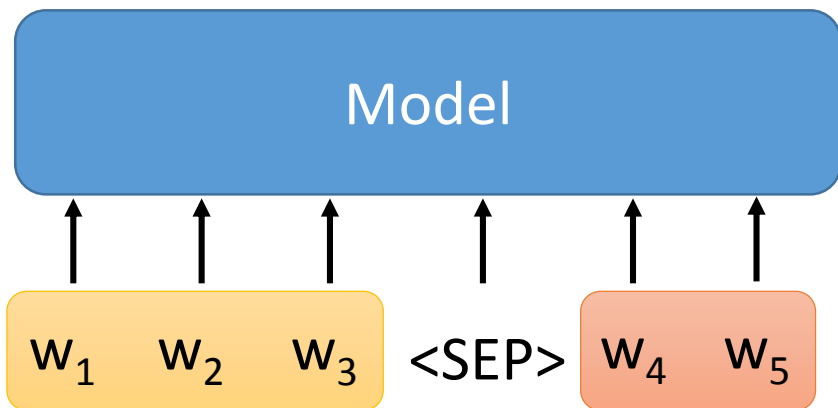


# Category

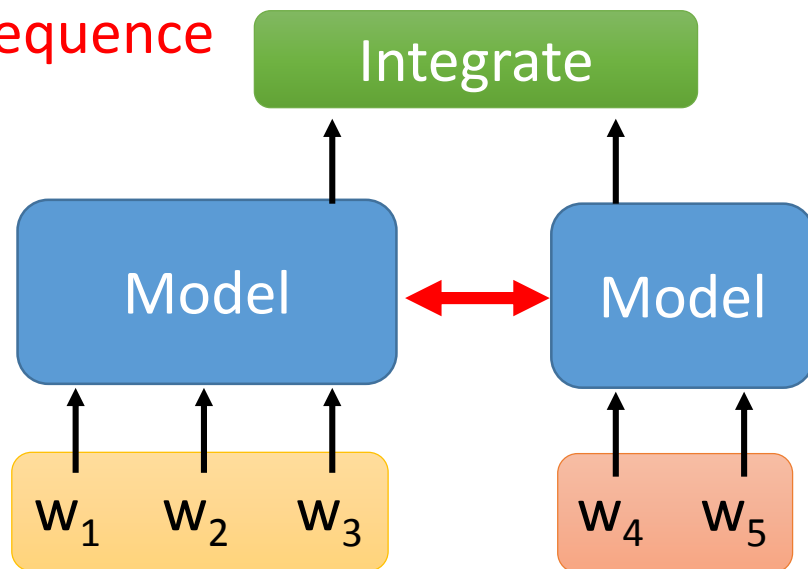
What happens if there are more than two input sequences?



Simply concatenate ...



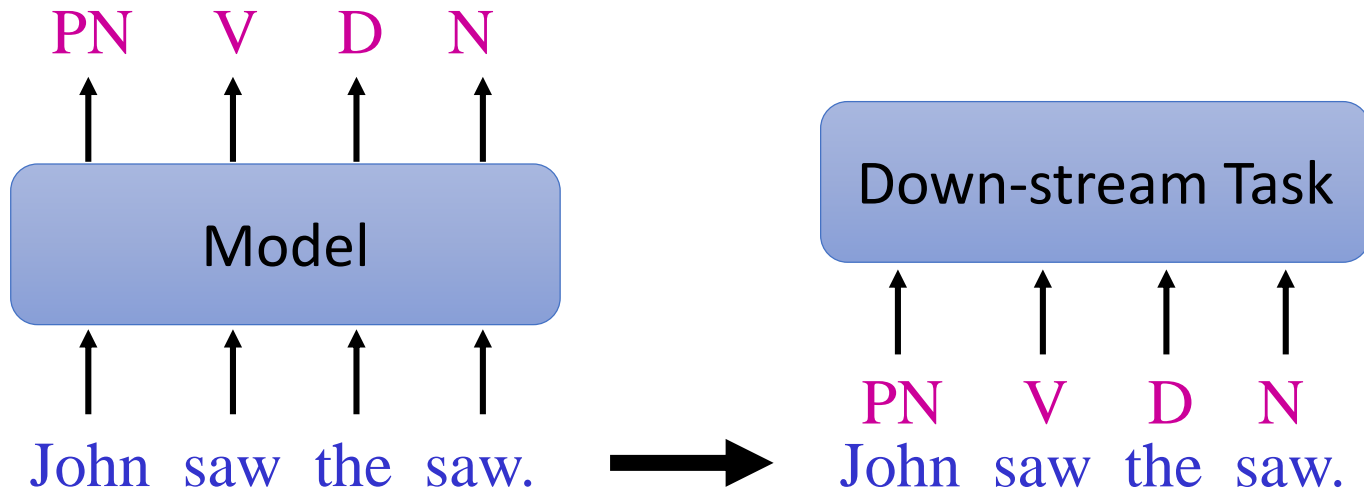
Attention between  
sequence



|                      | <b>One Sequence</b>   | <b>Multiple Sequences</b>                       |
|----------------------|---|---|
| One Class            | Sentiment Classification<br>Stance Detection<br>Veracity Prediction<br>Intent Classification<br>Dialogue Policy | NLI<br>Search Engine<br>Relation Extraction     |
| Class for each Token | POS tagging<br>Word segmentation<br>Extraction Summarization<br>Slotting Filling<br>NER                         |   |
| Copy from Input      |   | Extractive QA                                   |
| General Sequence     | Abstractive Summarization<br>Translation<br>Grammar Correction<br>NLG   | General QA<br>Task Oriented Dialogue<br>Chatbot |
| Other?               | Parsing, Coreference Resolution   |   |

# Part-of-Speech (POS) Tagging

- Annotate each word in a sentence with a part-of-speech (e.g. Verb, Adjective, Noun)



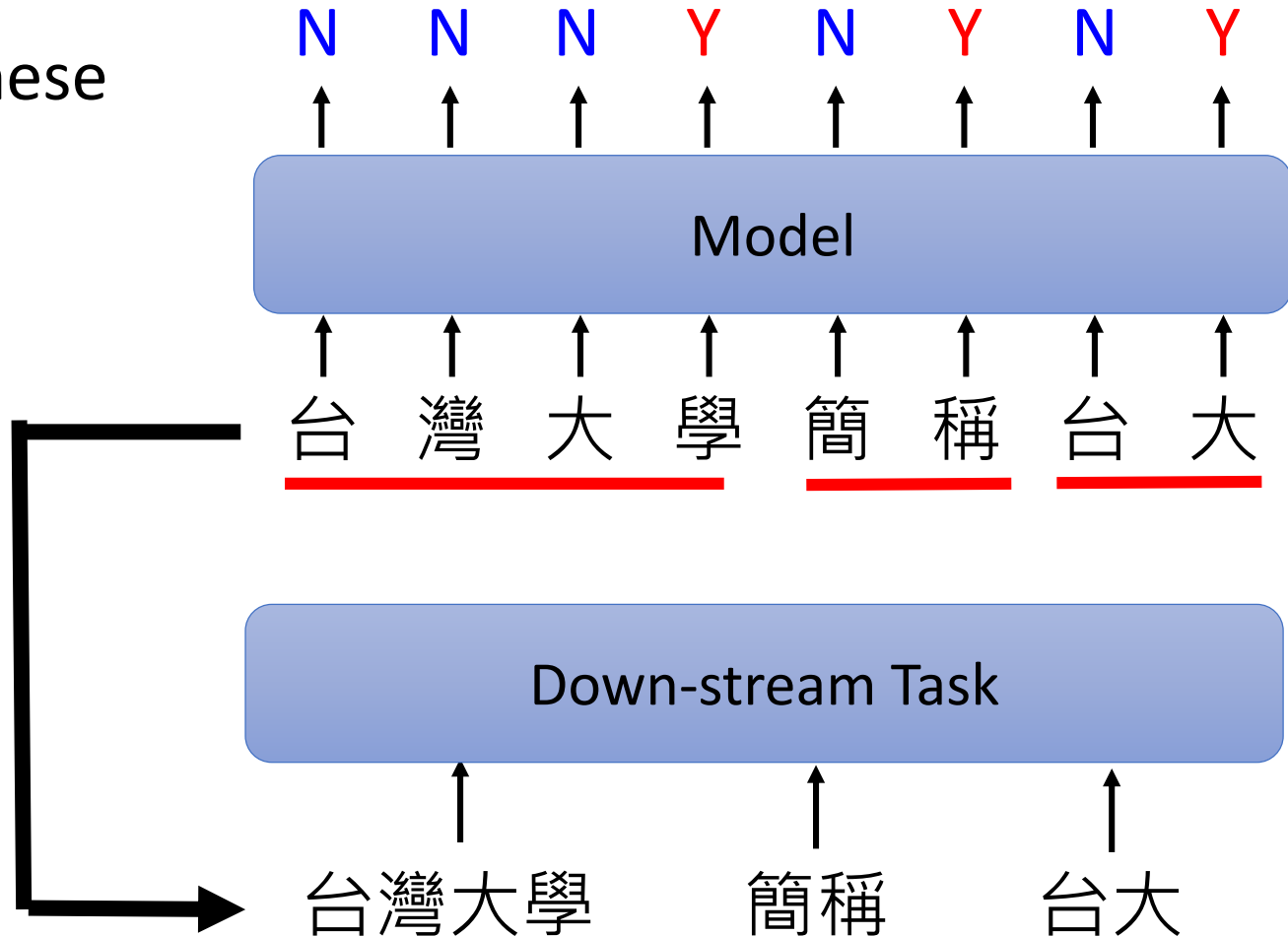
Input: sequence

Output: class for each token

# Word Segmentation

Input: sequence  
Output: class for each token

- for Chinese





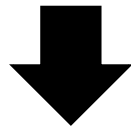
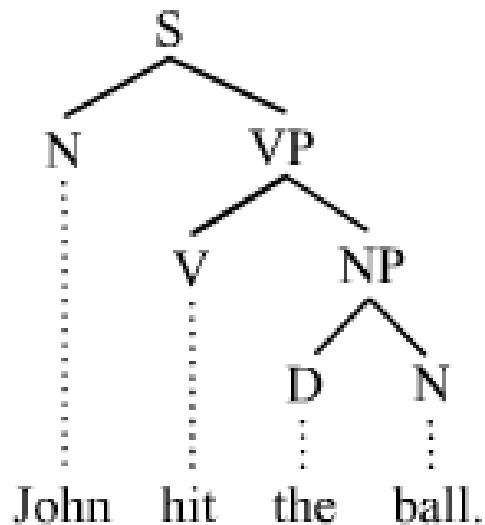
『楊公再興之神』

(出自《笑傲江湖》)

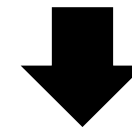
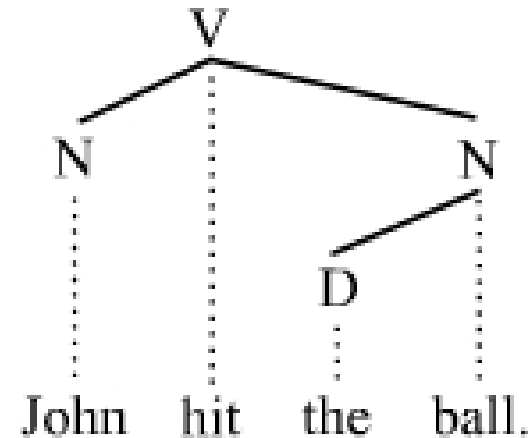
# Parsing

Outlier?

## Constituency Parsing



## Dependency Parsing



The results of parsing can be used in the downstream tasks.

# Coreference Resolution (指代消解)

Outlier?

Paul Allen was born on January 21, 1953. He attended Lakeside School, where he befriended Bill Gates. Paul and Bill used a teletype terminal at their high school, Lakeside, to develop their programming skills on several time-sharing computer systems.

Source of example:

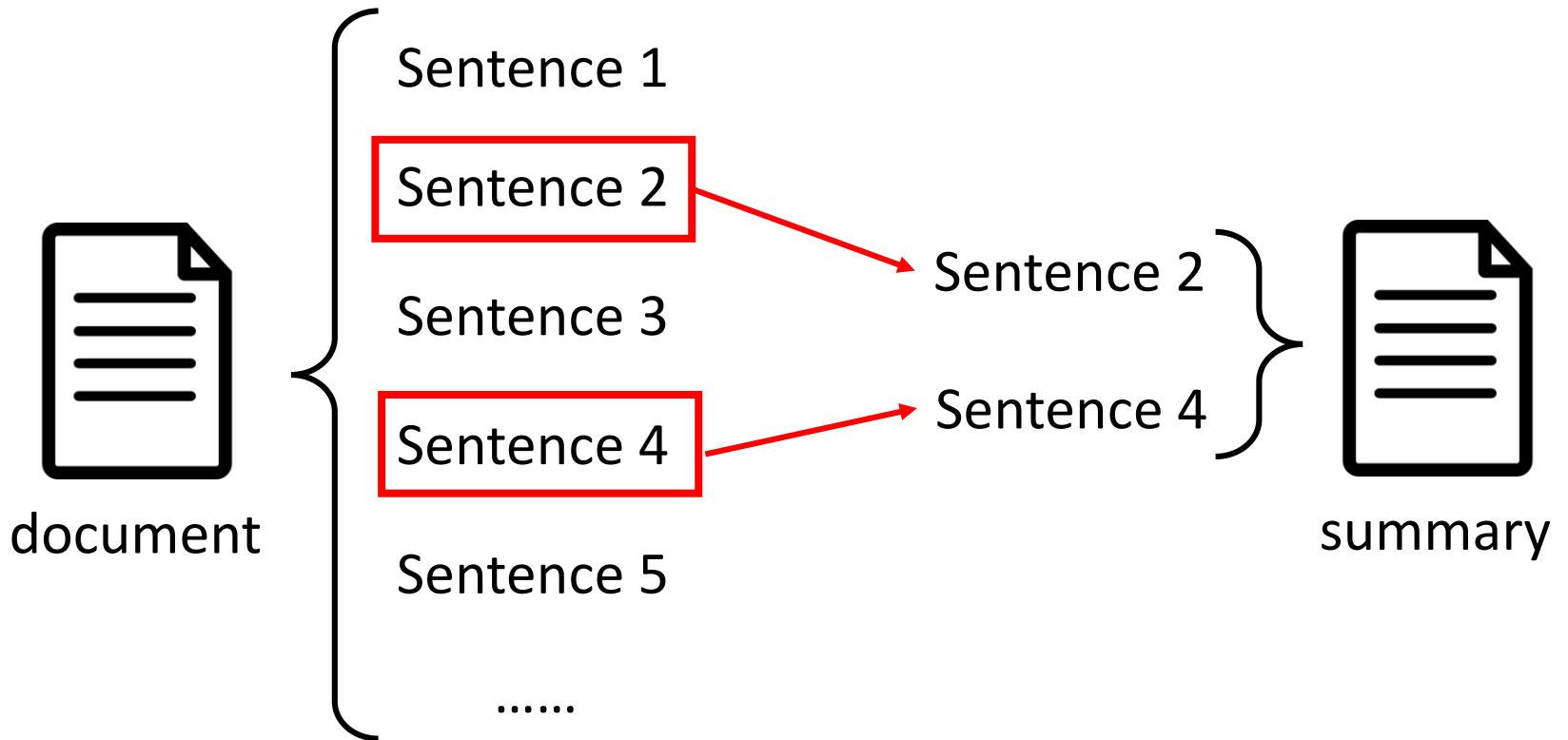
<https://demo.allennlp.org/coreference-resolution/>

# Summarization

Input: sequence  
Output: class for each token

- **Extractive** summarization

(Here a token is a sentence)

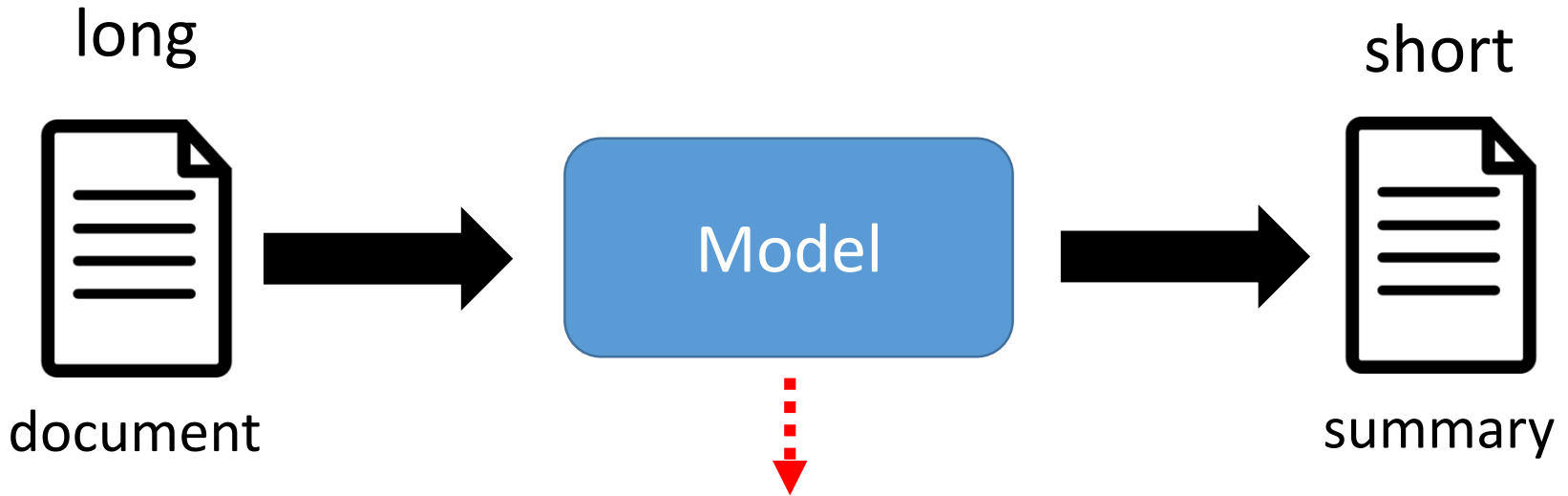


# Summarization

- **Abstractive** summarization

Input: sequence  
Output: sequence

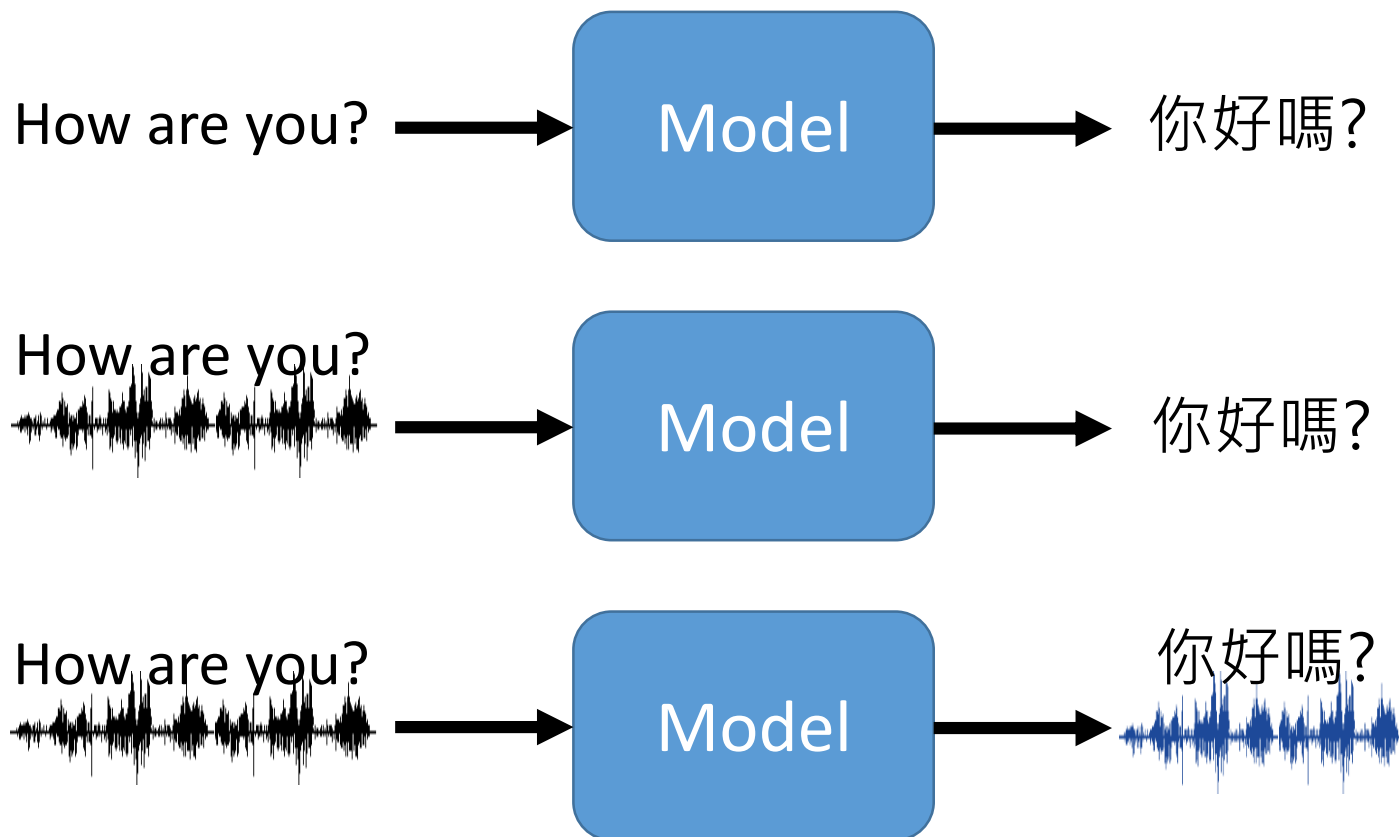
(copy is encouraged)



Pointer network:  
encouraging direct copy from input

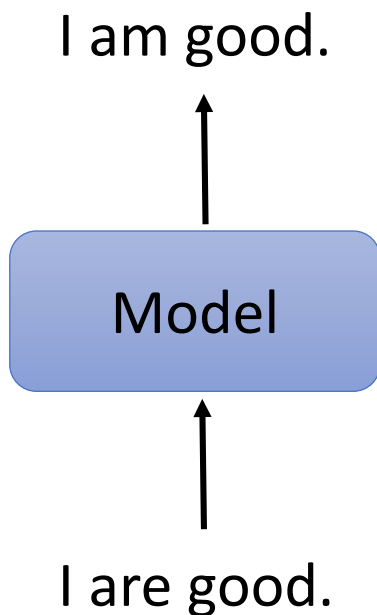
# Machine Translation

Input: sequence  
Output: sequence



Unsupervised machine translation is a critical research direction.

# Grammar Error Correction



Input: sequence  
Output: sequence

(copy is encouraged)

| Example 1 |  |
|-----------|--|
| x         | [ Bolt can have run race ]   |
| y         | [ Bolt could have run the race ]   |
| diff      | (C, []) (C,Bolt) (D,can) (I,can,could) (C,have)<br>(C,run) (I,run,the) (C,race) (C,) |
| e         | C C R(could) C A(the) C C<br>↑ ↑ ↑ ↑ ↑ ↑ ↑<br>[ Bolt can have run race ]             |
| Example 2 |  |
| x         | [ He still won race ! ]  |
| y         | [ However , he still won ! ]   |
| diff      | (C, []) (I, [,However, ) (D,He) (I,He,he) (C,still)<br>(C,won) (D,-race) (C,!) (C,)  |
| e         | A(However, ) T_case C C D C C<br>↑ ↑ ↑ ↑ ↑ ↑ ↑<br>[ He still won race ! ]            |

Source of image:

<https://www.aclweb.org/anthology/D19-1435.pdf>

# Sentiment Classification

Input: sequence  
Output: class

柯南劇場版《紺青之拳》還蠻有趣的



柯南劇場版《紺青之拳》槽點很多



柯南劇場版《紺青之拳》雖然槽點很多，但還蠻有趣的



柯南劇場版《紺青之拳》雖然還蠻有趣的，但槽點很多





# Stance Detection

Input: two sequences  
Output: a class

Source: 李宏毅是個型男 (post on twitter or FB)

Reply: 他只是個死臭酸宅



**DENIED**

Many systems use the Support, Denying, Querying, and Commenting (SDQC) labels for classifying replies.

Used in Veracity Prediction

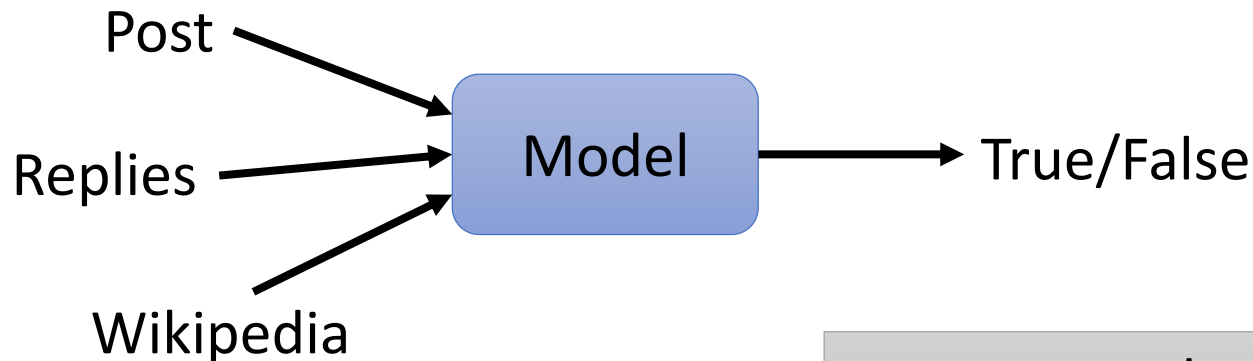
# Veracity Prediction

## Veracity prediction examples:

**u1:** Hostage-taker in supermarket siege killed, reports say. #ParisAttacks -LINK- [true]

**u1:** OMG. #Prince rumoured to be performing in Toronto today. Exciting! [false]

Source of image: <https://www.aclweb.org/anthology/S17-2006.pdf>



Input: several sequences  
Output: class

# Natural Language Inference (NLI)

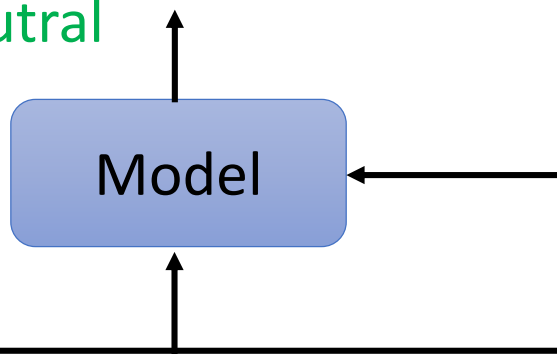
Input: two sequences

Output: a class

contradiction

entailment

neutral



premise: A person on a horse jumps over a broken down airplane

hypothesis: A person is at a diner.

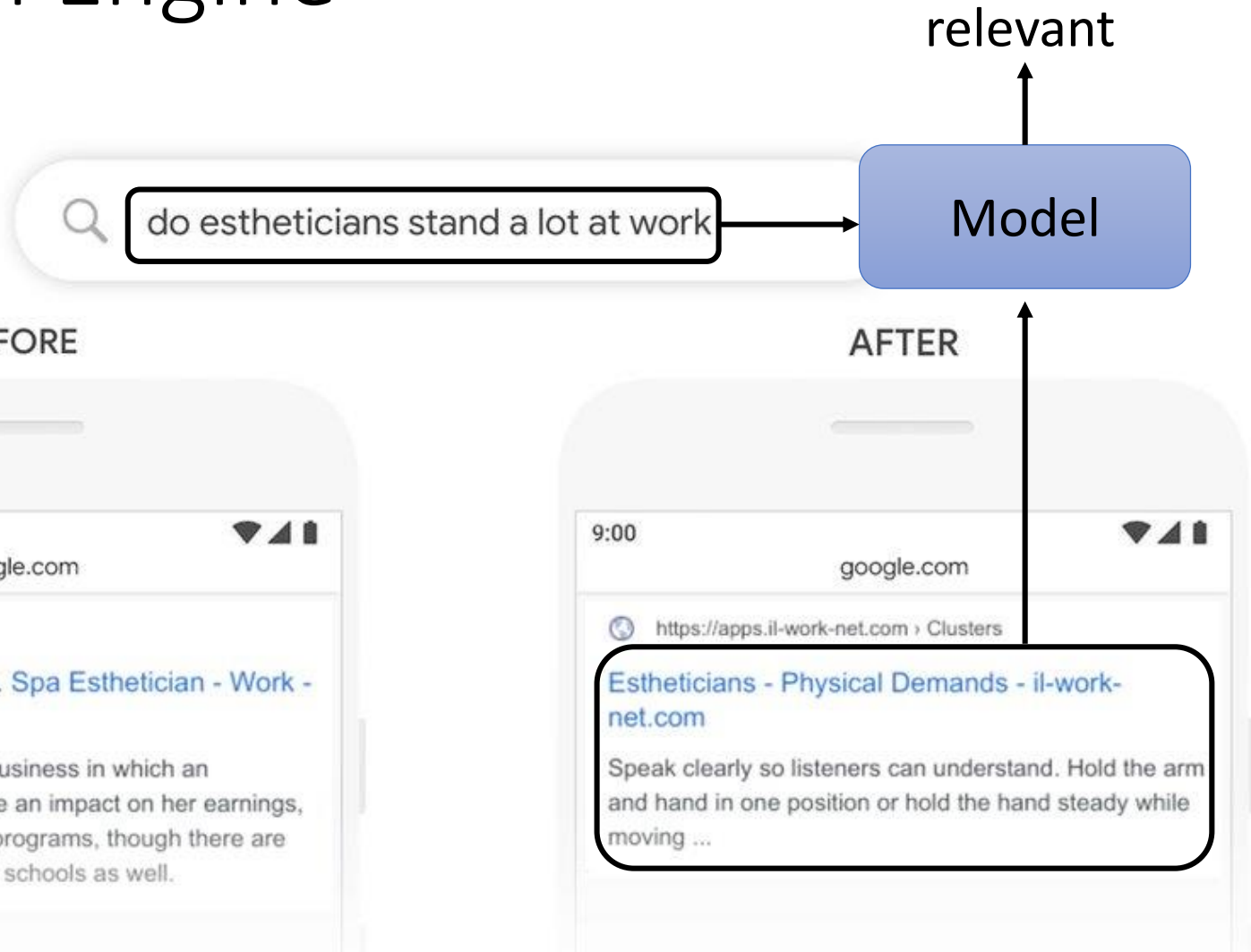
contradiction

hypothesis: A person is outdoors, on a horse. entailment

hypothesis: person is training his horse for a competition. neutral

# Search Engine

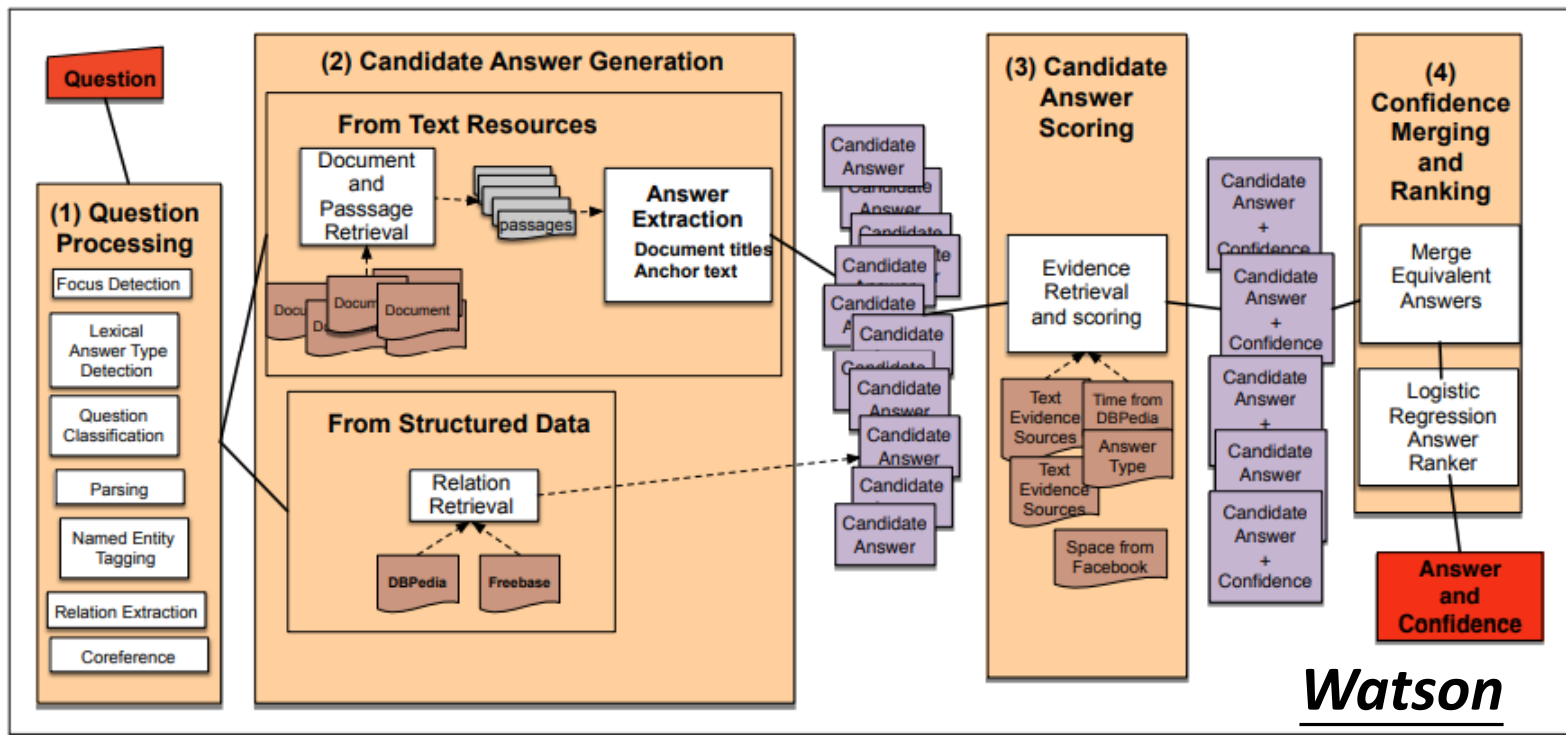
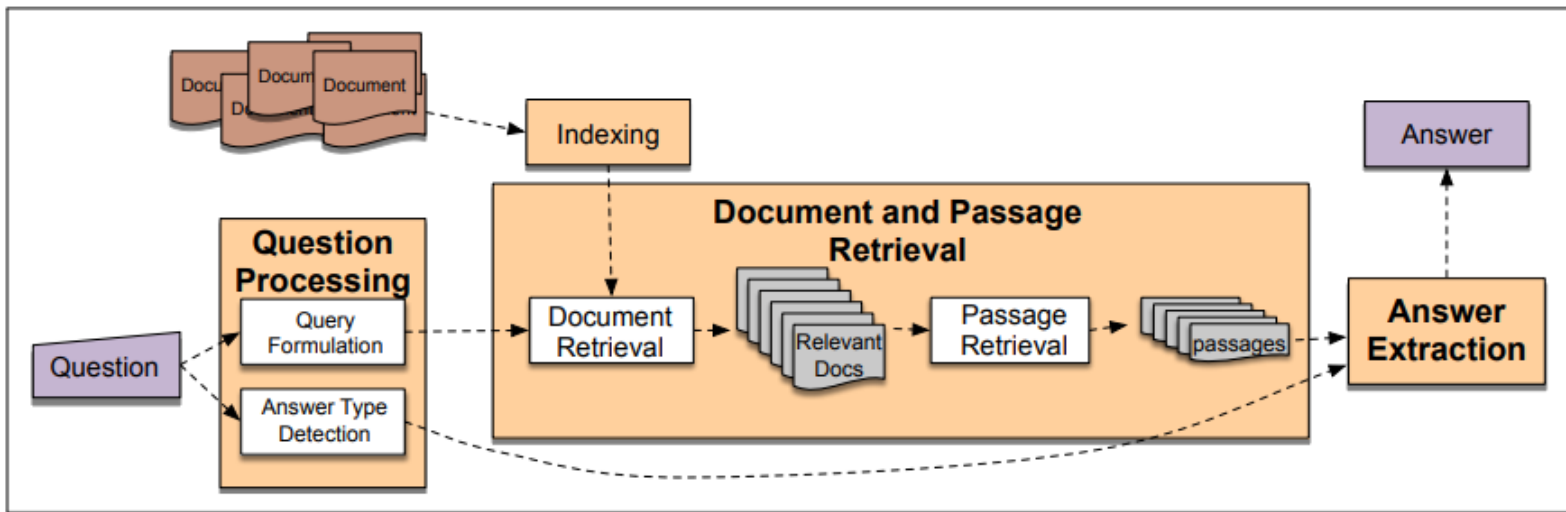
Input: two sequences  
Output: a class



Source of image: <https://www.blog.google/products/search/search-language-understanding-bert/>

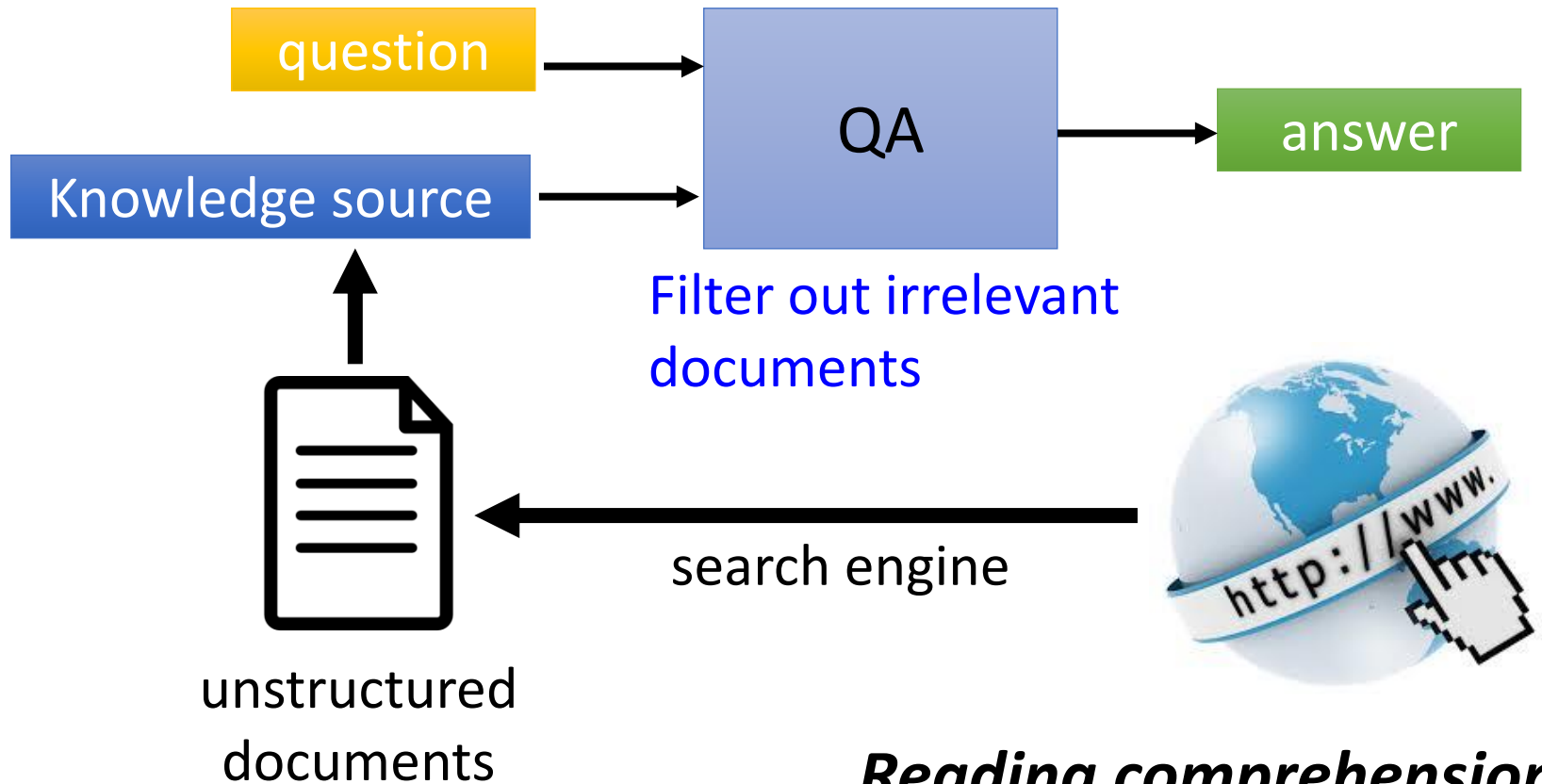
# Question Answering (QA)





# Question Answering

Input: several sequences  
Output: sequence

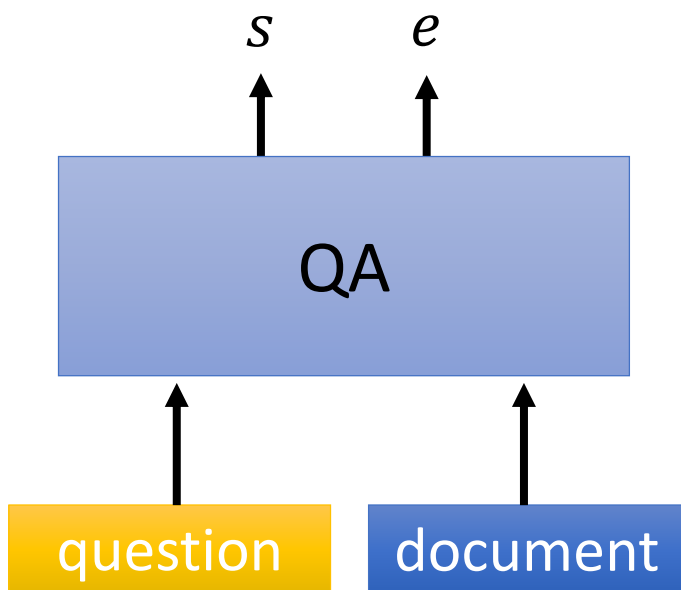


# Question Answering

Input: several sequences  
Output: sequence

(copy from input)

- Extractive QA: Answer in the document



In meteorology, precipitation is any product of the condensation of **17** spheric 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 **77** at **79** cations are called "showers".

What causes precipitation to fall?

**gravity**

$s = 17, e = 17$

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**

$s = 77, e = 79$



# Dialogue

**Chatting**

**Task-oriented**



# Chatting



<https://www.cleverbot.com/>

User: What is your name?

Cleverbot: Weil ich dich nicht kennw.

User: What is your name?

Cleverbot: Li.

User: Are you Li?

Cleverbot: No i'm Mara.

User: Are you Mara?

Cleverbot: No.

User: Who you are?

Cleverbot: I just said your name you don't need to know my name.

User: Do you know my name?

Cleverbot: No that is why I am asking you.

User: Can you tell me your name?

Cleverbot: I already told you.

User: You are Mara?

Cleverbot: I'm sorry.

# Chatting



<https://www.cleverbot.com/>

User: Sorry for what?

Cleverbot: For not remembering your name.

User: Do you remember your name?

Cleverbot: Yes I do...

User: What is your name?

Cleverbot: Once again, I can't tell you.

User: You have told me that you are Mara.

Cleverbot: I am a boy.

User: Are you a boy?

Cleverbot: Yes.

User: Are you a girl?

Cleverbot: No, are you?

User: No. How old are you?

Cleverbot: 2000 years old.

User: You are so old!

Cleverbot: No i'm quite young.

User: How old are you?

Cleverbot: Why do you keep asking me that?

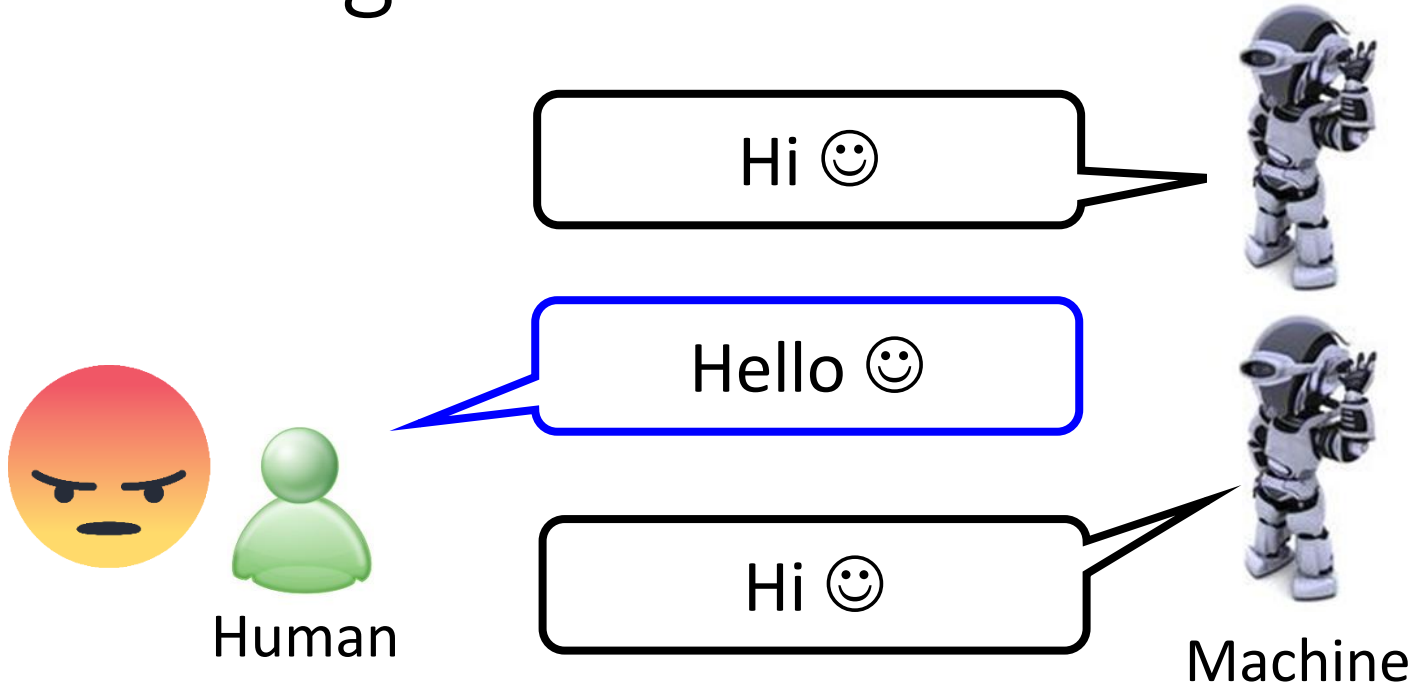
User: Because I want to know your age.

Cleverbot: What if I do not want to tell you.

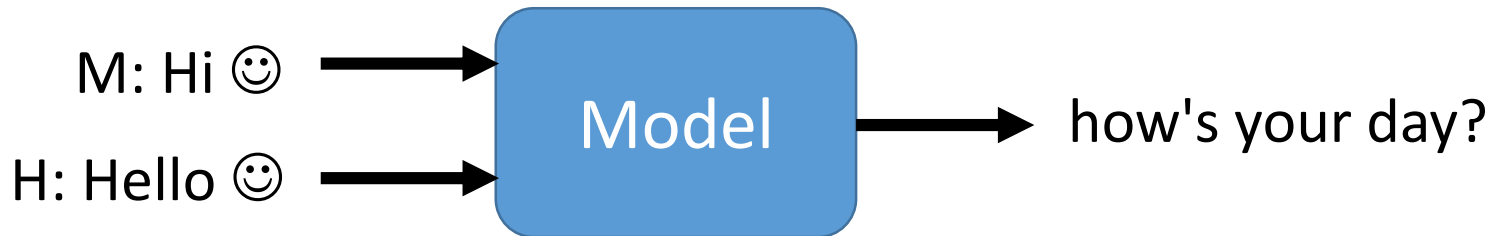
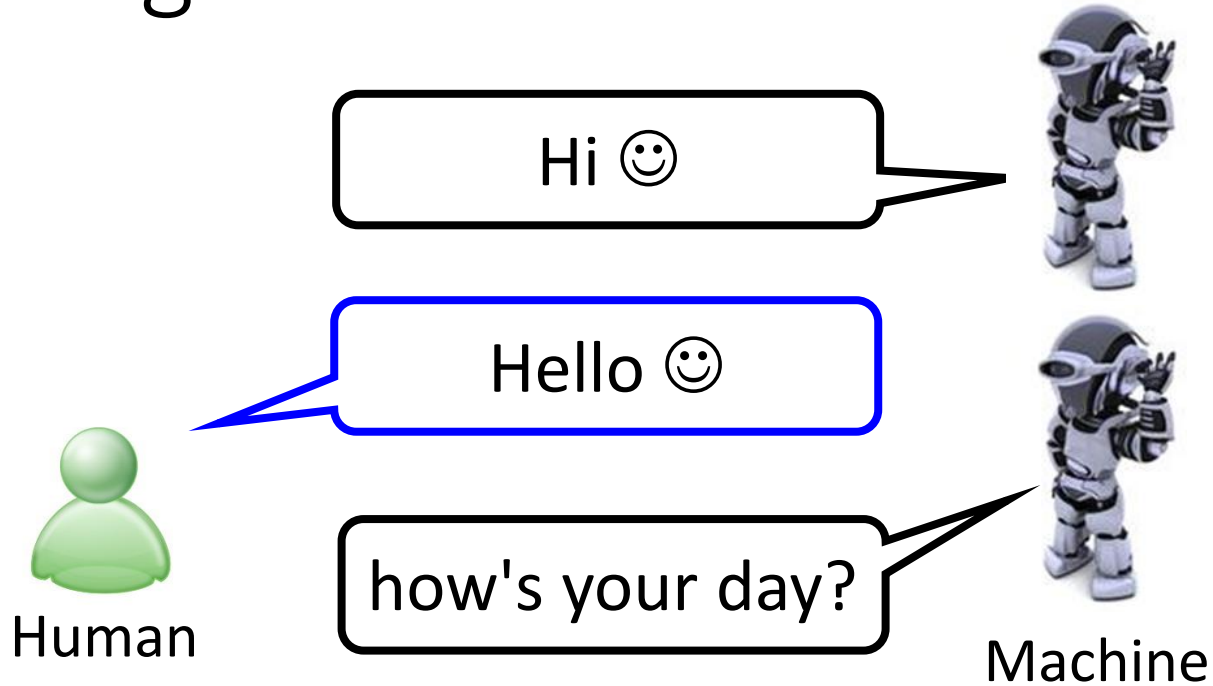
User: Then don't tell me your age.

Cleverbot: My name is Blue Monday.

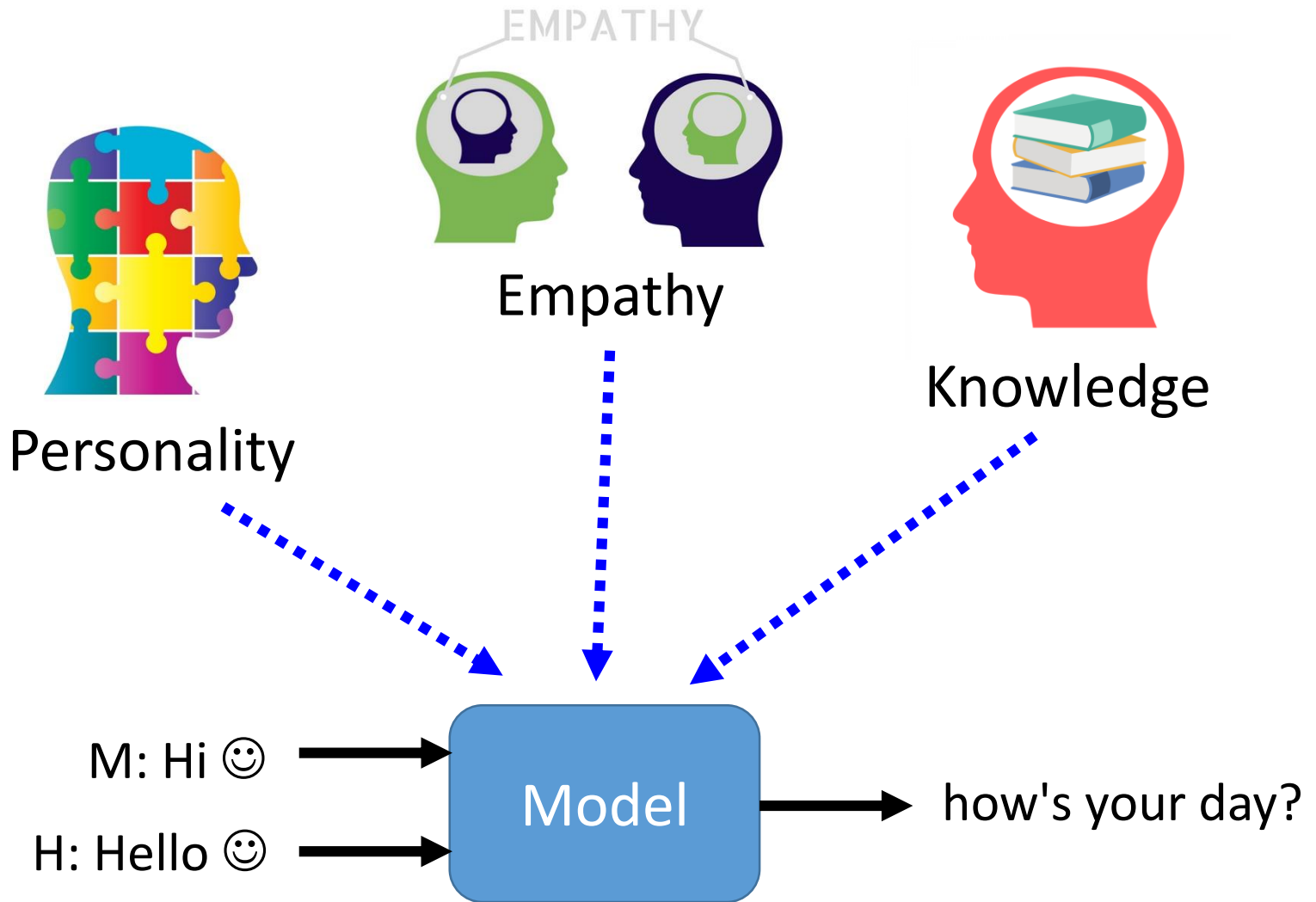
# Chatting



# Chatting



# Chatting



# Task-oriented

|    |       |  |
|----|-------|--|
| 1  | Agent | 您好，這裡是夏季商務旅館，請問有什麼能為您服務的嗎？               |
| 2  | User  | 我要訂房                                     |
| 3  | Agent | 沒問題，請問怎麼稱呼您？                             |
| 4  | User  | 我姓林                                      |
| 5  | Agent | 林小姐您好，請問您預計何時入住？                         |
| 6  | User  | 9月9日晚上入住                                 |
| 7  | Agent | 請問要住幾個晚上呢？                               |
| 8  | User  | 2晚                                       |
| 9  | Agent | 請問訂房人數總共是幾位呢？                            |
| 10 | User  | 5位                                       |
| 11 | Agent | 請問您要預定的房型？                               |
| 12 | User  | 一間豪華套房、一間豪華客房                            |
| 13 | Agent | 好的，您入住的日期是9月9日星期一，退房時間為9月11日星期三。請問確認訂房嗎？ |
| 14 | User  | ok                                       |

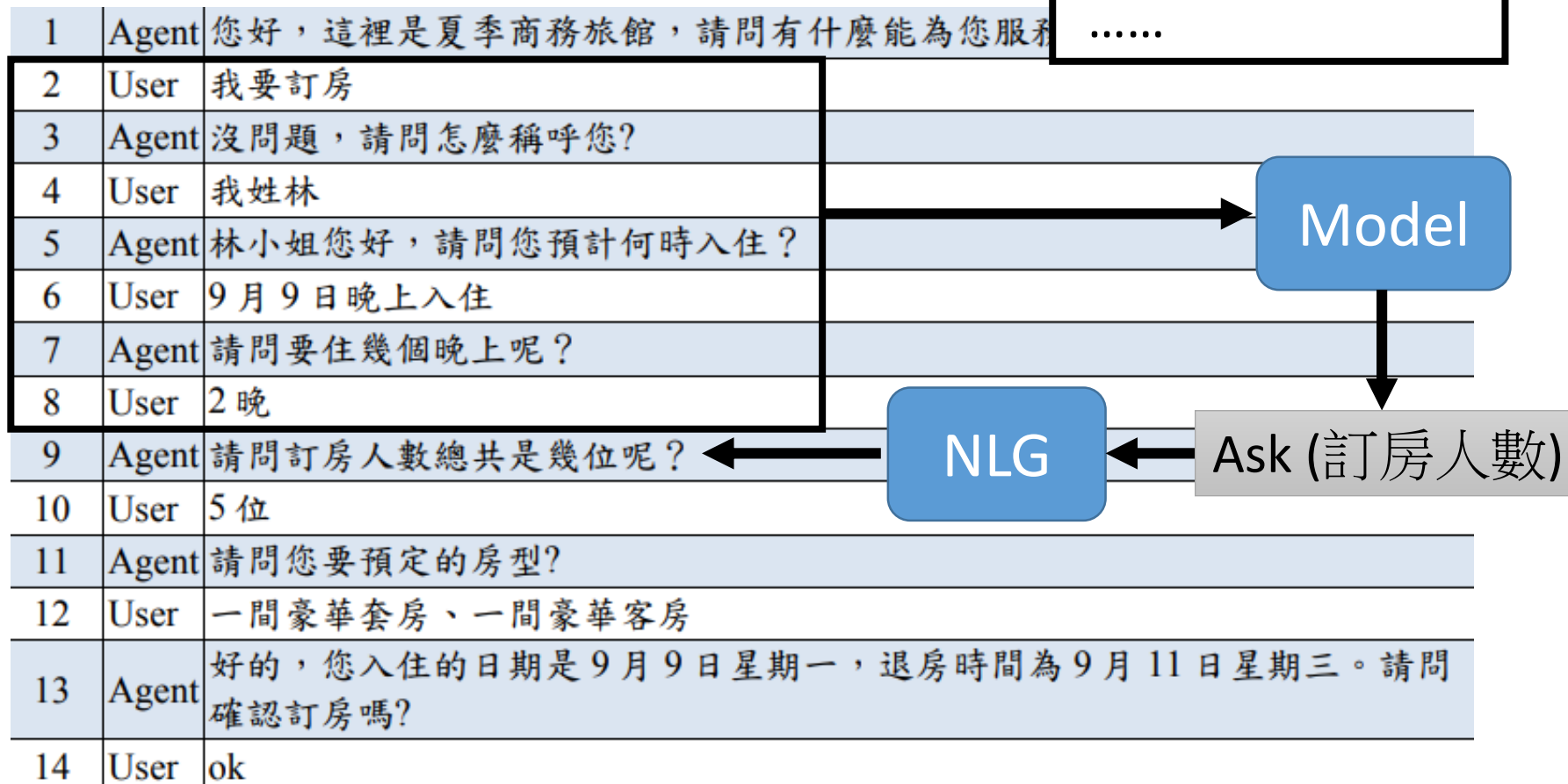
The diagram shows a blue rounded rectangle labeled 'Model' positioned to the right of the dialogue table. Three black arrows originate from the 'Model' box: one points to row 5, one points to row 8, and one points to row 9. Additionally, a black arrow points from row 8 to row 9, indicating a flow in the dialogue.

(source of image: 科技大擂台 2020 簡章範例)

# Natural Language Generation (NLG)

Action:

Greeting  
Ask (入住日)  
Ask (退房日)  
.....

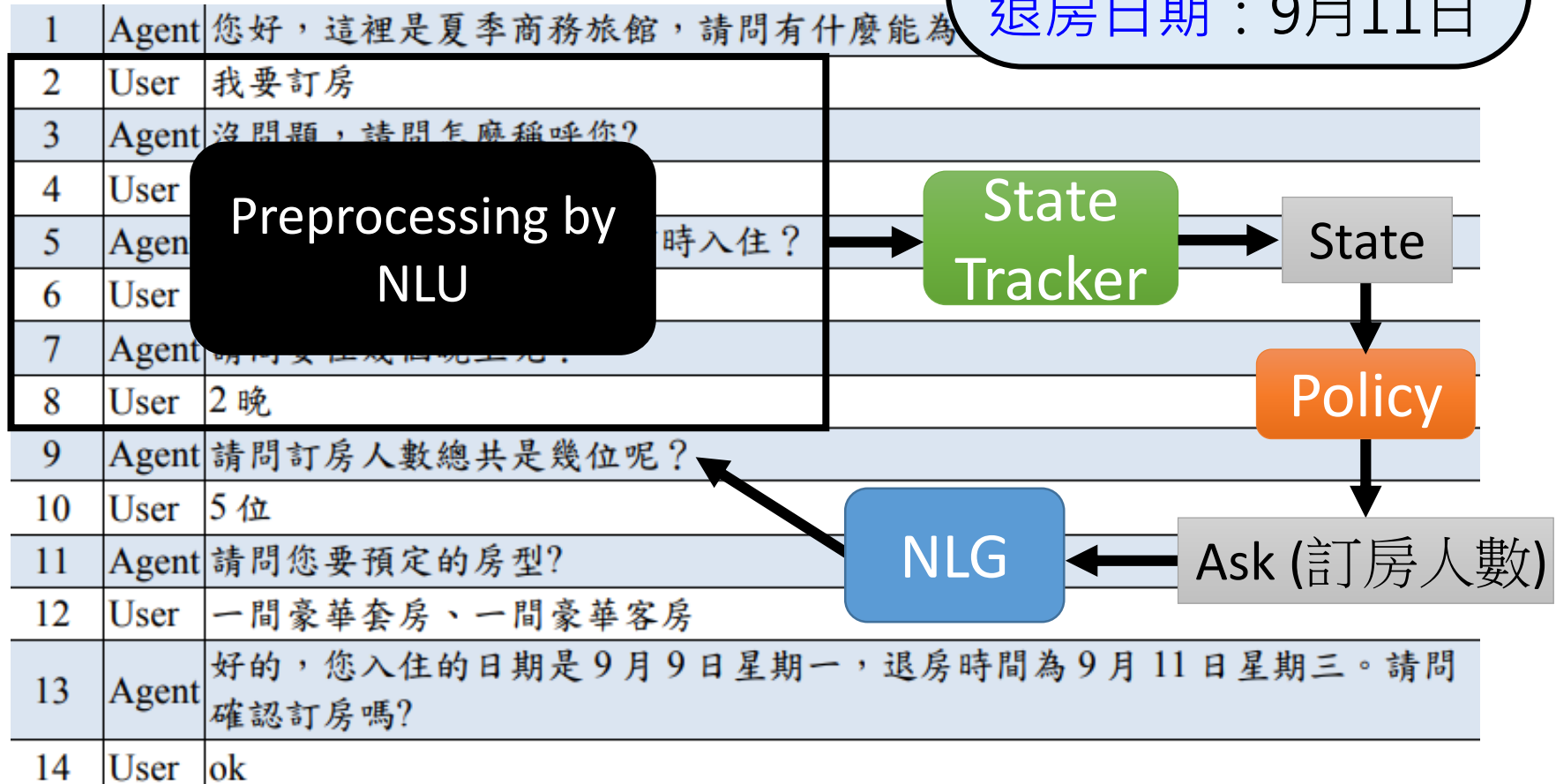




# Policy & State Tracker

**State**: What has happened in this dialogue

連絡人: 林XX  
入住人數: None  
連絡電話: None  
房型: None  
入住日期: 9月9日  
退房日期: 9月11日



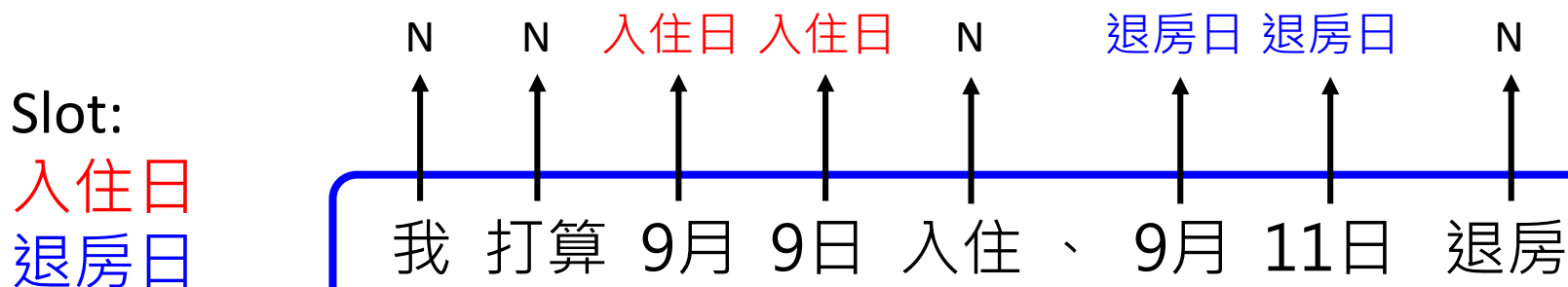
# Natural Language Understanding (NLU)

- Intent Classification

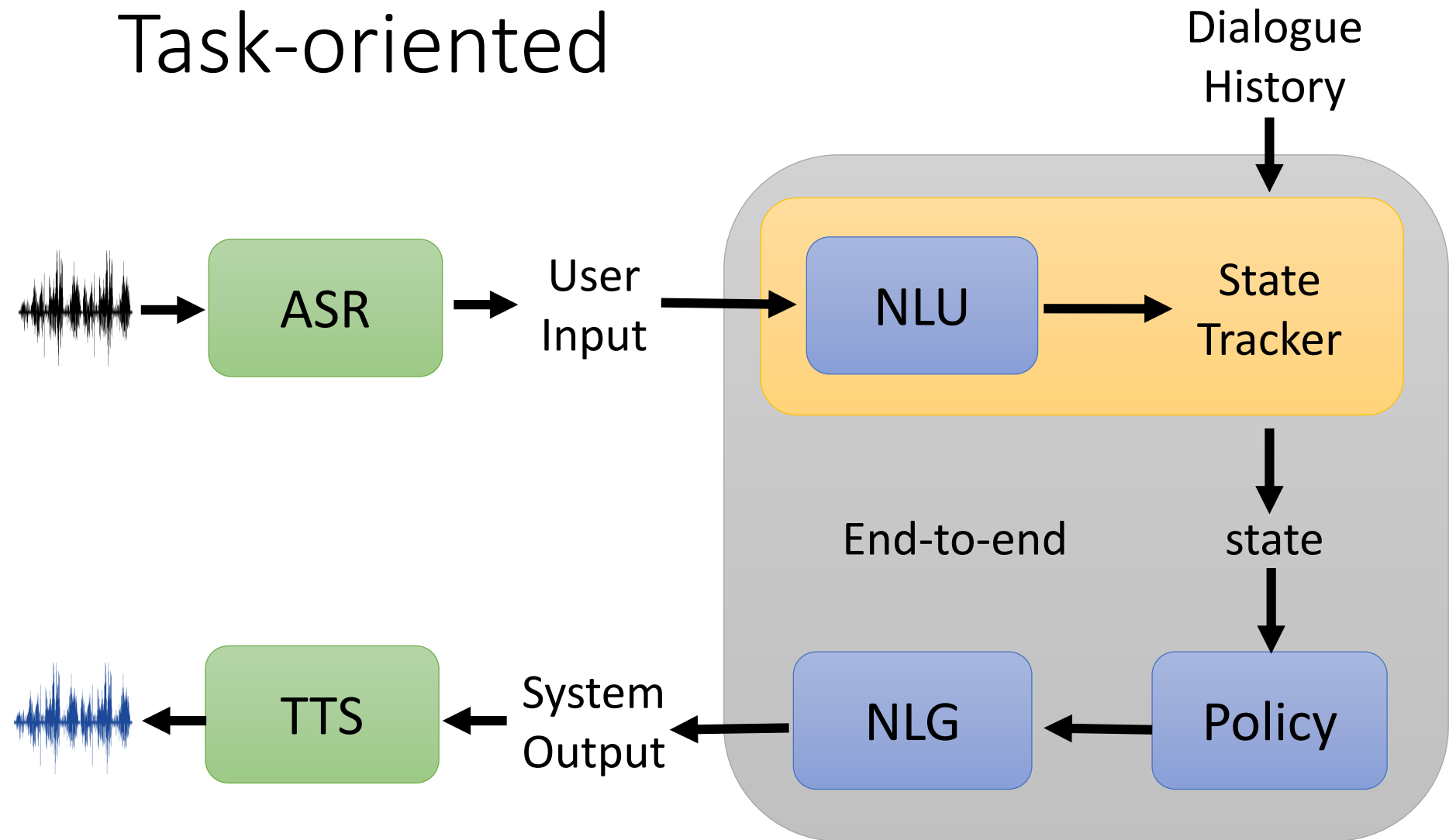
Provide Information

我 打算 9月 9日 入住、 9月 11日 退房

- Slot Filling

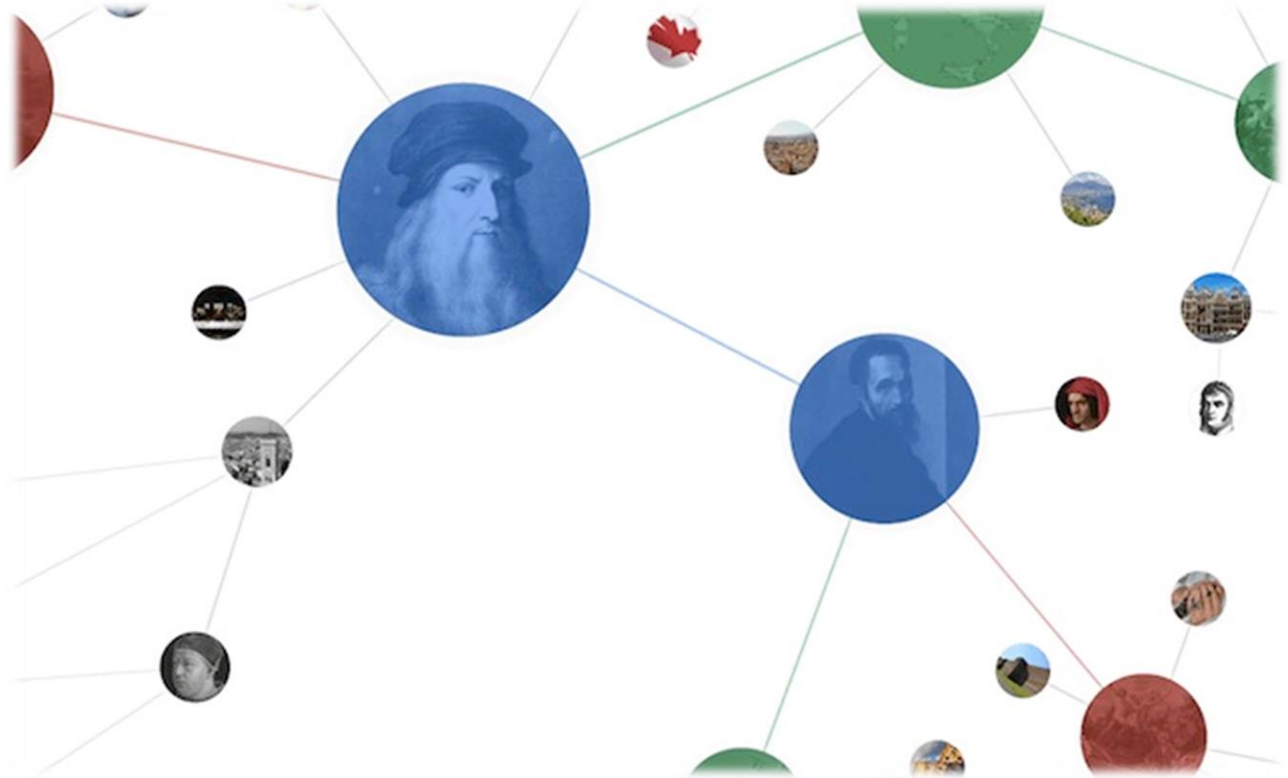


# Task-oriented

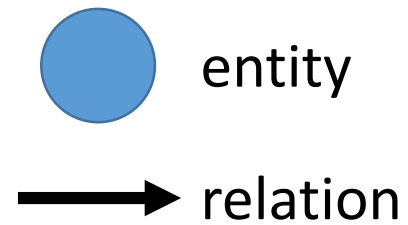


Input: several sequences, Output: a sequence

# Knowledge Extraction



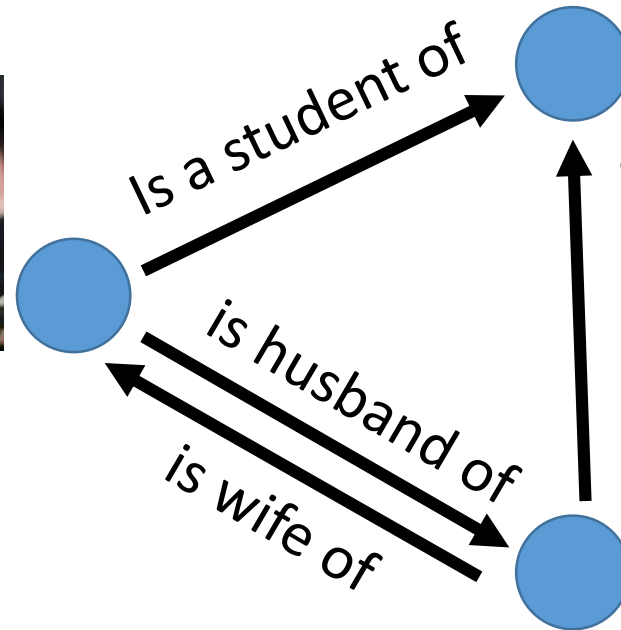
# Knowledge



Ron Weasley



Hogwarts



Hermione Granger

Step 1: Extract Entity

Step 2: Extract Relation

(warning: The following description oversimplifies the task)

# Name Entity Recognition (NER)

Harry Potter is a student of Hogwarts and lived on Privet Drive.



## Name entity recognition

people, organizations, places are usually  
name entity

Harry Potter is a student of Hogwarts and lived on Privet Drive.

Input: sequence

Output: class for each token

(just like POS tagging,  
slot filling)

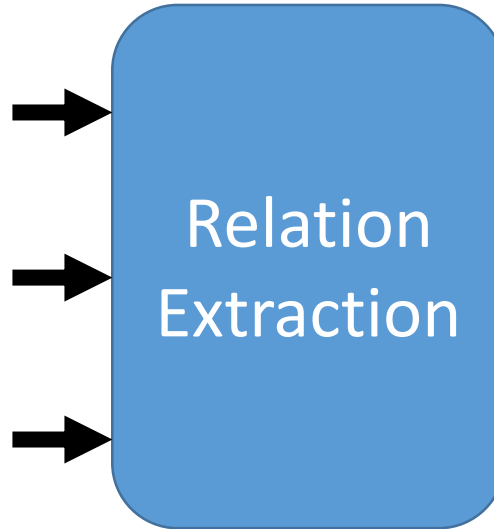
# Relation Extraction

Classification Problem

Harry Potter is a student  
of Hogwarts and lived on  
Privet Drive.

Harry Potter

Hogwarts

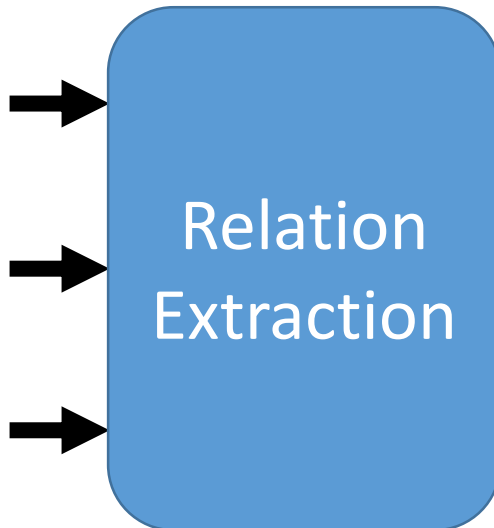


“is a student of”

Harry Potter is a student  
of Hogwarts and lived on  
Privet Drive.

Hogwarts

Privet Drive



“none”

# GLUE

## General Language Understanding Evaluation (GLUE)

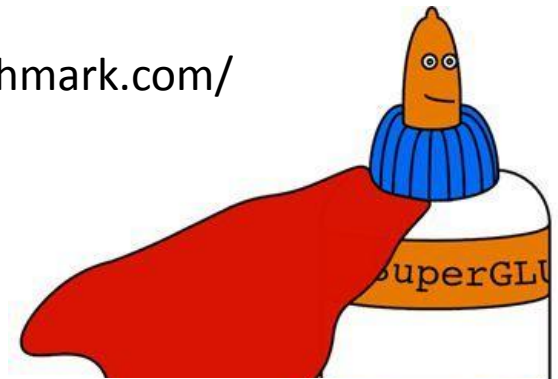
<https://gluebenchmark.com/>

- Corpus of Linguistic Acceptability (CoLA)
- Stanford Sentiment Treebank (SST-2)
- Microsoft Research Paraphrase Corpus (MRPC)
- Quora Question Pairs (QQP)
- Semantic Textual Similarity Benchmark (STS-B)
- Multi-Genre Natural Language Inference (MNLI)
- Question-answering NLI (QNLI)
- Recognizing Textual Entailment (RTE)
- Winograd NLI (WNLI)

GLUE also has Chinese version (<https://www.cluebenchmarks.com/>)



# Super GLUE



**BoolQ** **Passage:** *Barq's – Barq's is an American soft drink. Its brand of root beer is notable for having caffeine. Barq's, created by Edward Barq and bottled since the turn of the 20th century, is owned by the Barq family but bottled by the Coca-Cola Company. It was known as Barq's Famous Olde Tyme Root Beer until 2012.*

**Question:** *is barq's root beer a pepsi product*    **Answer:** No

**CB** **Text:** *B: And yet, uh, I we-, I hope to see employer based, you know, helping out. You know, child, uh, care centers at the place of employment and things like that, that will help out. A: Uh-huh. B: What do you think, do you think we are, setting a trend?*

**Hypothesis:** *they are setting a trend*    **Entailment:** Unknown

**COPA** **Premise:** *My body cast a shadow over the grass.*    **Question:** *What's the CAUSE for this?*

**Alternative 1:** *The sun was rising.*    **Alternative 2:** *The grass was cut.*

**Correct Alternative:** 1

**MultiRC** **Paragraph:** *Susan wanted to have a birthday party. She called all of her friends. She has five friends. Her mom said that Susan can invite them all to the party. Her first friend could not go to the party because she was sick. Her second friend was going out of town. Her third friend was not so sure if her parents would let her. The fourth friend said maybe. The fifth friend could go to the party for sure. Susan was a little sad. On the day of the party, all five friends showed up. Each friend had a present for Susan. Susan was happy and sent each friend a thank you card the next week*

**Question:** *Did Susan's sick friend recover?*    **Candidate answers:** *Yes, she recovered (T), No (F), Yes (T), No, she didn't recover (F), Yes, she was at Susan's party (T)*

# Super GLUE



**ReCoRD** **Paragraph:** *(CNN) Puerto Rico on Sunday overwhelmingly voted for statehood. But Congress, the only body that can approve new states, will ultimately decide whether the status of the US commonwealth changes. Ninety-seven percent of the votes in the nonbinding referendum favored statehood, an increase over the results of a 2012 referendum, official results from the State Electoral Commission show. It was the fifth such vote on statehood. "Today, we the people of Puerto Rico are sending a strong and clear message to the US Congress ... and to the world ... claiming our equal rights as American citizens, Puerto Rico Gov. Ricardo Rossello said in a news release. @highlight Puerto Rico voted Sunday in favor of US statehood*

**Query** For one, they can truthfully say, "Don't blame me, I didn't vote for them," when discussing the <placeholder> presidency    **Correct Entities:** US

**RTE** **Text:** *Dana Reeve, the widow of the actor Christopher Reeve, has died of lung cancer at age 44, according to the Christopher Reeve Foundation.*

**Hypothesis:** *Christopher Reeve had an accident.*    **Entailment:** False

**WiC** **Context 1:** *Room and board.*    **Context 2:** *He nailed boards across the windows.*

**Sense match:** False

**WSC** **Text:** *Mark told Pete many lies about himself, which Pete included in his book. He should have been more truthful.*    **Coreference:** False

# DecaNLP

<https://decanlp.com/>

- 10 NLP tasks
- Solving by the same model

Decathlon



| <u>Question</u>   | <u>Context</u>   | <u>Answer</u>  |
|---|--|--|
| What is a major importance of Southern California in relation to California and the US?                             | ...Southern California is a <b>major economic center</b> for the state of California and the US....    | <b>major economic center</b>                                     |
| What is the translation from English to German?   | Most of the planet is ocean water.   | Der Großteil der Erde ist Meerwasser                             |
| What is the summary?  | <b>Harry Potter star Daniel Radcliffe</b> gains access to a reported <b>£320 million fortune</b> ...   | <b>Harry Potter star Daniel Radcliffe gets £320M fortune</b> ... |
| Hypothesis: Product and geography are what make cream skimming work. <b>Entailment</b> , neutral, or contradiction? | Premise: Conceptually cream skimming has two basic dimensions – product and geography.                 | <b>Entailment</b>  |
| Is this sentence <b>positive</b> or negative?   | A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film. | <b>positive</b>  |

|                             | <b>One Sequence</b>   | <b>Multiple Sequences</b>                       |
|-----------------------------|---|---|
| <b>One Class</b>            | Sentiment Classification<br>Stance Detection<br>Veracity Prediction<br>Intent Classification<br>Dialogue Policy | NLI<br>Search Engine<br>Relation Extraction     |
| <b>Class for each Token</b> | POS tagging<br>Word segmentation<br>Extraction Summarization<br>Slotting Filling<br>NER                         |   |
| <b>Copy from Input</b>      |   | Extractive QA                                   |
| <b>General Sequence</b>     | Abstractive Summarization<br>Translation<br>Grammar Correction<br>NLG   | General QA<br>Task Oriented Dialogue<br>Chatbot |
| <b>Other?</b>               | Parsing, Coreference Resolution   |   |