CONVERSATIONAL UI

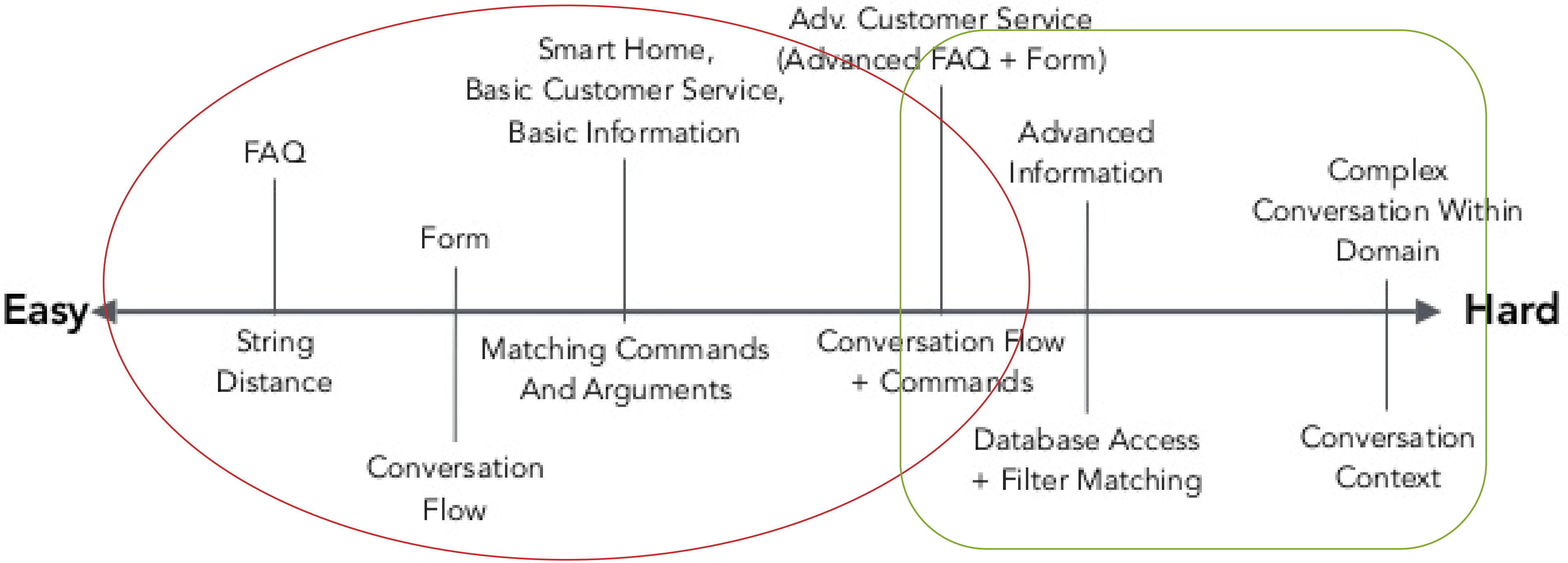
Natural Language Understanding

Dr. Aobo Wang
Email: isswan@nus.edu.sg

Agenda

- Types of CUIs
- CUI Workflow
- Task-Oriented CUI
  - Intent Detection
  - Slots Filling
- CUI for QA
  - Natural Language Understanding for FAQ
  - Natural Language Understanding for Machine-Reading QA

Types of CUI



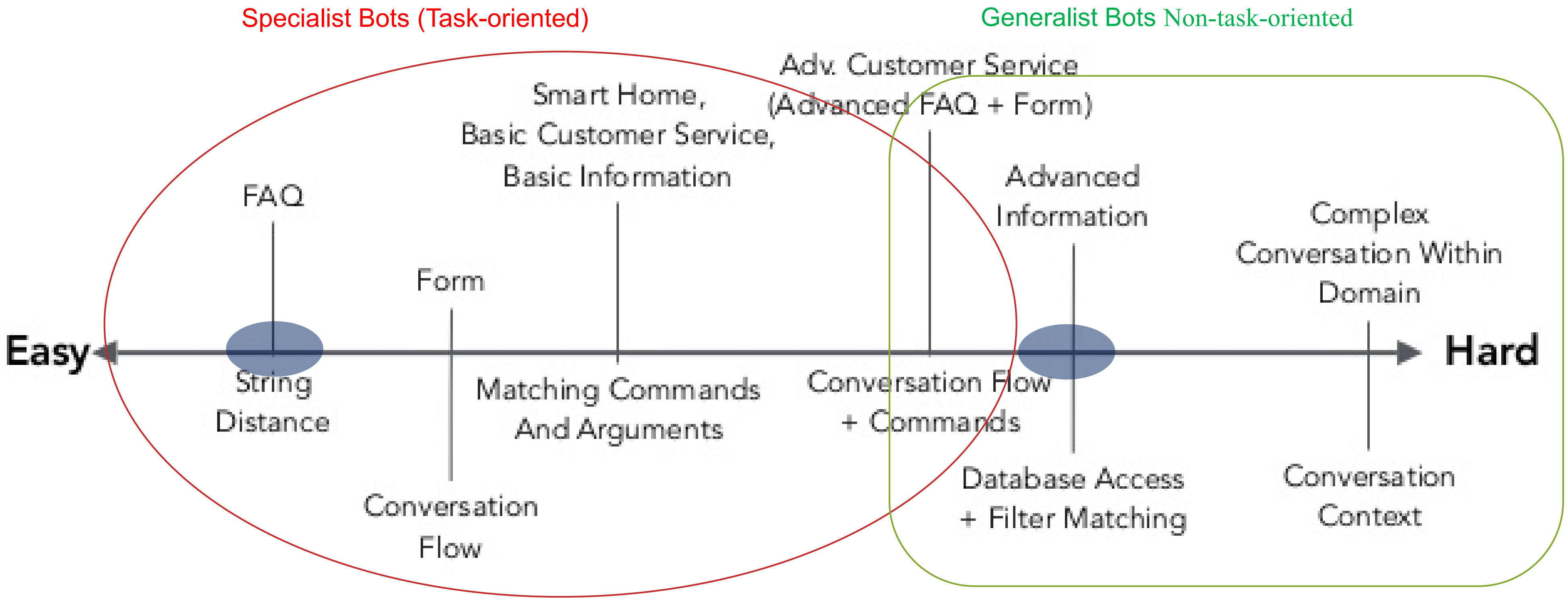
Specialist Bots

Task-oriented

Generalist Bots

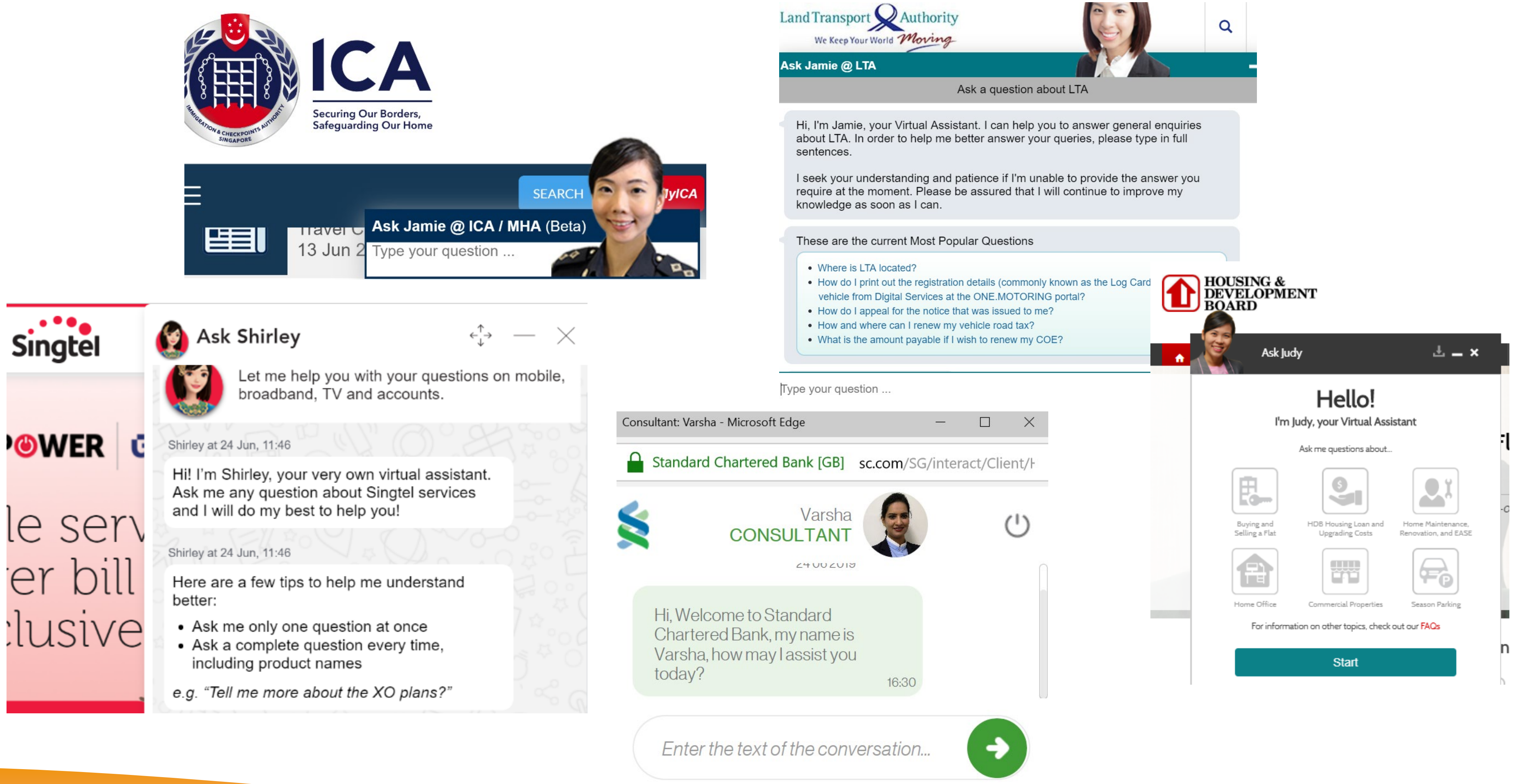
Non-task-oriented

Types of CUI



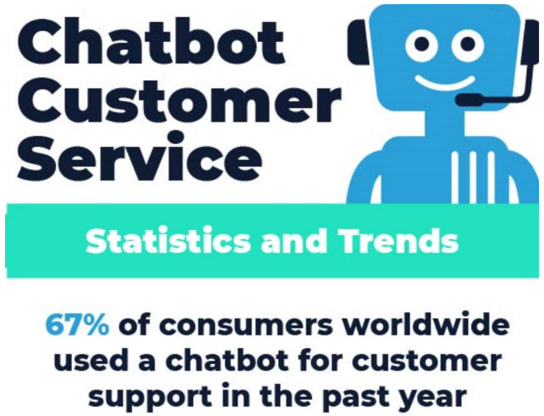
FAQ
Machine Reading Comprehension

CUI for QA



ICA
Securing Our Borders, Safeguarding Our Home
SEARCH
MyICA
Ask Jamie @ ICA / MHA (Beta)
Type your question ...
Land Transport Authority
We Keep Your World Moving
Ask Jamie @ LTA
Ask a question about LTA
Hi, I'm Jamie, your Virtual Assistant. I can help you to answer general enquires about LTA. In order to help me better answer your queries, please type in full sentences.
I seek your understanding and patience if I'm unable to provide the answer you require at the moment. Please be assured that I will continue to improve my knowledge as soon as I can.
These are the current Most Popular Questions
Where is LTA located?
How do I print out the registration details (commonly known as the Log Card vehicle from Digital Services at the ONE.MOTORING portal?
How do I appeal for the notice that was issued to me?
How and where can I renew my vehicle road tax?
What is the amount payable if I wish to renew my COE?
Type your question ...
Singtel
Ask Shirley
Let me help you with your questions on mobile, broadband, TV and accounts.
Shirley at 24 Jun, 11:46
Hi! I'm Shirley, your very own virtual assistant. Ask me any question about Singtel services and I will do my best to help you!
Shirley at 24 Jun, 11:46
Here are a few tips to help me understand better:
Ask me only one question at once
Ask a complete question every time, including product names
e.g. "Tell me more about the XO plans?"
Consultant: Varsha - Microsoft Edge
Standard Chartered Bank [GB] sc.com/SG/interact/Client/
Varsha
CONSULTANT
Hi, Welcome to Standard Chartered Bank, my name is Varsha, how may I assist you today?
16:30
Enter the text of the conversation...
HOUSING & DEVELOPMENT BOARD
Ask Judy
Hello!
I'm Judy, your Virtual Assistant
Ask me questions about...
Buying and Selling a Flat
HDB Housing Loan and Upgrading Costs
Home Maintenance, Renovation, and EASE
Home Office
Commercial Properties
Season Parking
For information on other topics, check out our FAQs
Start

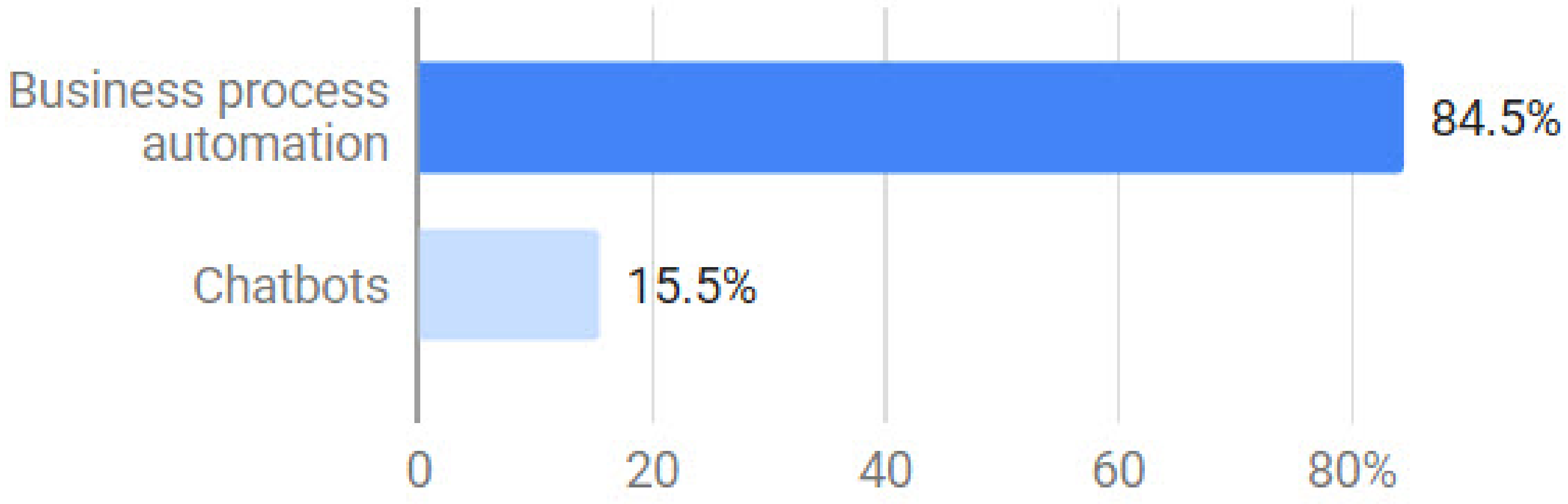
Why or Why Not CUI



What Can We Realistically Expect

1. Which of these two technology trends is more likely to make a positive impact on your business in the next 3 years?

200 respondent(s)

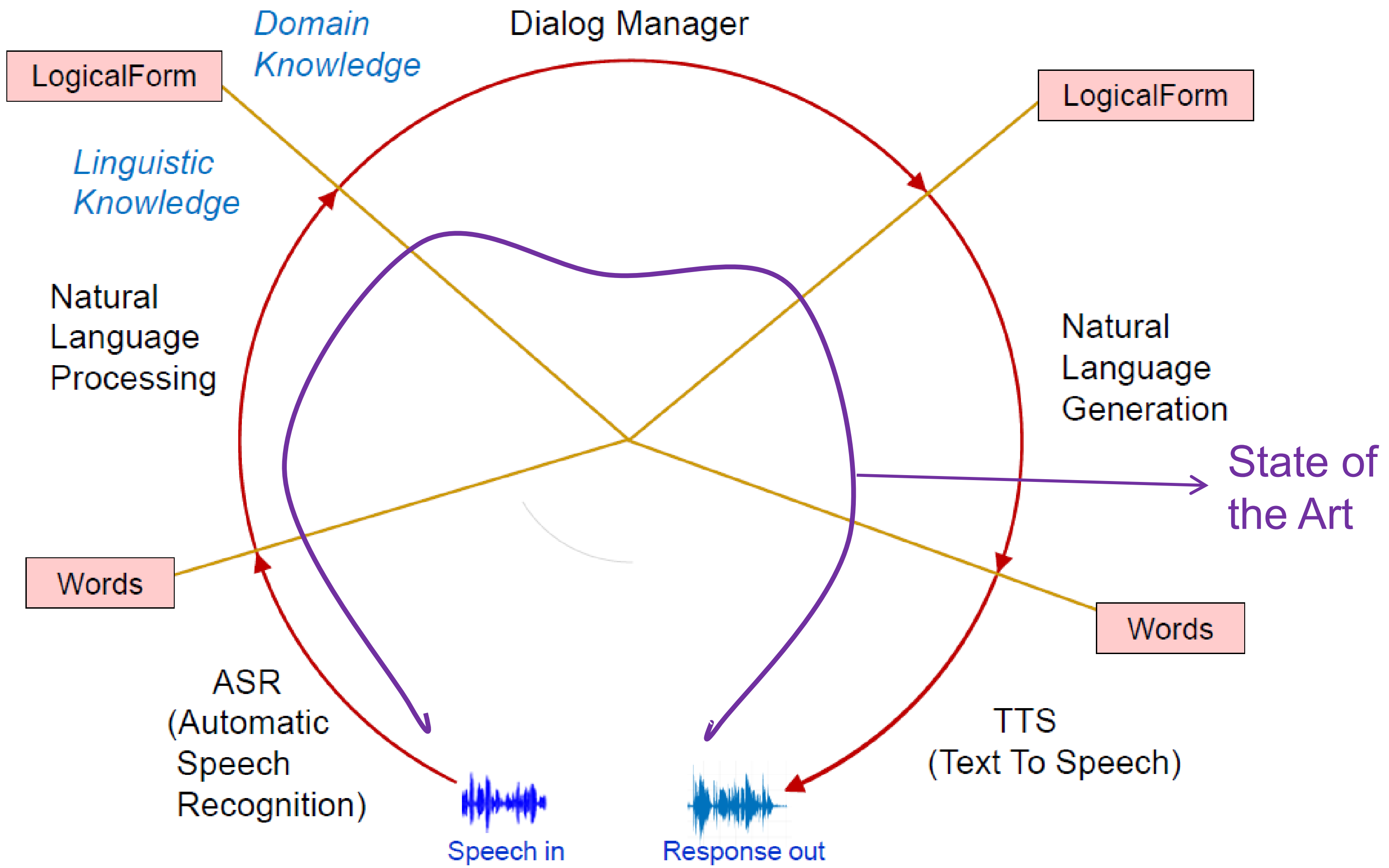


How Does CUI Work

CUI Workflow

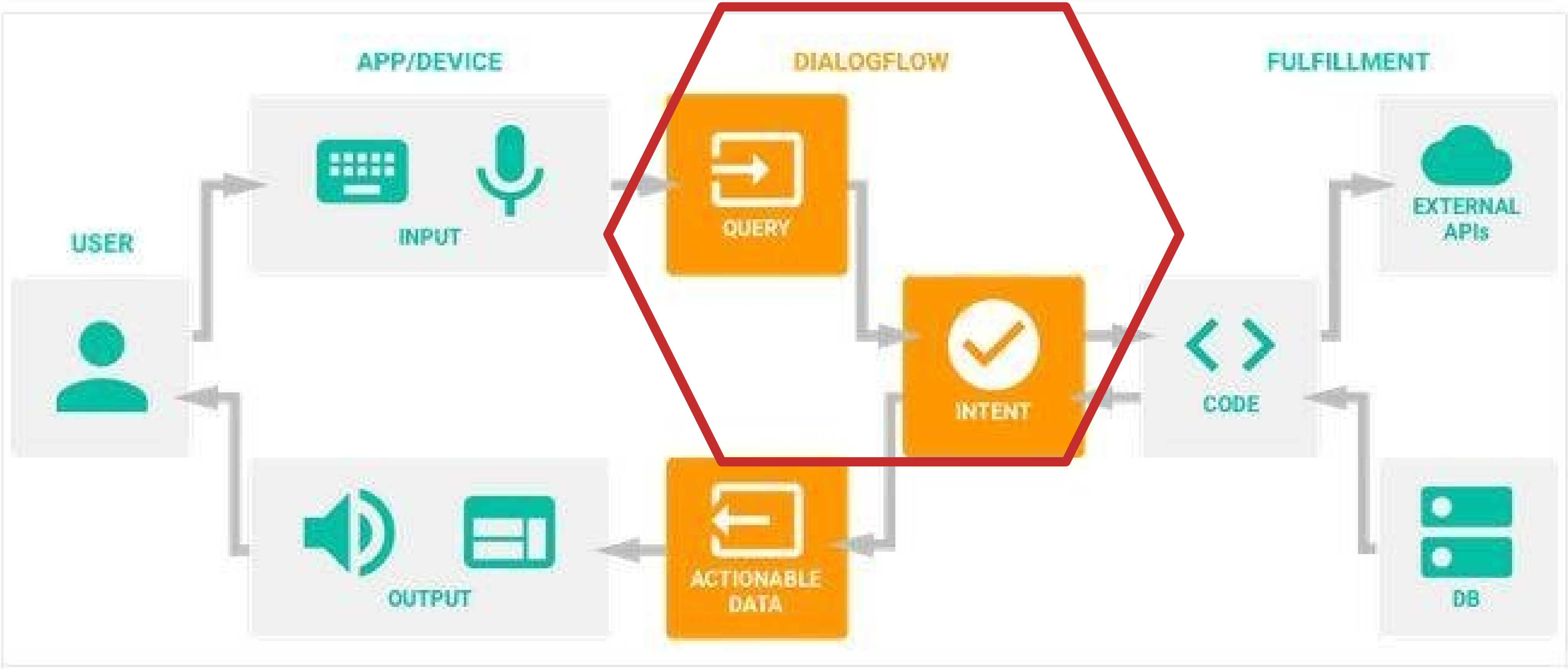
Task-Oriented CUI Workflow

- Alexa



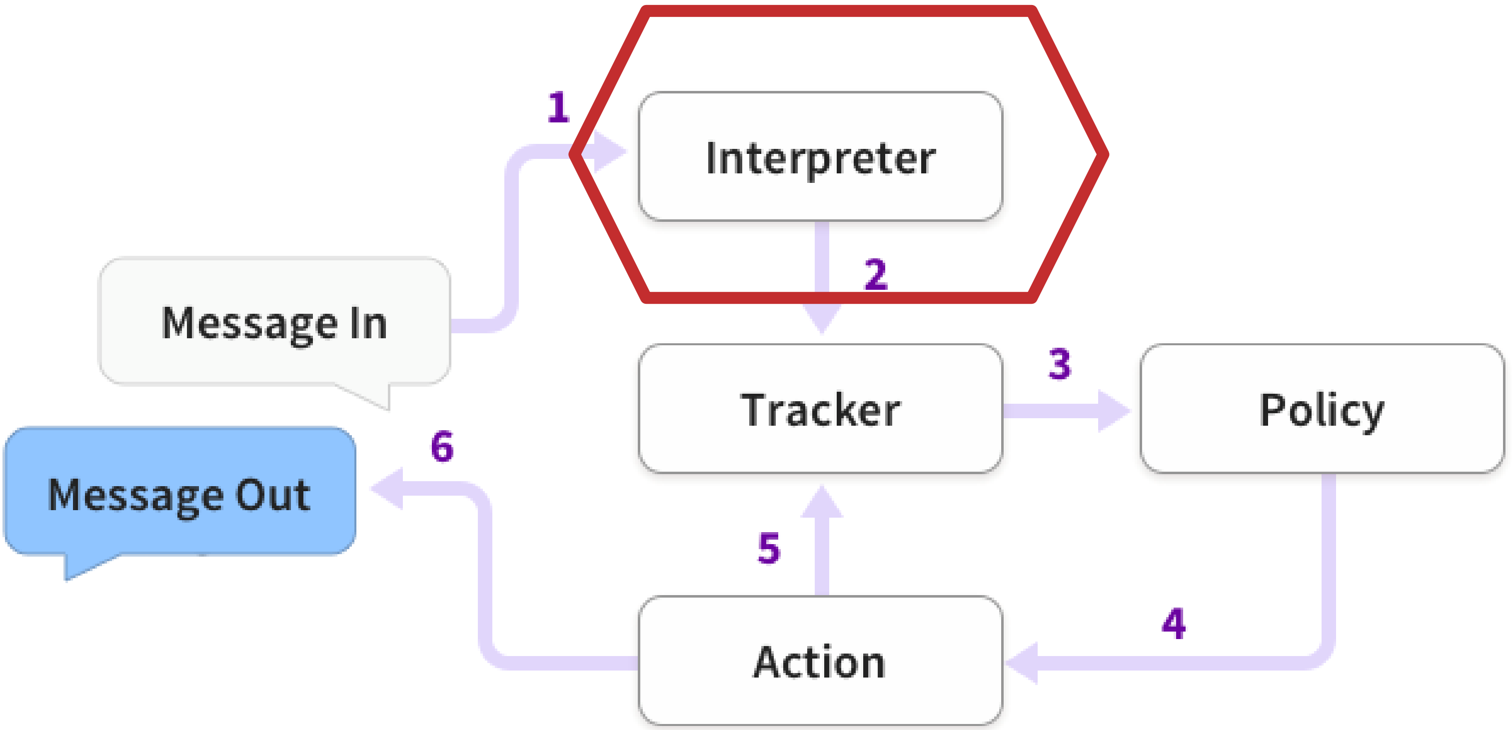
Task-Oriented CUI Workflow

- Google Dialogflow



Task-Oriented CUI Workflow

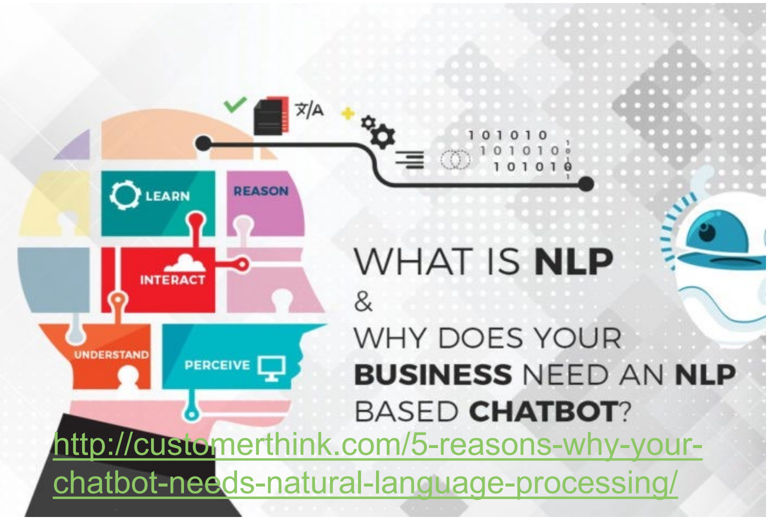
- RASA Architecture



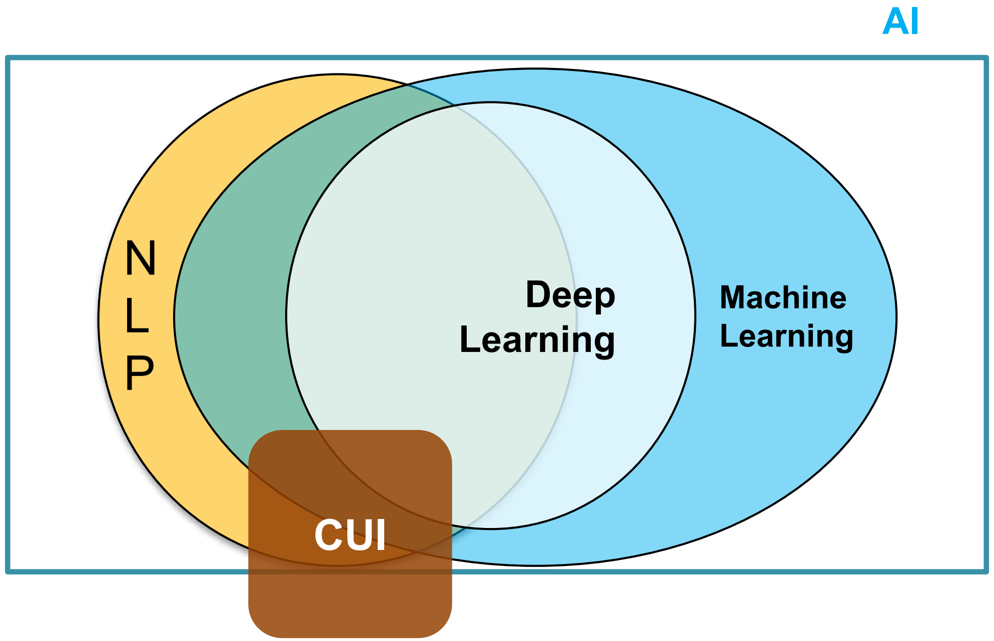
How is NLP relevant to CUI

Interpretable NLU

- **NLP Driven Approach**
  - Natural Language Processing： how to program computers to ***process*** and ***analyze*** large amounts of natural language data.



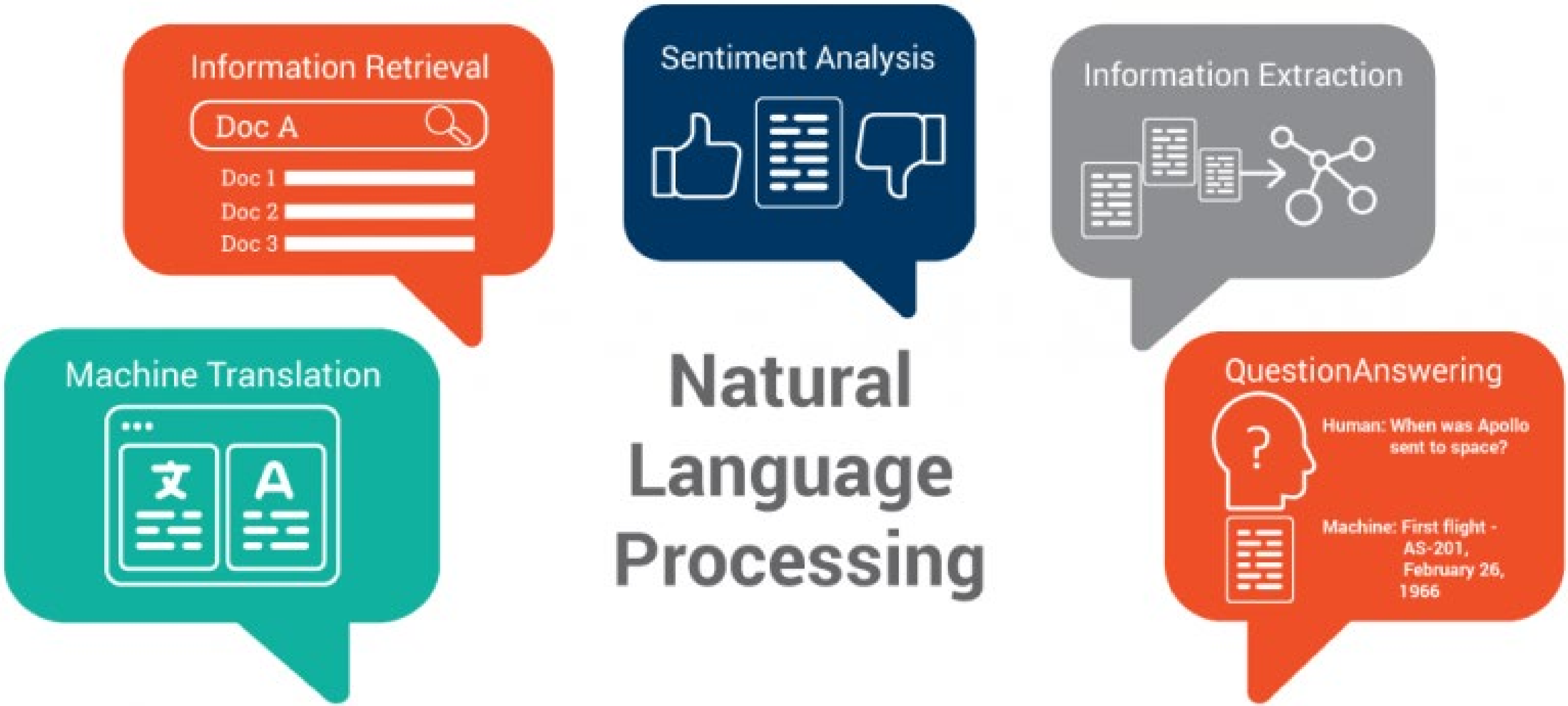
http://customerthink.com/5-reasons-why-your-chatbot-needs-natural-language-processing/



  - Using ML/DL, **NLP** is able to analyse ***unseen data***, especially when a *(huge)* volume of **example** *(training)* **data is available**.

How is NLP relevant to CUI

- **Typical Applications**



More Deeper Application of NLP

Group 1	Group 2	Group 3
Cleanup, Tokenization	Information Retrieval and Extraction (IR)	Machine Translation
Stemming	Relationship Extraction	Automatic Summarization/ Paraphracing
Lemmatization	Named Entity Recognation (NER)	Natural Language Generation
Part of Speech Tagging	Sentiment Analysis/Sentance Boundary Dismbiguation	Reasoning over Knowledge Based
Query Expansion	World sense and Dismbiguation	Quation Answering System
Parsing	Text Similarity	Dialog System
Topic Segmentationand Recognation	Coreference Resolution	Image Captioning & other Multimodel Tasks
Morphological Degmentation (Word/Sentences)	Discourse Analysis	

What Does NLP Do Exactly?

- **Q&A Related Task**

Paraphrase Identification

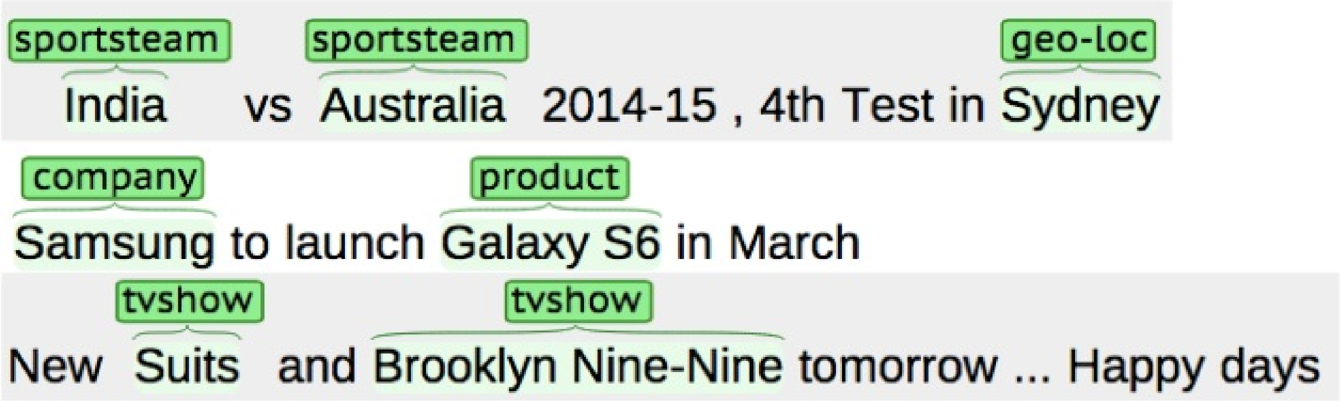
S: she struck a deal with RH to pen a book today
$+$: she signed a contract with RH to write a book
$-$: she denied today that she struck a deal with RH

Answer Sentence Selection

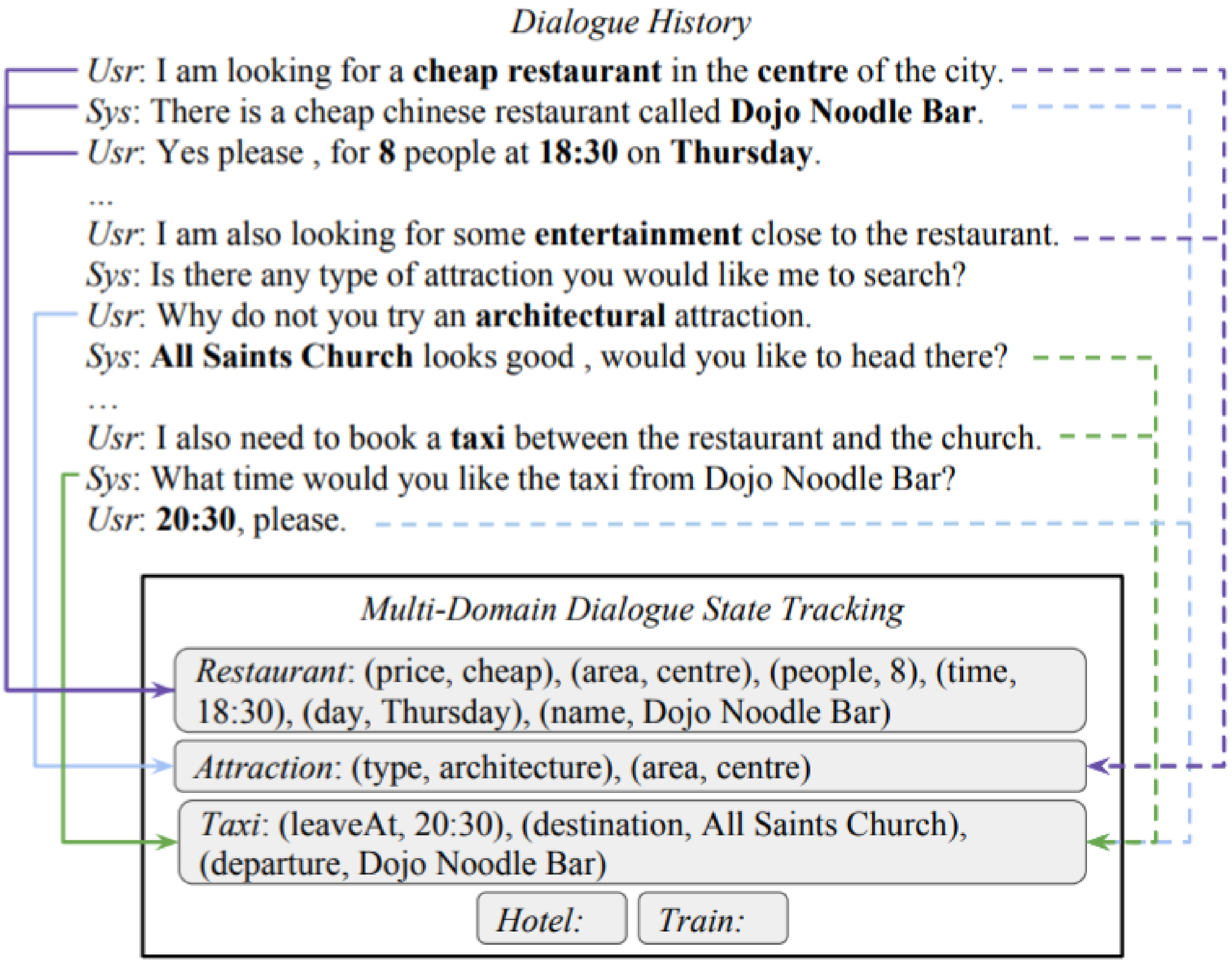
Q: how are glacier caves formed ?
$+$: a glacier cave is a cave formed within the ice of a glacier.
$-$: the ice facade is approximately 60m high

What Does NLP Do Exactly?

- **Named Entity Recognition**

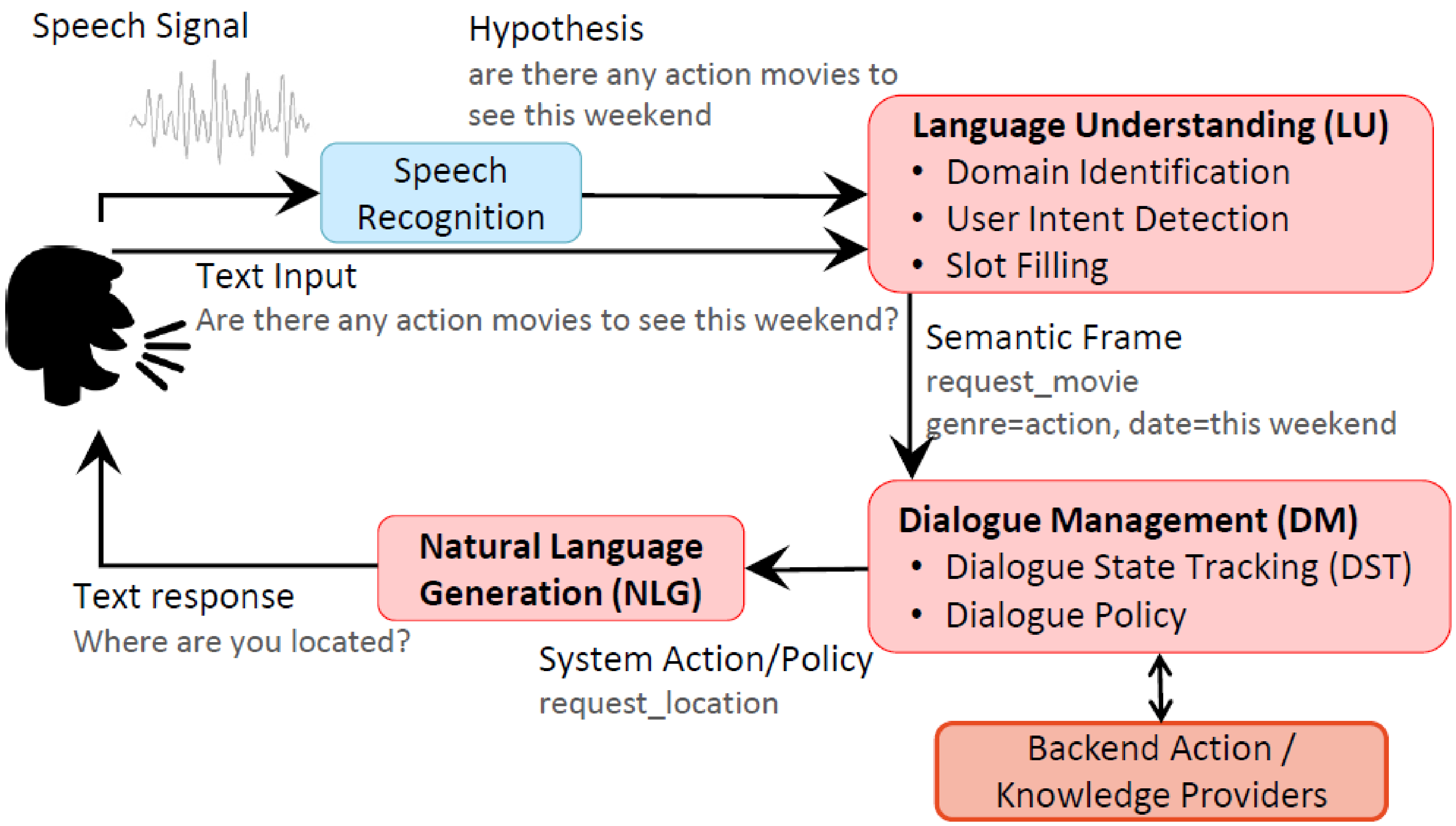


Dialog State Tracking



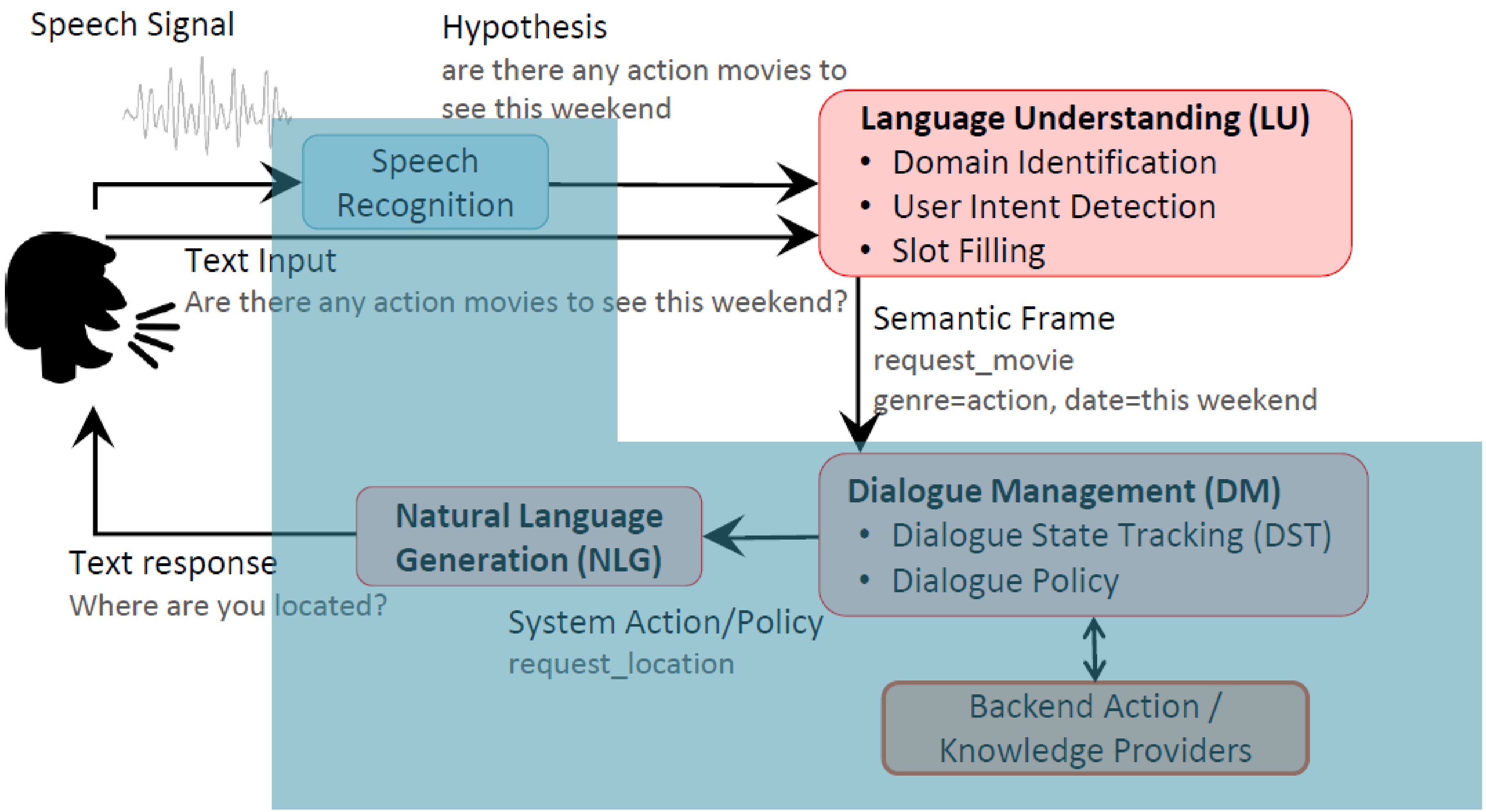
Task-oriented CUI

Core Architecture



https://sites.google.com/view/deepdial/

Core Architecture

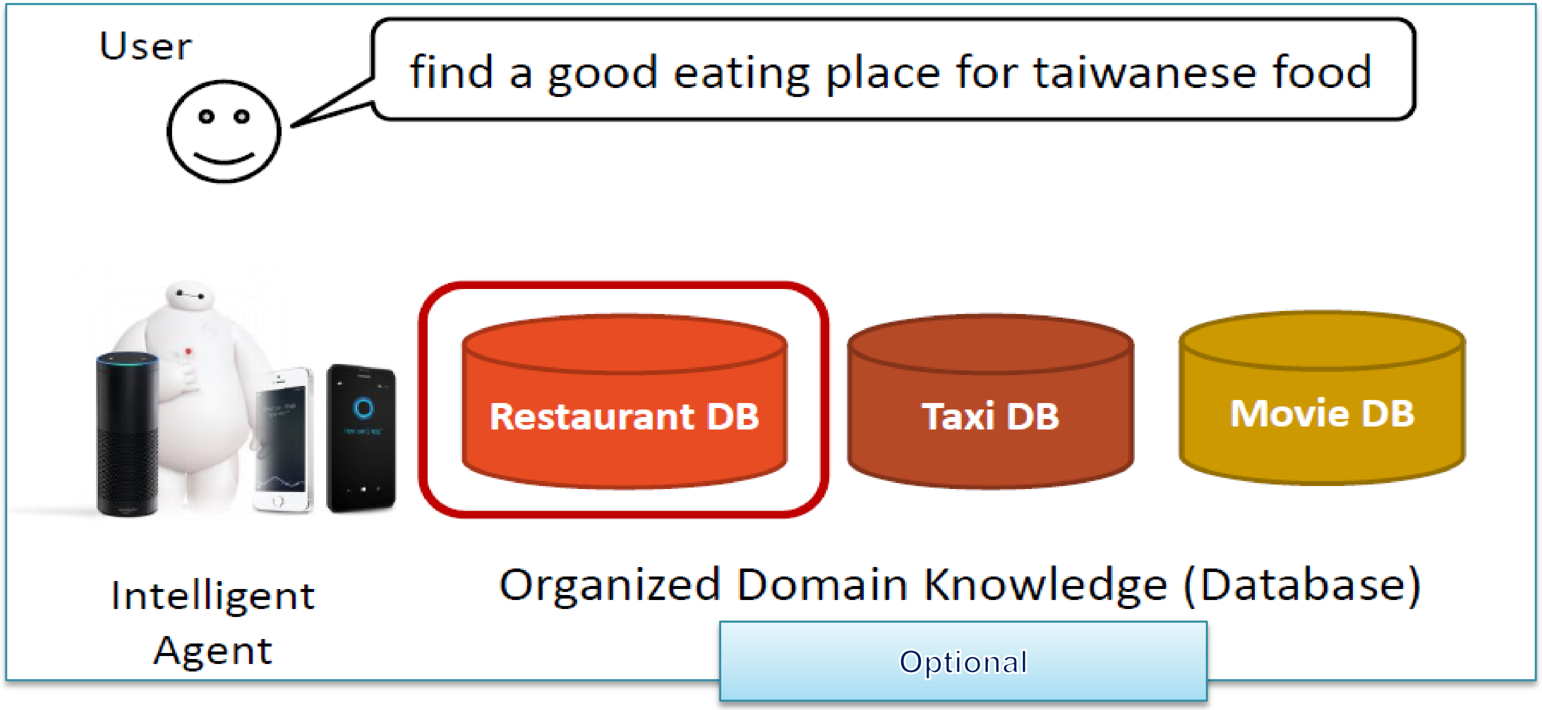


https://sites.google.com/view/deepdial/

Natural Language Understanding

Language Understanding

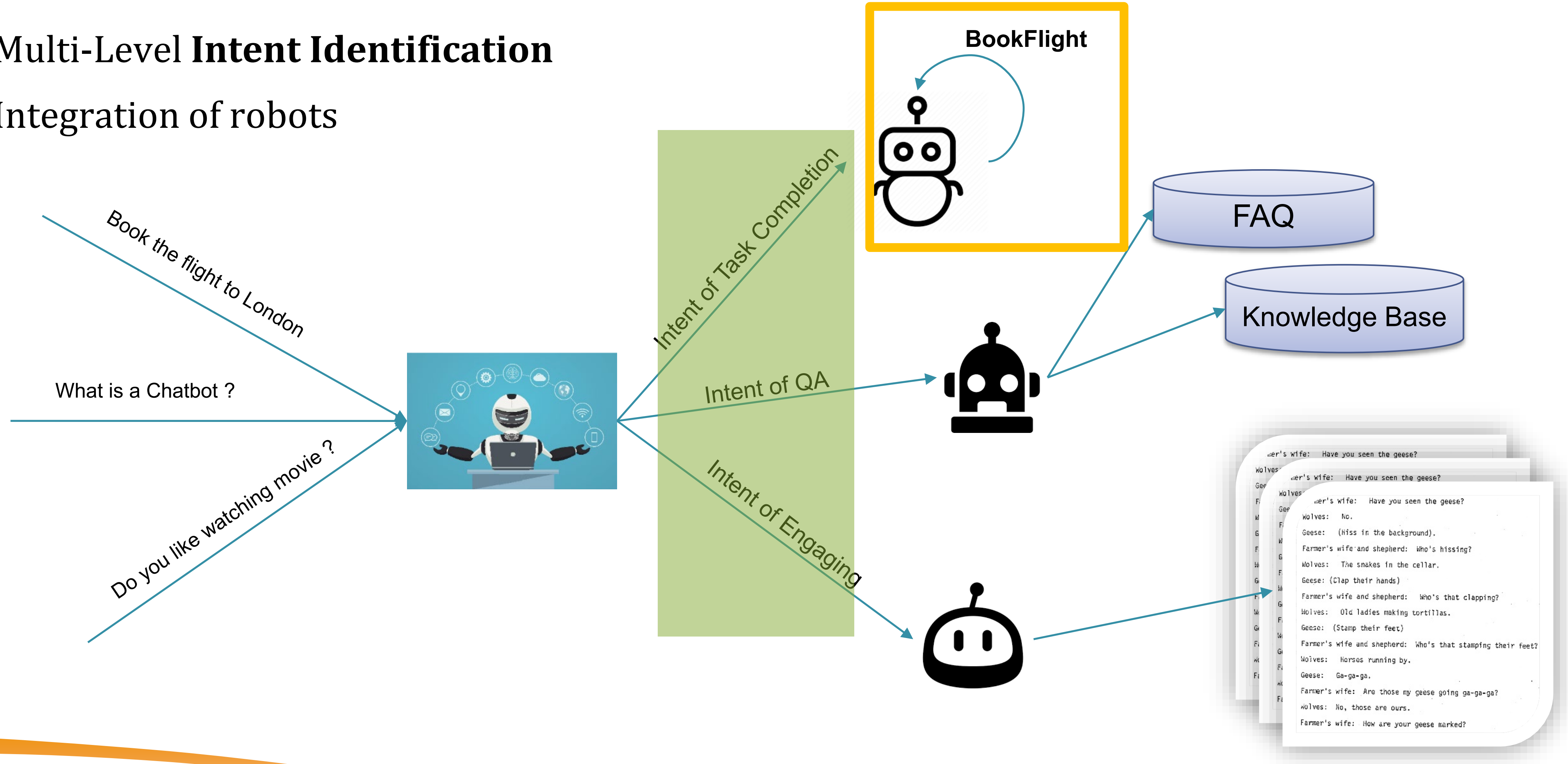
- **Domain Identification**
  - Classification (Supervised)



https://sites.google.com/view/deepdial/

Language Understanding

- Multi-Level **Intent Identification**
- Integration of robots



Natural Language Understanding for Task Completion

NLU for Task Completion

- **To understand the request:**
  - What is the weather in Seattle today?
  - What's the weather in London?
  - Tell me the temperature now.

sample utterances

Question-word
Command-word
Intent-word

WeatherForecast

Intent detection

Location-word
Time-word

Seattle / London
Today / Now

Slots Filling

.

Intent Detection

NLU Pipeline

- Deterministic : ***FST*** to compile sample ***utterances***
- Stochastic : Machine learning models for ***entity***, ***slot*** and ***intent*** prediction

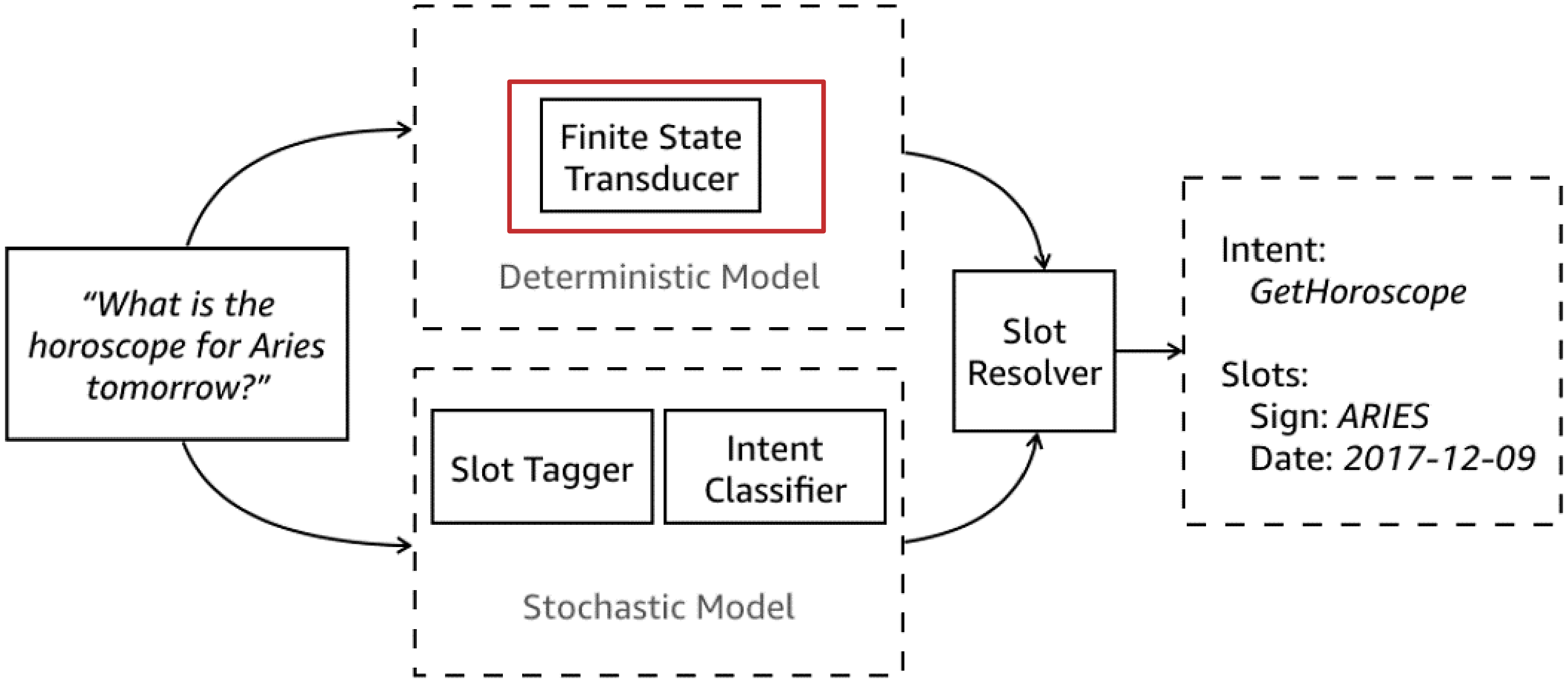
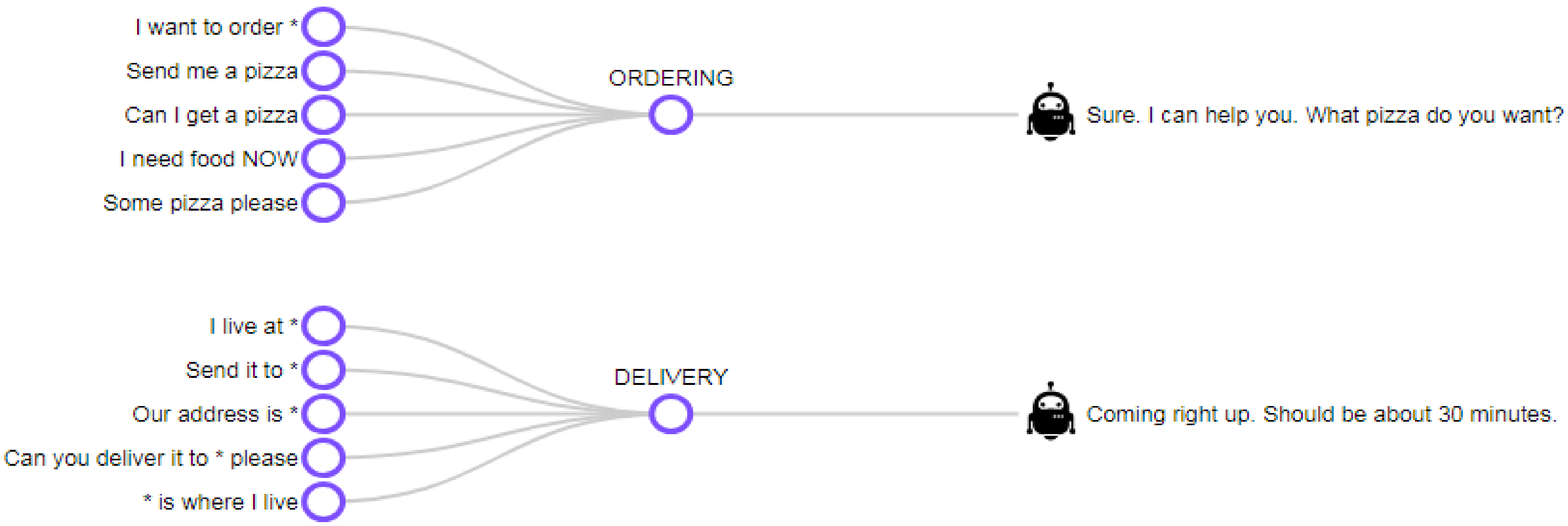


Image from 8. "Just ASK: building an architecture for extensible self-service spoken language understanding."

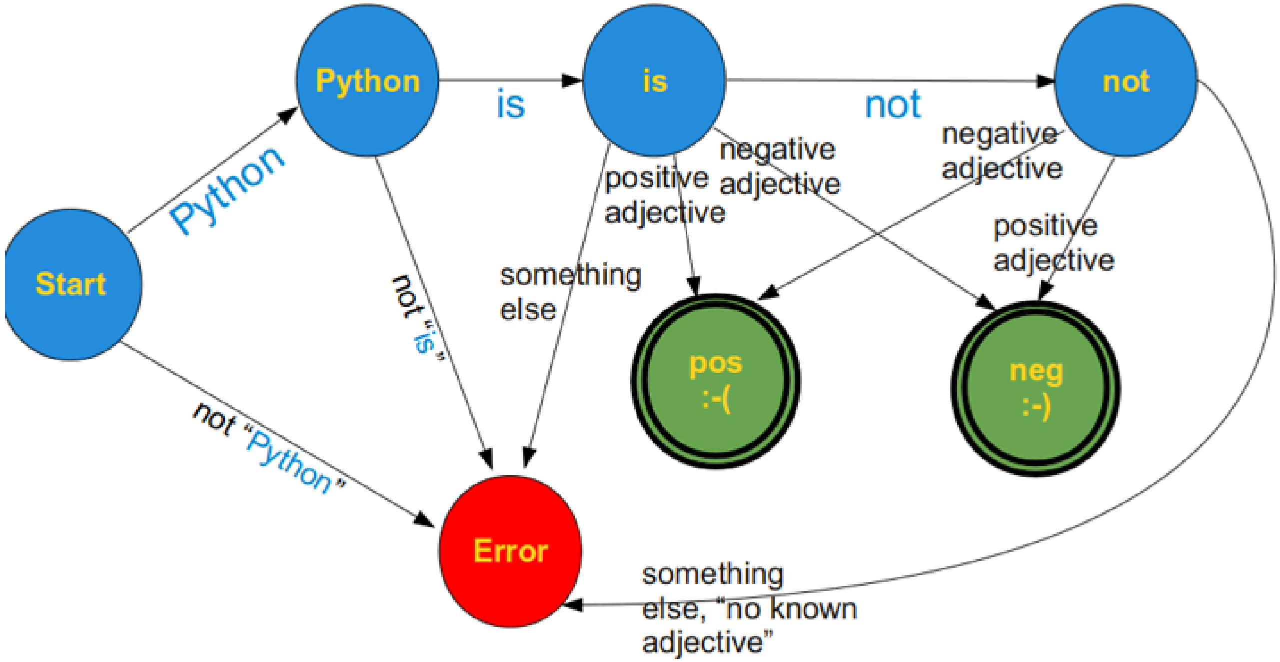
Deterministic Intent Detection

- **Rule based Approach**
  - Utterance and Response pairs are pre-defined in graphed architecture
  - Expensive human labour to build hand-craft rules
  - Templates are provided to minimize the coding effort



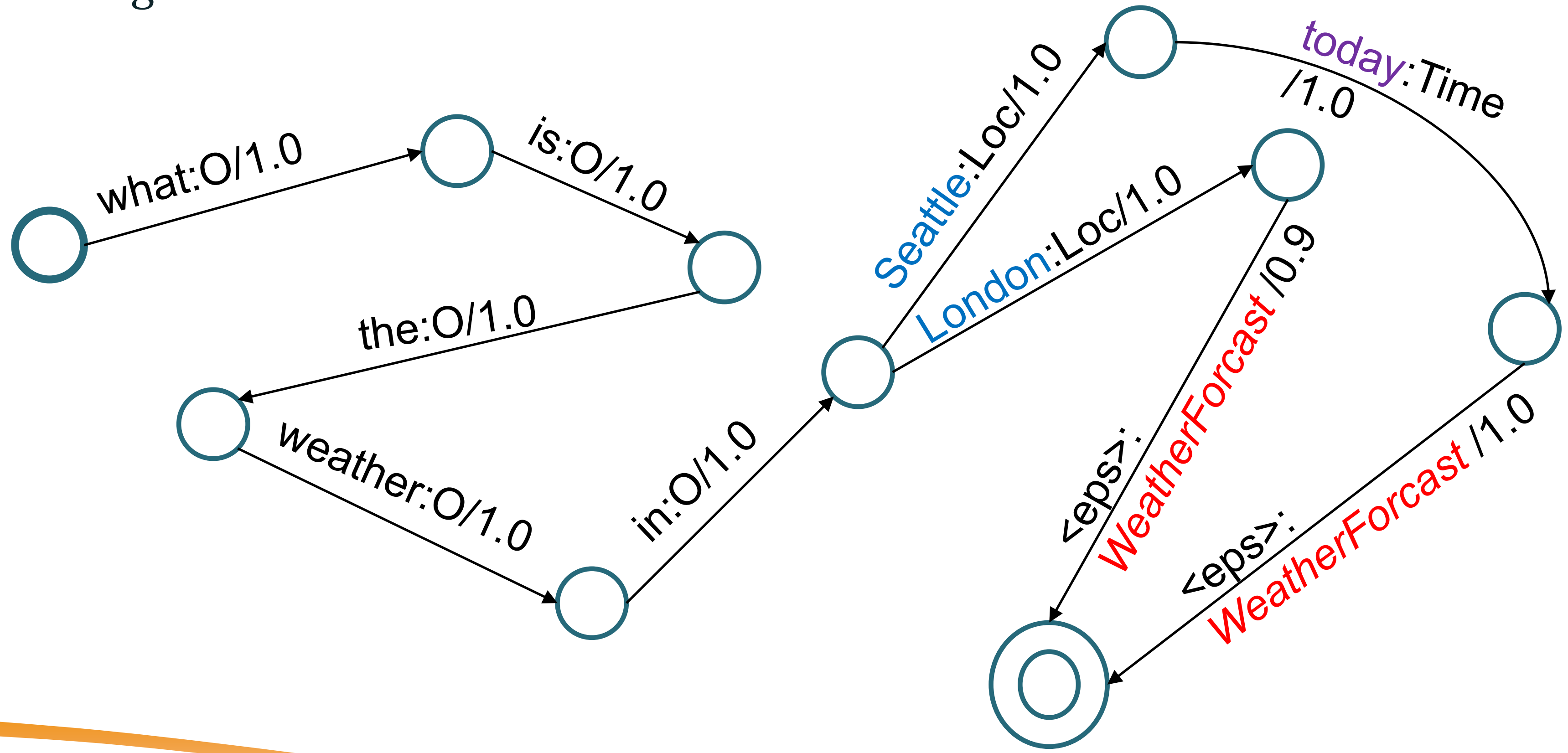
Deterministic Intent Detection

- Finite State Transducer
  - Starting state
  - Final states
  - Reaching any of the final states by consuming all the inputs



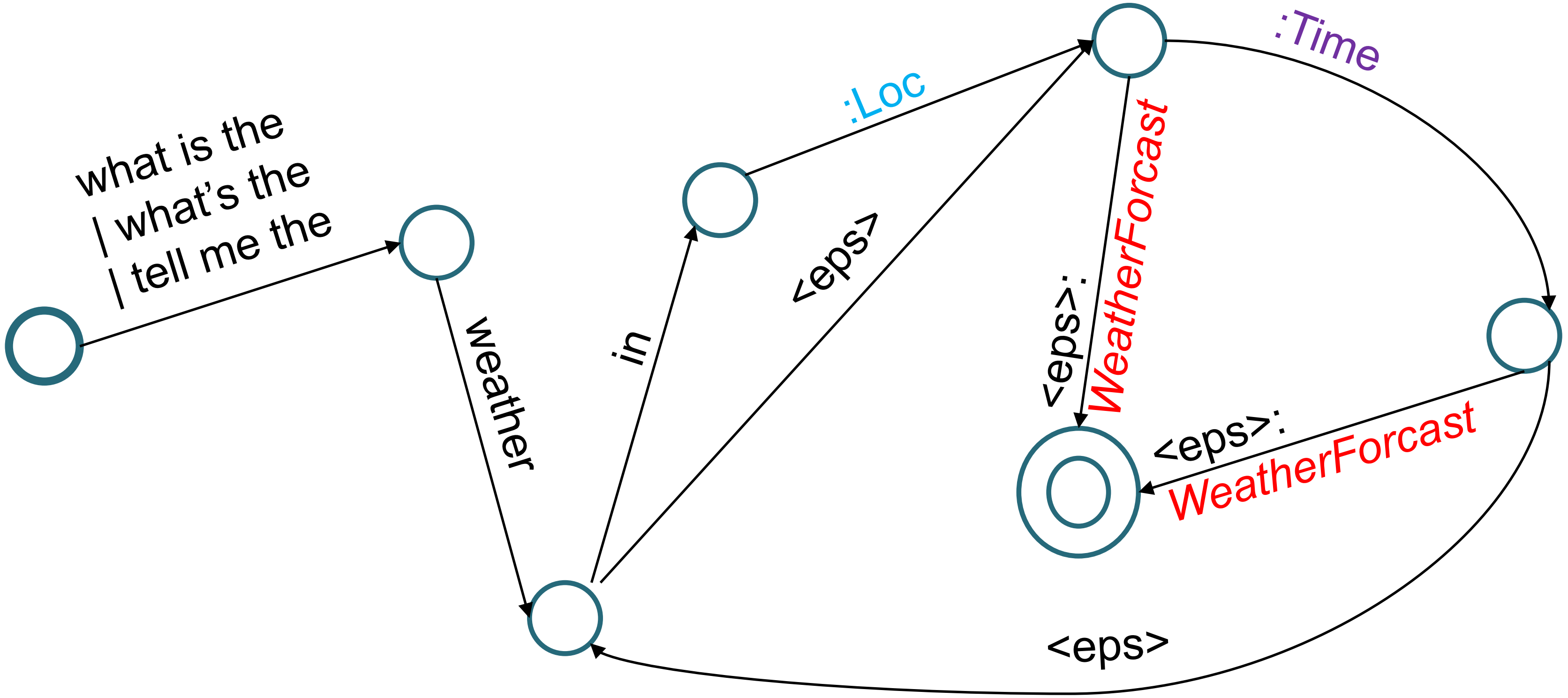
Deterministic Intent Detection

- Weighted Finite State Transducers
  - an easy but powerful way to represent data under a weighted grammar



Deterministic Intent Detection

- Simplified Finite State Transducers
  - Generalised by Named Entity Recognition
  - Dictionaries enrich the Knowledge Base



NLU Pipeline

- Deterministic : ***FST*** to compile sample ***utterances***
- Stochastic : Machine learning models for ***entity***, ***slot*** and ***intent*** prediction

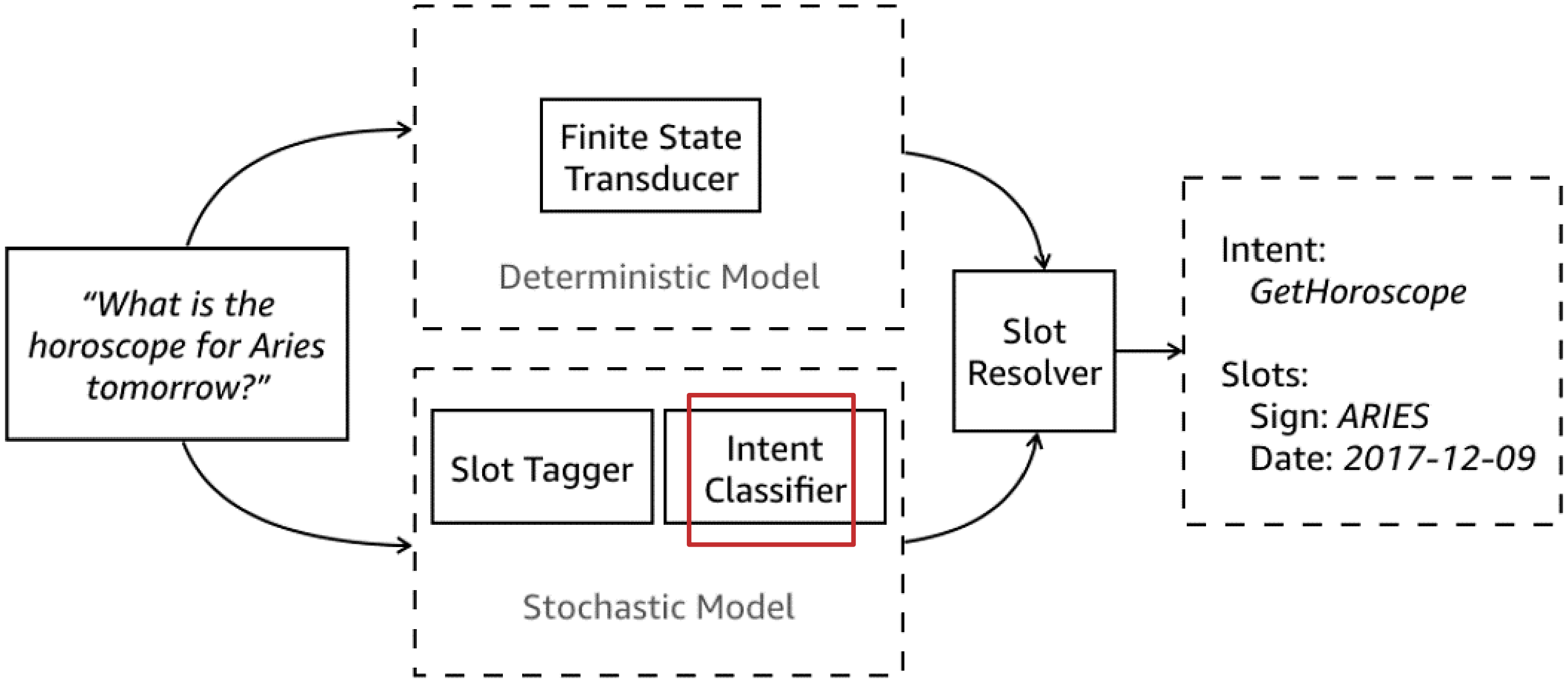
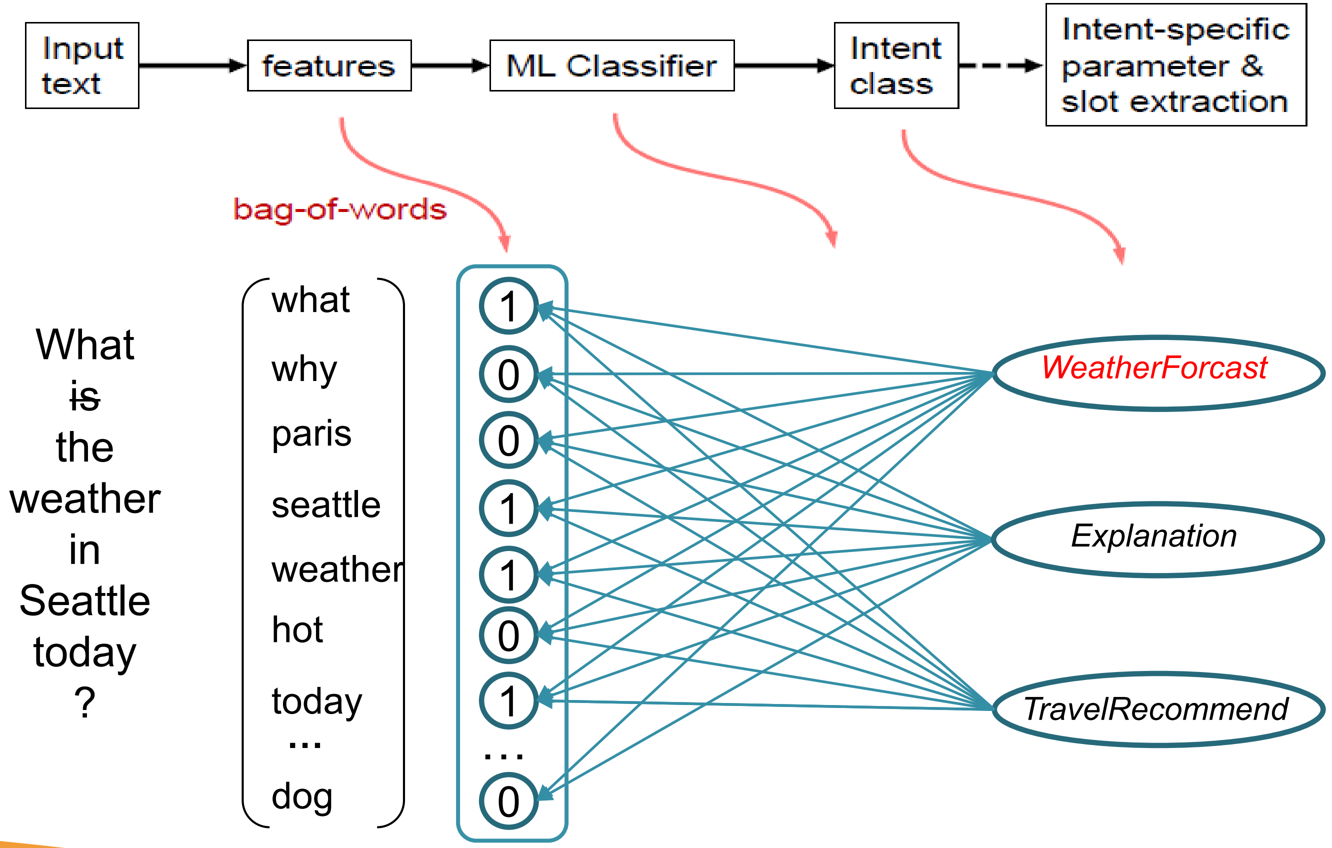


Image from 8. "Just ASK: building an architecture for extensible self-service spoken language understanding."

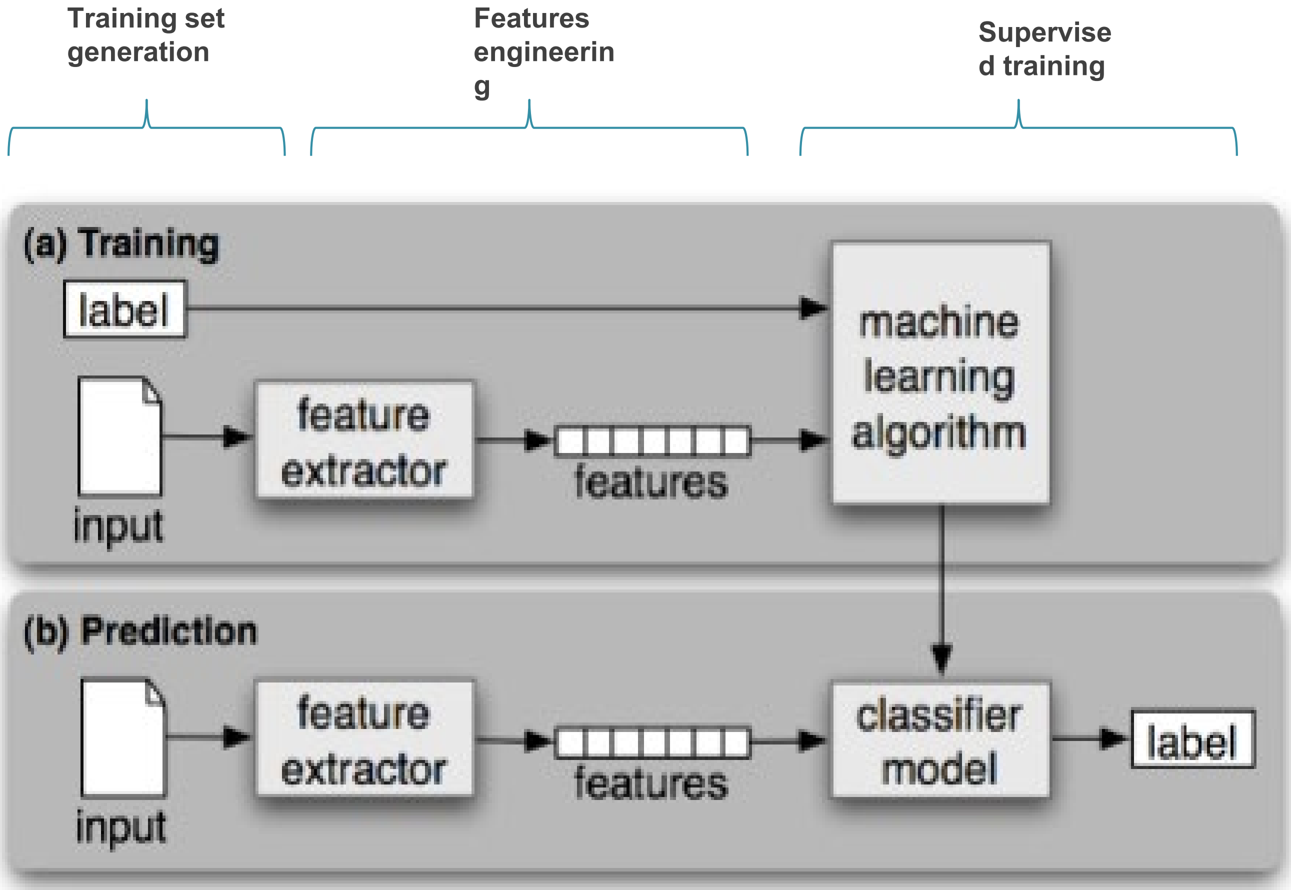
Stochastic Intent Detection

- Learn to organize the knowledge through modelling
  - ***Intent*** classification through machine learning (**supervised**) models



Stochastic Intent detection

- Overview of Supervised Classifier System
  - Data with Labels
  - Features engineering/embedding
  - **ML algorithm**
  - Evaluation



Naïve Bayes Classifier

- Bayes Rule classifier
    - $Prob(C_k|x_i) = Prob(x_i|C_k)Prob(C_k)$

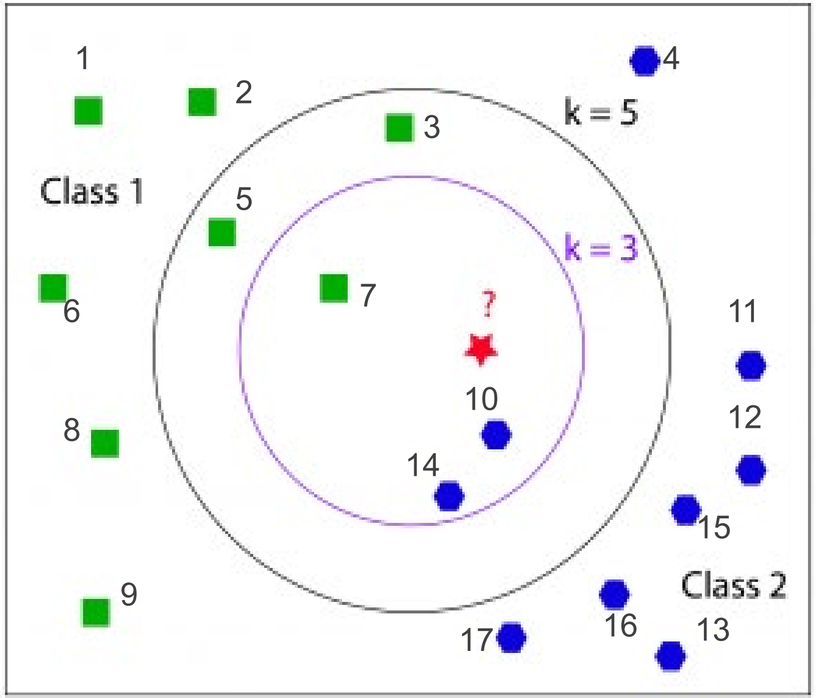
- Above for only *one* feature and class. Consider all classes and features. The class with the highest value is the most likely classifier category.

    - $\tilde{y} = \max_{k \in \{1,\ldots,K\}} \prod_i^n Prob(x_i|C_k)Prob(C_k)$

Here, C_k refers to class k. x_i is the feature.
So for example, k can mean ***Weather***, ***Health*** …
x_i can refer to any lexicon features or other feature 'engineered'

KNN Classifier

- Not to be confused with K-means clustering method



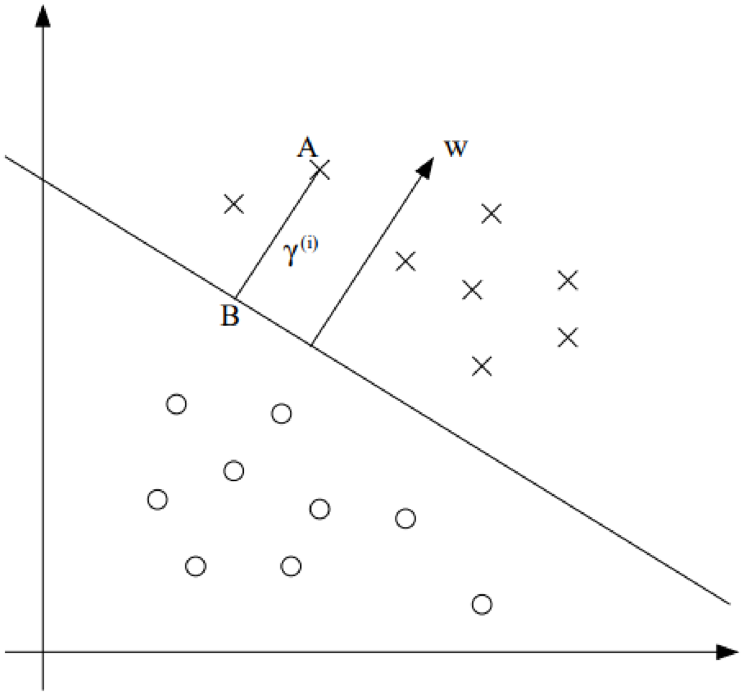
Which are the 3 closest points are to the red star?

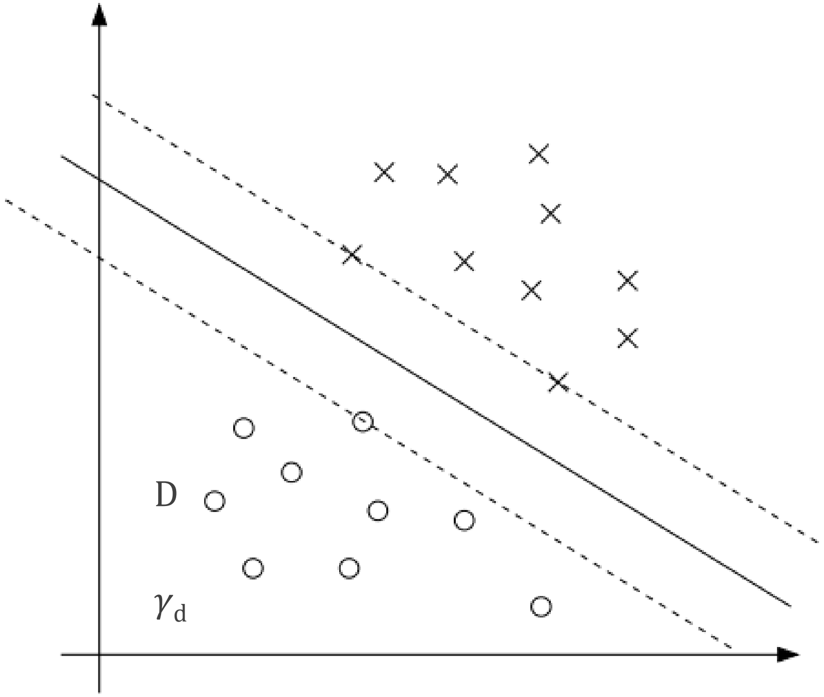
How about 5 closest?

If they take on the same importance in influencing the red star, what class will the red star be?

SVM classifier

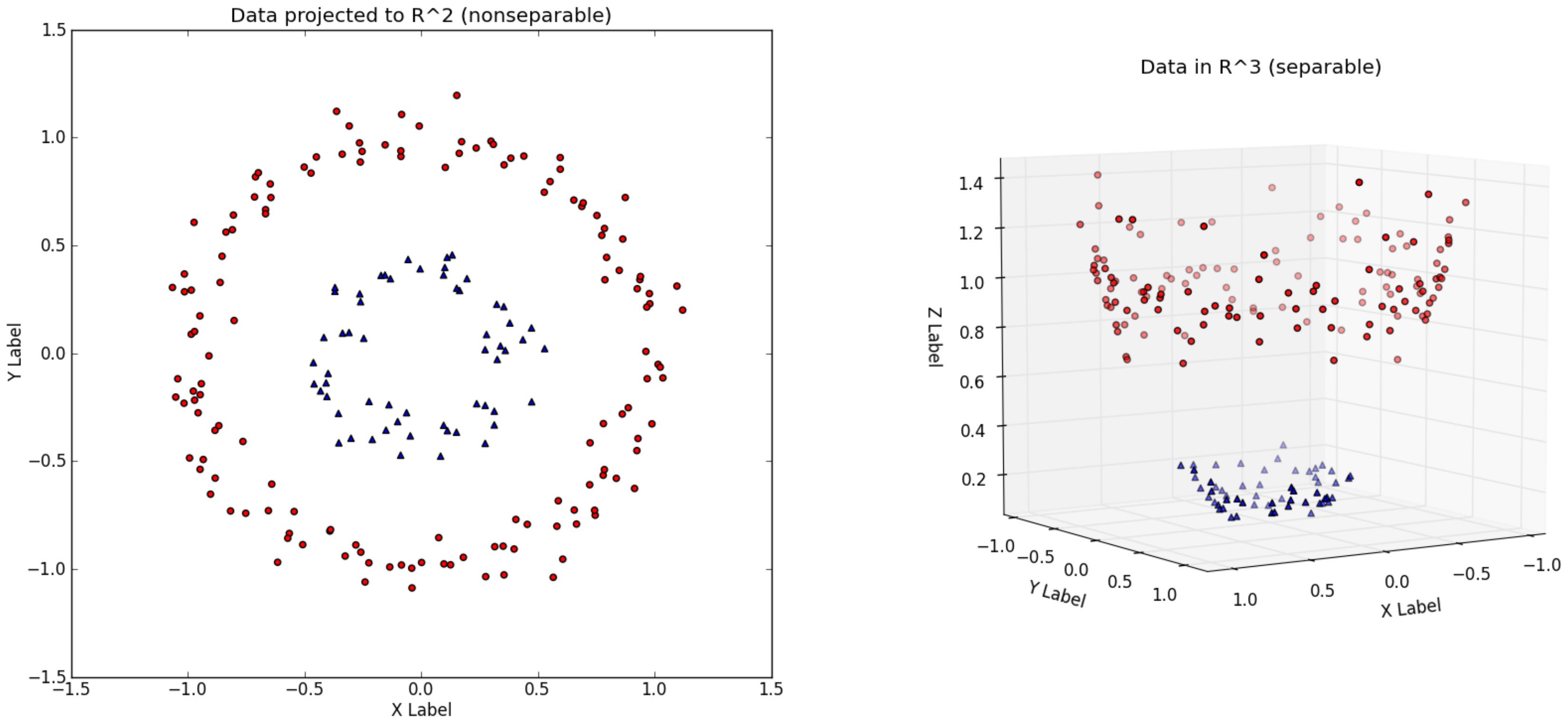
- The SVC is the most used amongst classifiers for NLP
- margin classifier





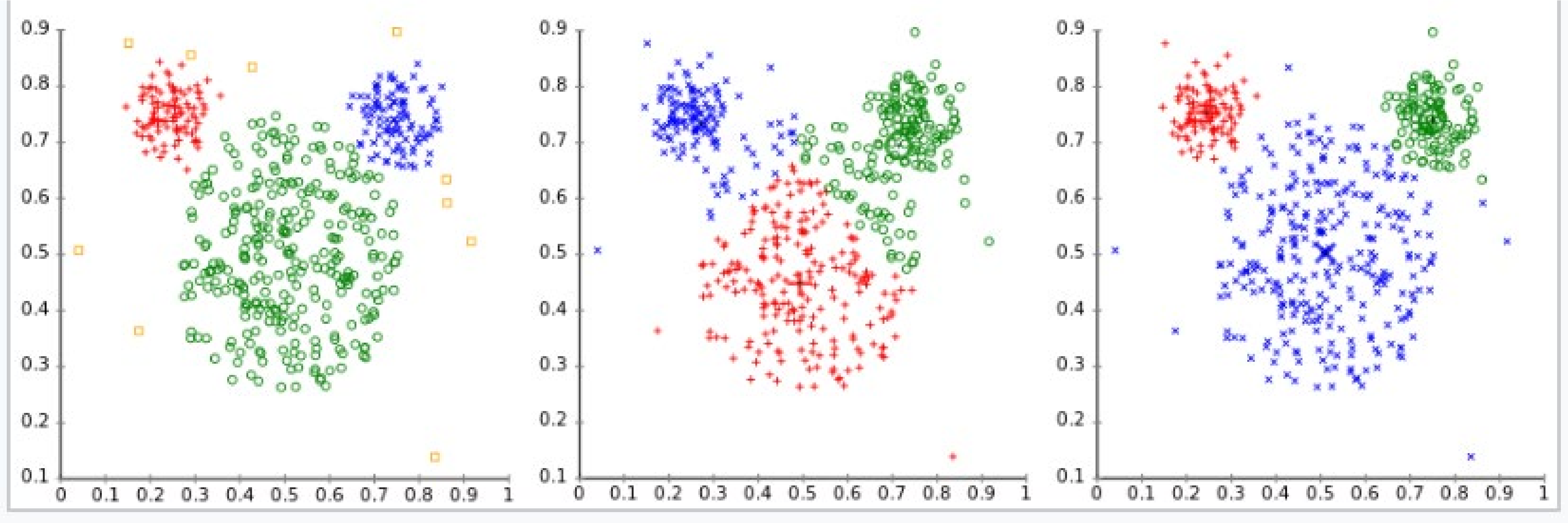
SVM classifier

- Kernel trick for separation



Comparing Classifiers

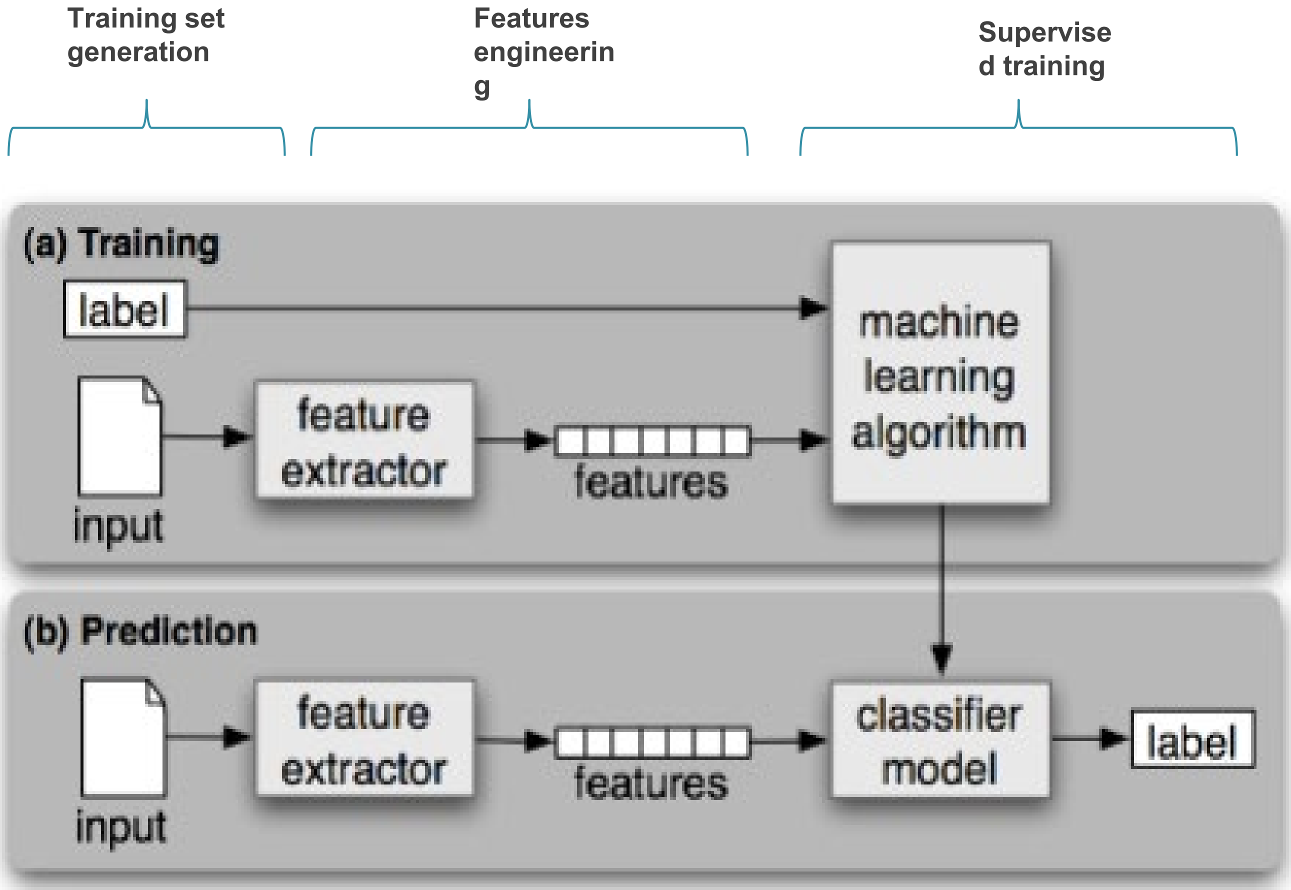
- Margin classifier is preferred



Original data KNN SVM

Features for Intent detection

- Overview of Supervised Classifier System
  - Data with Labels
  - **Features engineering/embedding**
  - ML algorithm
  - Evaluation



Utterances are Short

- **Features for Short Text Classification**
  - **Unigram/ Bigram**
  - Term Frequency
  - **TFIDF**
  - **POS tags**
  - Dependency Parsing
  - **Knowledge based Patterns**

Utterances are Short

- **Knowledge based Patterns**
  - Declarative Knowledge (*binary feature*)
    - "(***I*** | ***we*** | ***they***) (***want*** | ***like***) (to) (*)" → <u>*seeking-Pattern*</u>
    - "(***how to/why/what is***) (*)" → <u>*asking-Pattern*</u>
  - Social Behaviour Knowledge (*word count feature*)

Determiners (the)
Determiners (a, an)
Subject pronouns (she, he, we, they)
Mixed subjects/object pronouns but centered on individual (my, I, me)
Relative pronouns (that, this, these, those)
Possessive (mine, yours, his, hers, ours, theirs)
Relative pronouns (who, what, which, whom, whose)
Intensive/reflexive pronouns (myself, yourself, himself, herself, itself, ourselves, themselves, yourselves)
Dialogue management indicators (thanks, yes, ok, sorry, hi, hello, bye, anyway, how about, so, what do you mean, please, {could, would, should, can, will} followed by pronoun)
Hedge words (kinda, sorta)
Ambiguous pronoun (you)
Ambiguous pronoun (it)
Object pronouns (us, them, him, her)

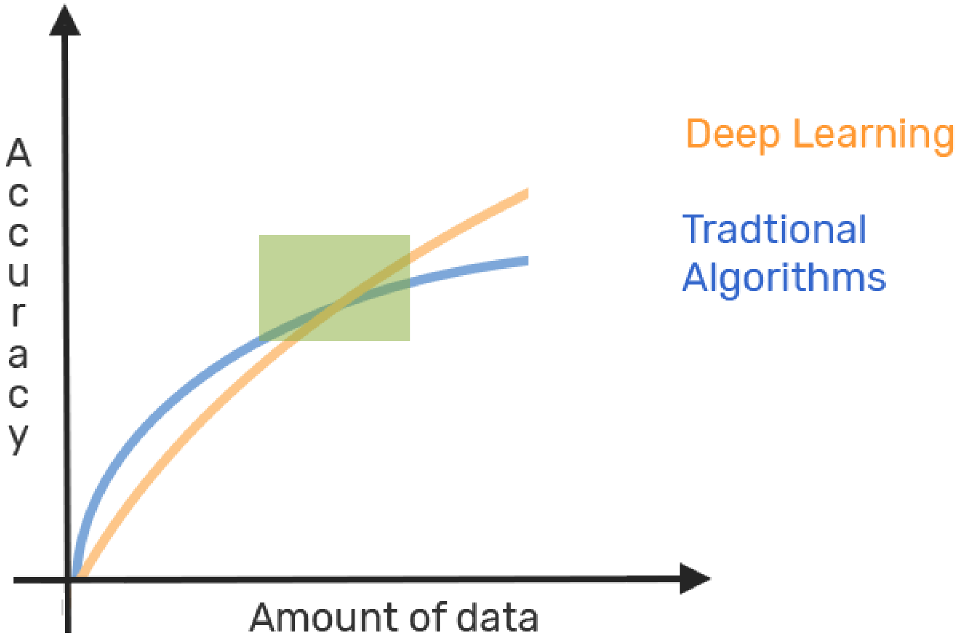
TABLE II
CONVERSATION INDICATORS FOR SOCIAL BEHAVIOR KNOWLEDGE

Utterances can be Embedded

- **Features engineering** + Classic machine learning classifiers
  - good performance, easy and fast training, interpretable, task-driven feature design
  - time consuming for feature engineering

- **Embedding** + Classic machine learning classifiers
  - good performance, easy and fast training, no feature engineering
  - Pre-trained ***wordtovec*** sometimes needs domain adaptation, difficult to interoperate

- **Embedding + NN**
  - usually better performance, no feature engineering
  - performance improvement could be limited, could be expensive on training, difficult to interoperate



CNN and RNN

- CNN and BiLSTM

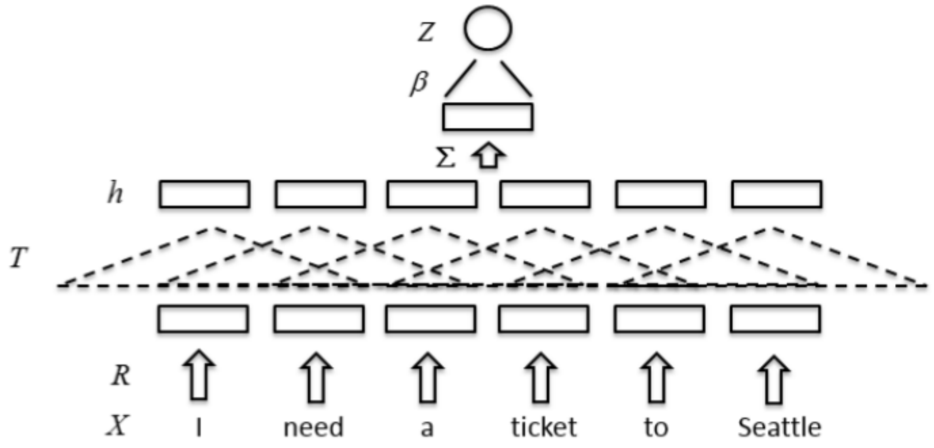
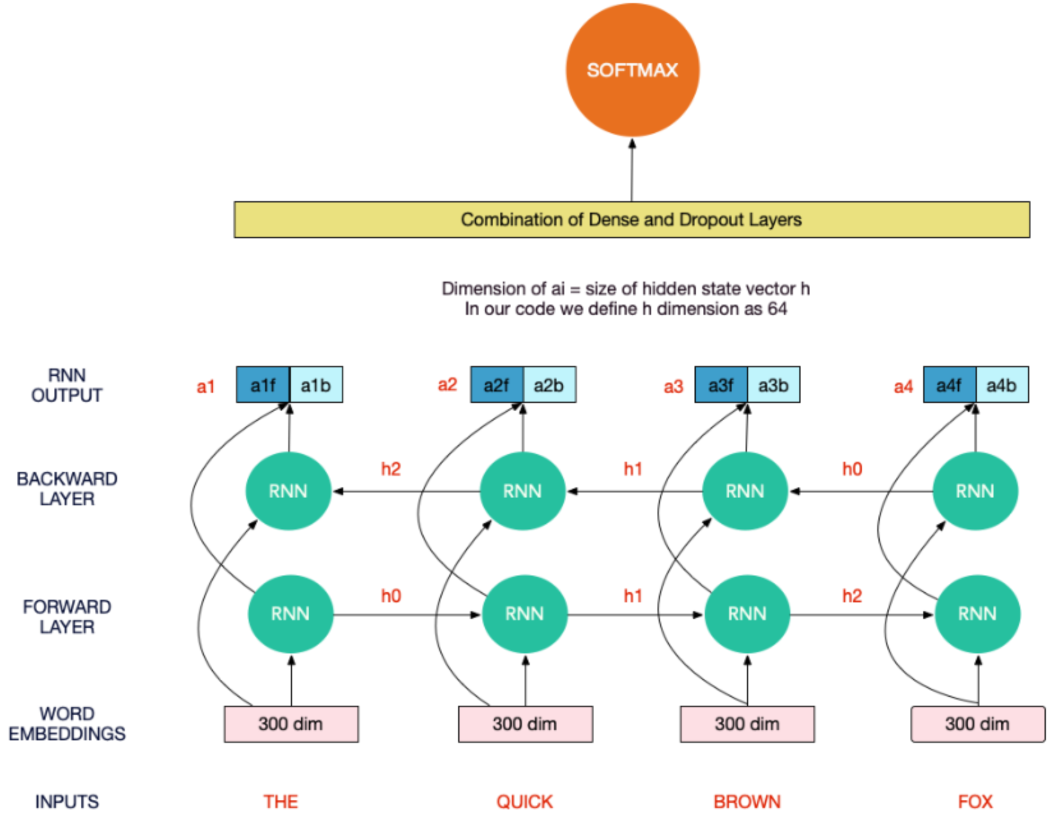


Fig. 3. CNN based intent classifier. The top layer is the same as standard linear classifiers such as logistic regression. The feature vectors in h are summed.



For a most simplistic explanation of Bidirectional RNN, think of RNN cell as taking as input a hidden state(a vector) and the word vector and giving out an output vector and the next hidden state.

Bi-Directional RNN

- Contextual BiLSTM for Classification

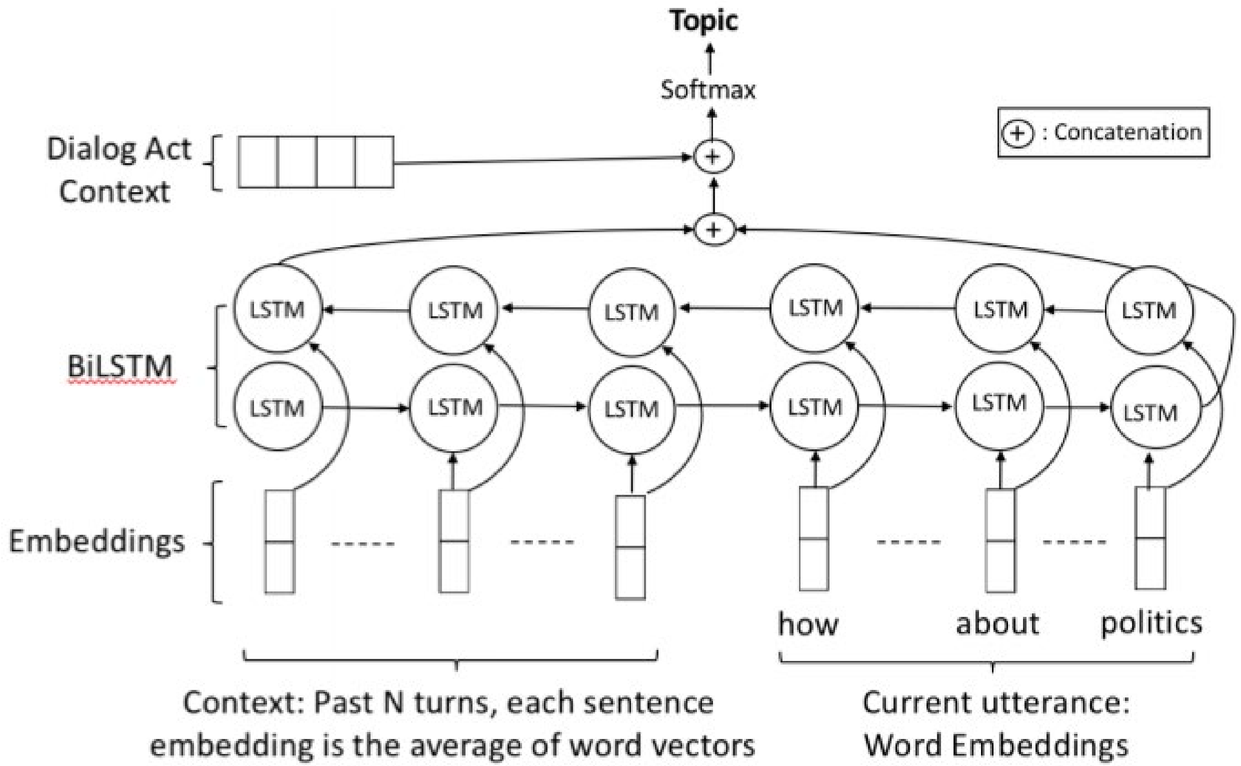
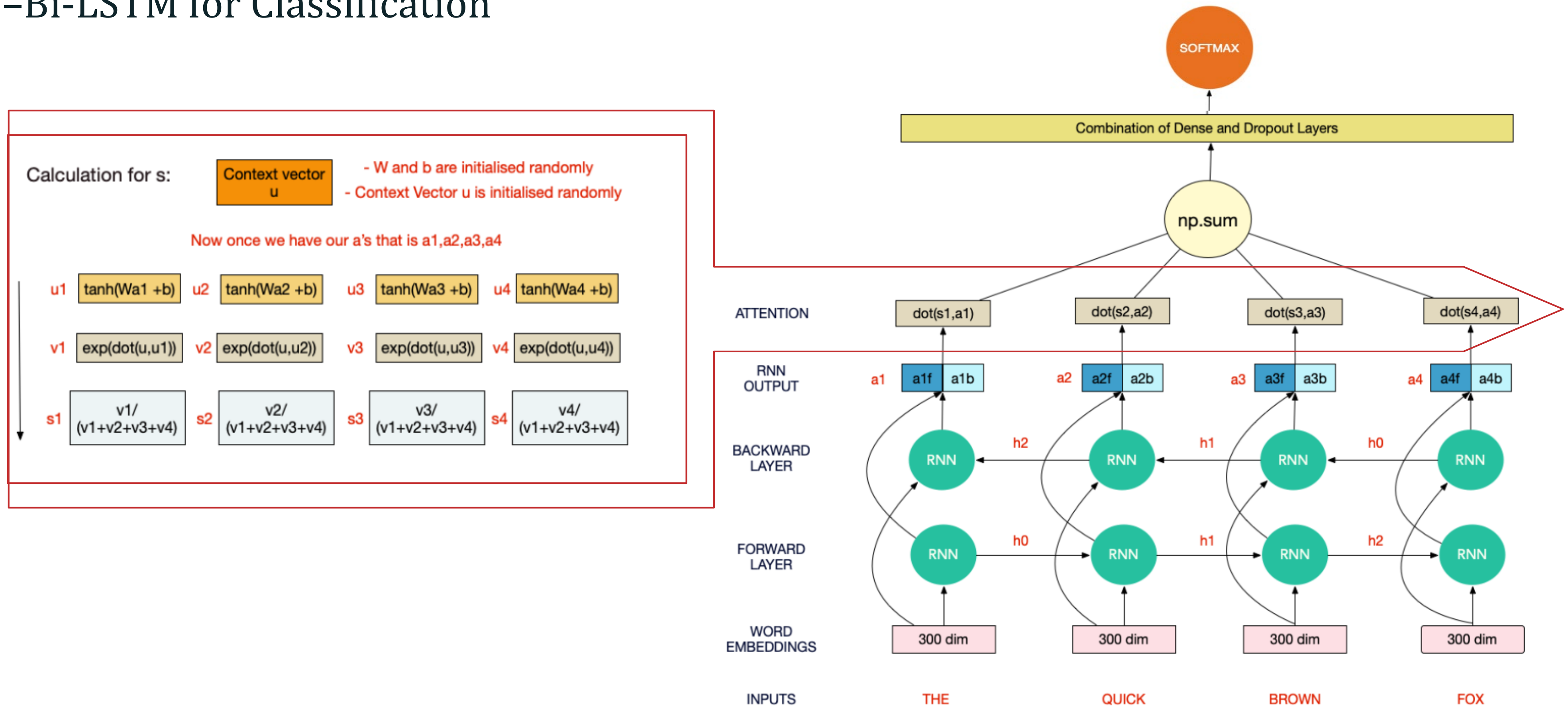


Fig. 5: Contextual BiLSTM for Classification

Need Attention?

- Attention –Bi-LSTM for Classification



NLU Pipeline

- **Slots Filling**

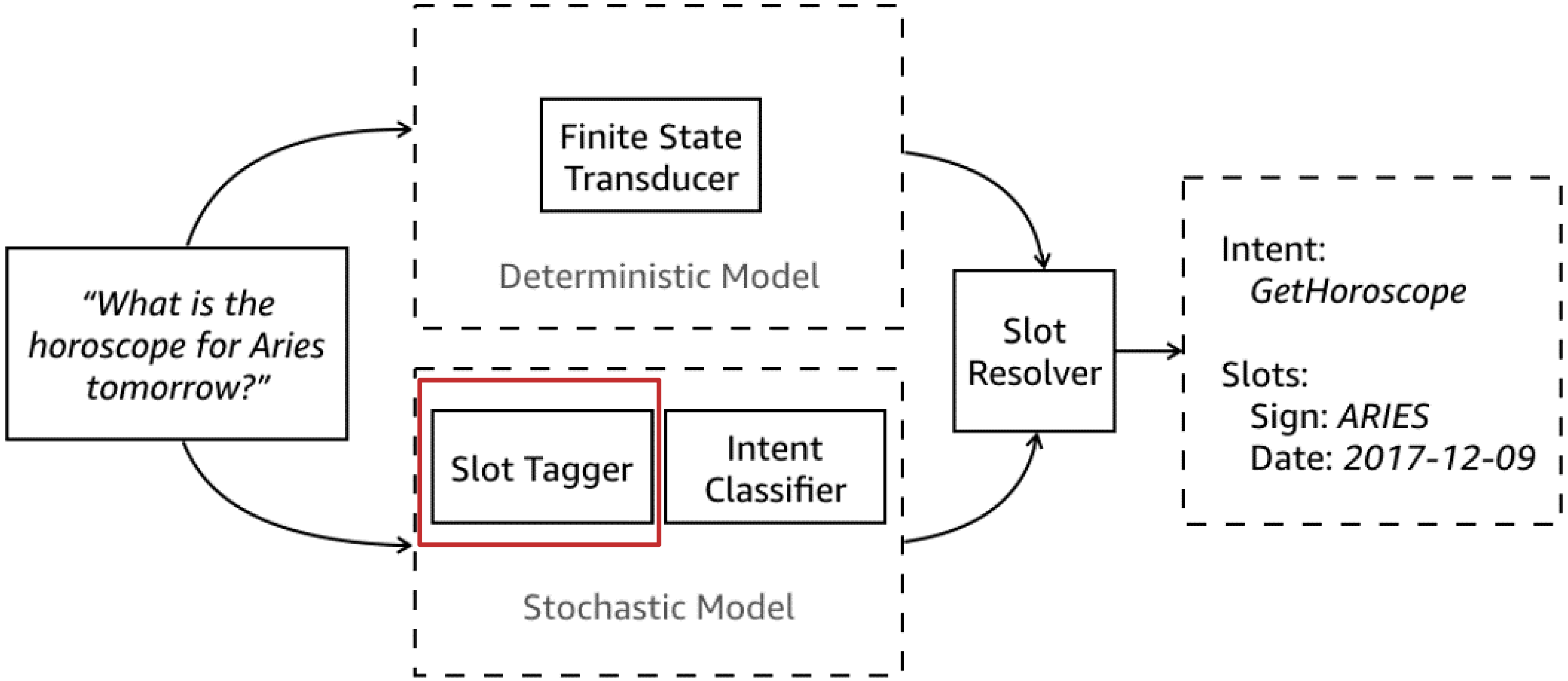
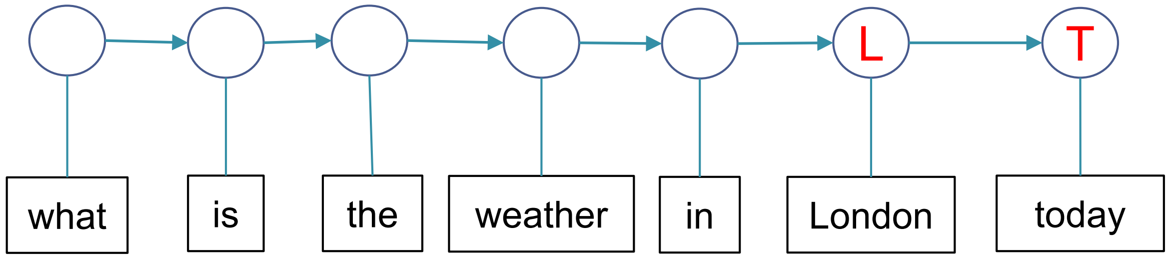


Image from 8. "Just ASK: building an architecture for extensible self-service spoken language understanding."

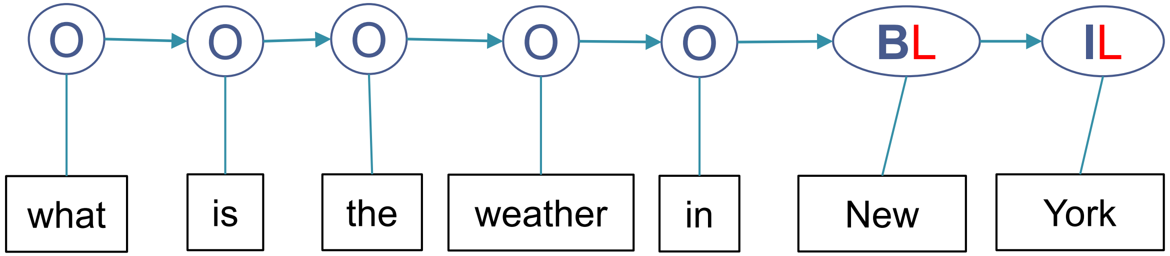
Slots Filling

Slots Filling

- Sequence Labelling Problem
  - Observations in sequence
  - Labels in sequence
    - L for ***Location***
    - T for ***Time***



  - BIO Labels to identify the boundary
    - **B**eginning of Entity/Slot
    - **I**nside of Entity/Slot
    - **O**utside of Entity/Slot



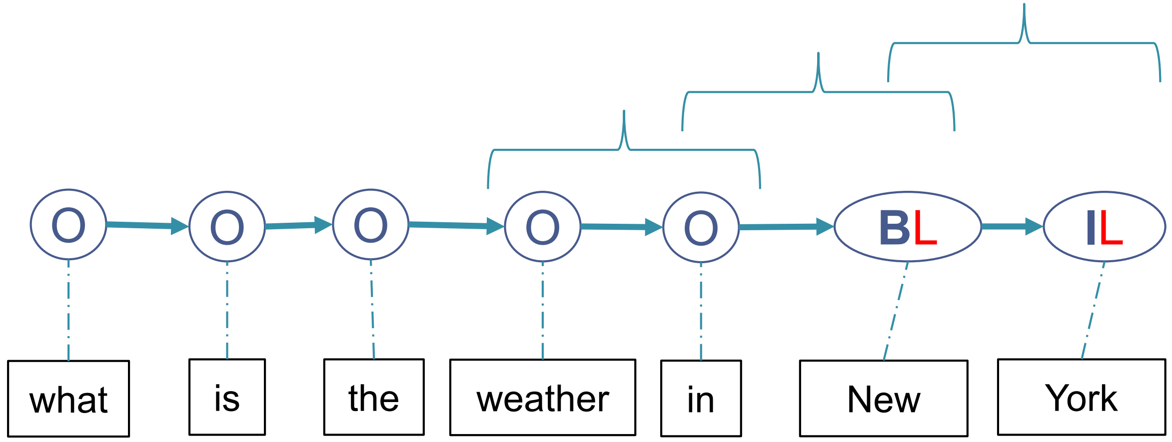
Slots Filling with Pattern Matching

- **Regular Expressions**
  - Date & Time
  - Address
  - Phone Number

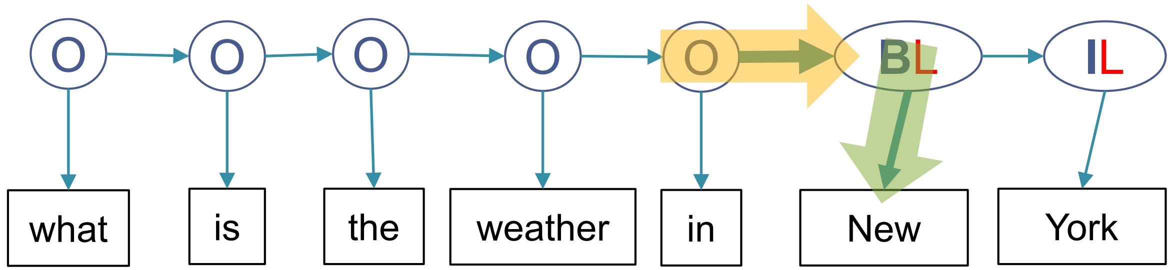
- **Dictionary**
  - Location
  - Names of Person and Organization
  - Domain specific entities

Slots Filling with Machine Learning

- Markov Chain
  - depends only on the **previous** *state/label*
  - States directly observed

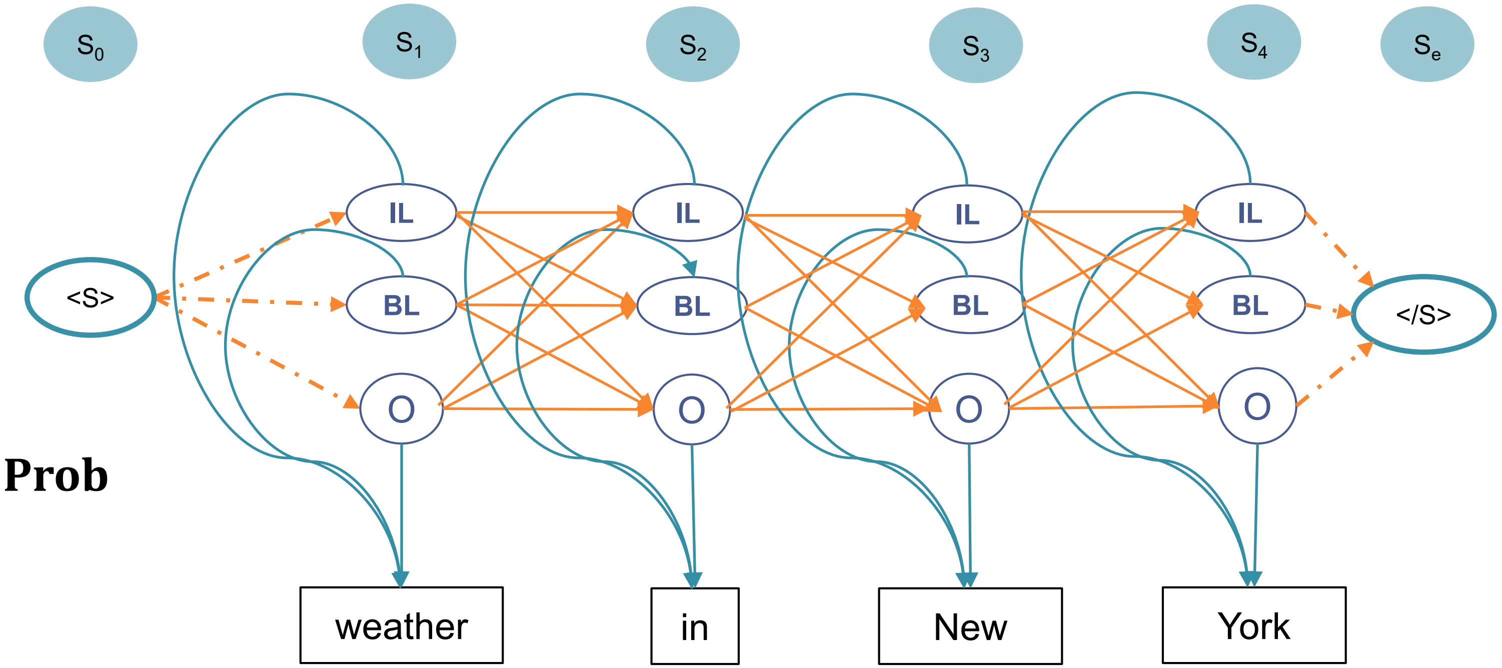


- **H**idden **M**arkov **M**odel
  - depends on the **previous** *state/label and* ***current*** *observation*
  - Considering the observations which are related to the state/label



HMM

- Decoding
  - ***O**bservation* = {*weather,in,new,york*}
  - ***State = {**O,BL,IL**}***
  - Learn likelihood from **training** data
    - *prob(weather|O), prob(BL|O)*
  - Decode Path to maximize the overall **Prob**
    - in total 3*3*3*3 = 81 paths
    - Brute Force



- *P(path) = P(S_1|S_0) * P(O_1|S_1)* P(S_2|S_1) * P(O_2|S_2)* P(S_3|S_2) * P(O_3|S_3)* P(S_4|S_3) * P(O_4|S_4) * P(S_E|S_4)*
- *Path = argmax P(All Paths)*

Viterbi decoding Example

- HMM + Viterbi
  - Training Data

V	DT	N	
read	the	book	
N	PP	DT	N
food	on	the	table
V	PP	V	ADV
try	to	book	now
DT	ADJ	N	V
the	big	table	fell

	book	the	table
V	1/4	0	0
DT	0	3/3	0
N	1/4	0	2/4
PP	0	0	0
ADJ	0	0	0
ADV	0	0	0

What is the ***POS*** tags for
"book the table"

Viterbi decoding Example

- HMM + Viterbi
  - Training Data

V	DT	N	
read	the	book	
N	PP	DT	N
food	on	the	table
V	PP	V	ADV
try	to	book	now
DT	ADJ	N	V
the	big	table	fell

	book	the	table
V	1/4	0	0
DT	0	3/3	0
N	1/4	0	2/4
PP	0	0	0
ADJ	0	0	0
ADV	0	0	0

	V	DT	N	PP	ADJ	ADV	</S>
<S>	2	1	1	0	0	0	0
V	0	1	0	1	0	1	1
DT	0	0	2	0	1	0	0
N	1	0	0	1	0	0	3
PP	1	1	0	0	0	0	0
ADJ	0	0	1	0	0	0	0
ADV	0	0	0	0	0	0	1

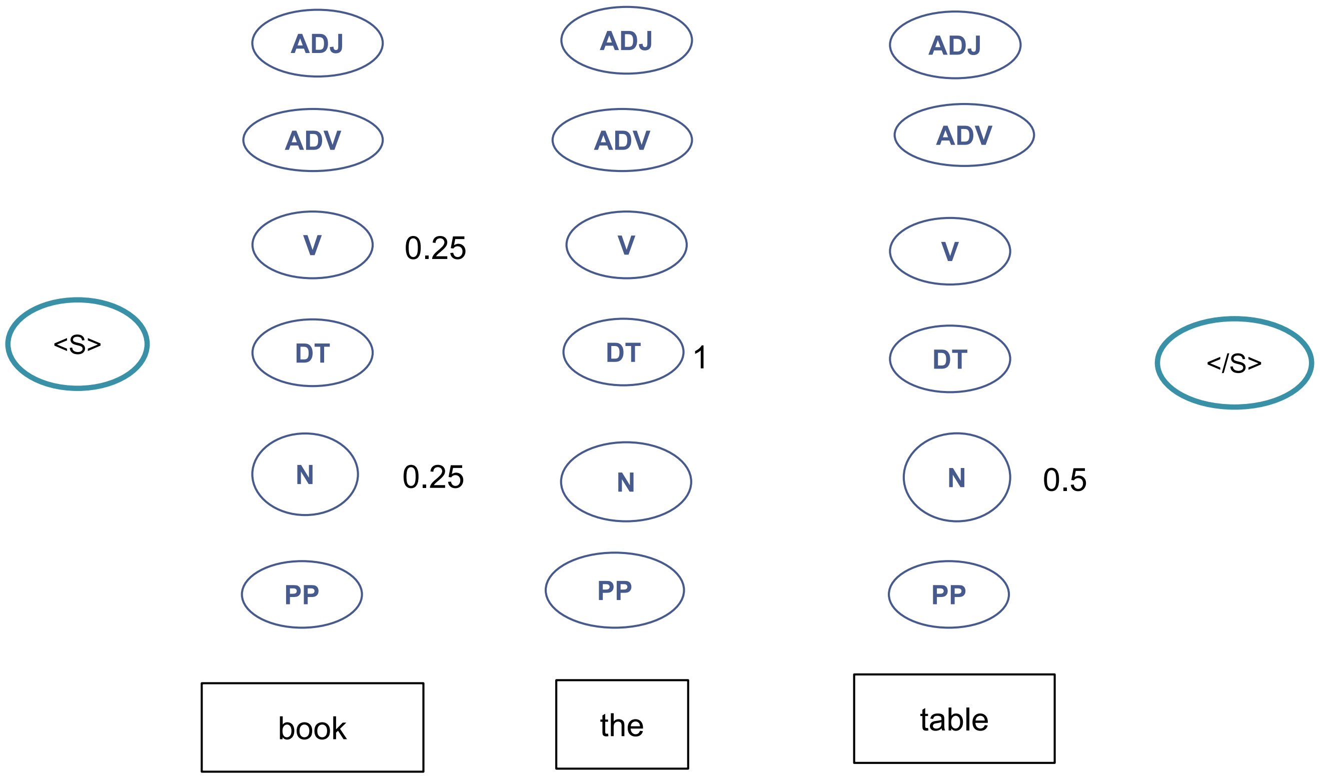
  - What is the ***POS*** tags for
    - “book the table”

Viterbi decoding Example

- Viterbi
  - Likelihood of observations
  - Smoothing

	book	the	table
V	1/4	0.01	0.01
DT	0.01	3/3	0.01
N	1/4	0.01	2/4
PP	0.01	0.01	0.01
ADJ	0.01	0.01	0.01
ADV	0.01	0.01	0.01

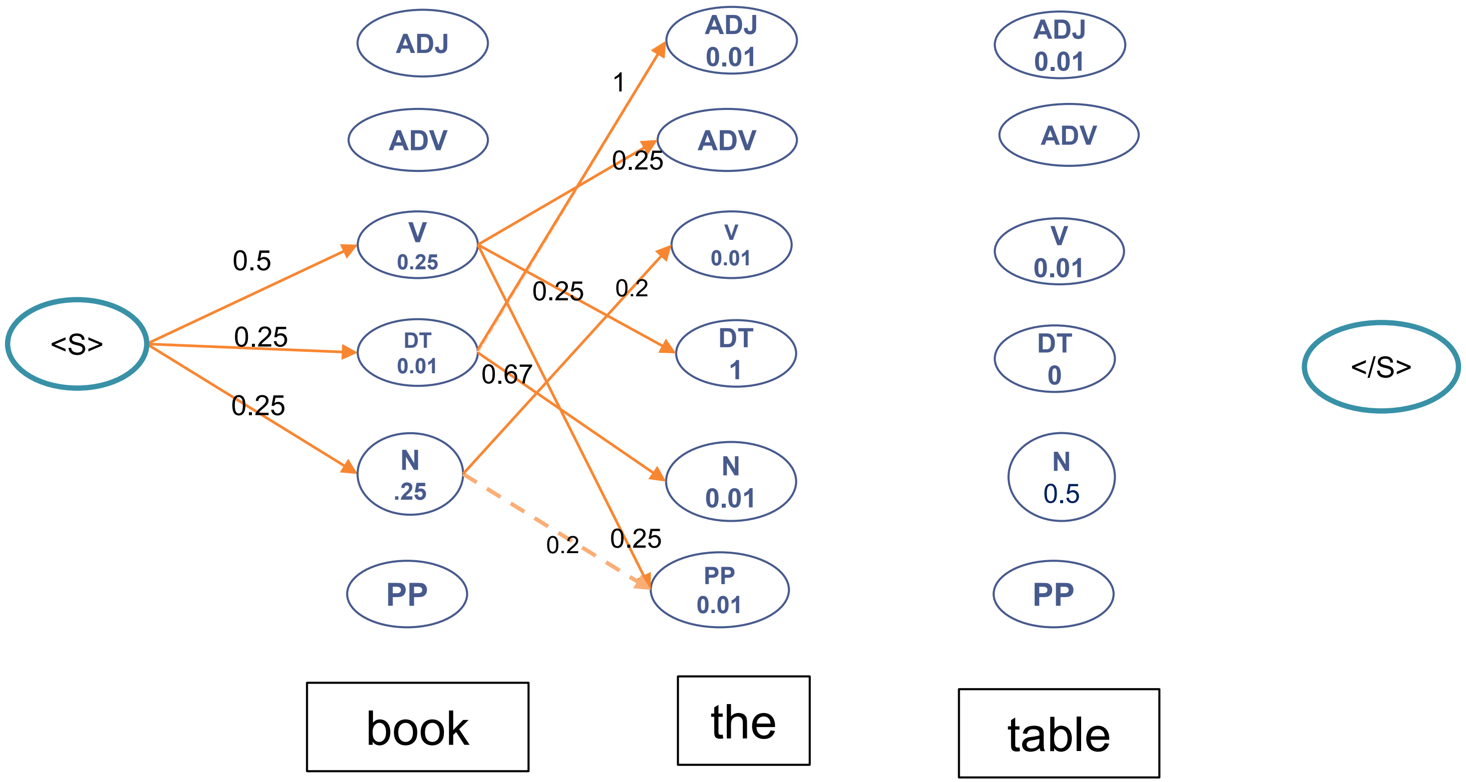
  - in total 6*6*6 = 216 paths



Viterbi decoding Example

- Viterbi
  - To reach all possible destinations at $\mathbf{S_{n+1}}$
  - Keep the best starting point at $\mathbf{S_n}$

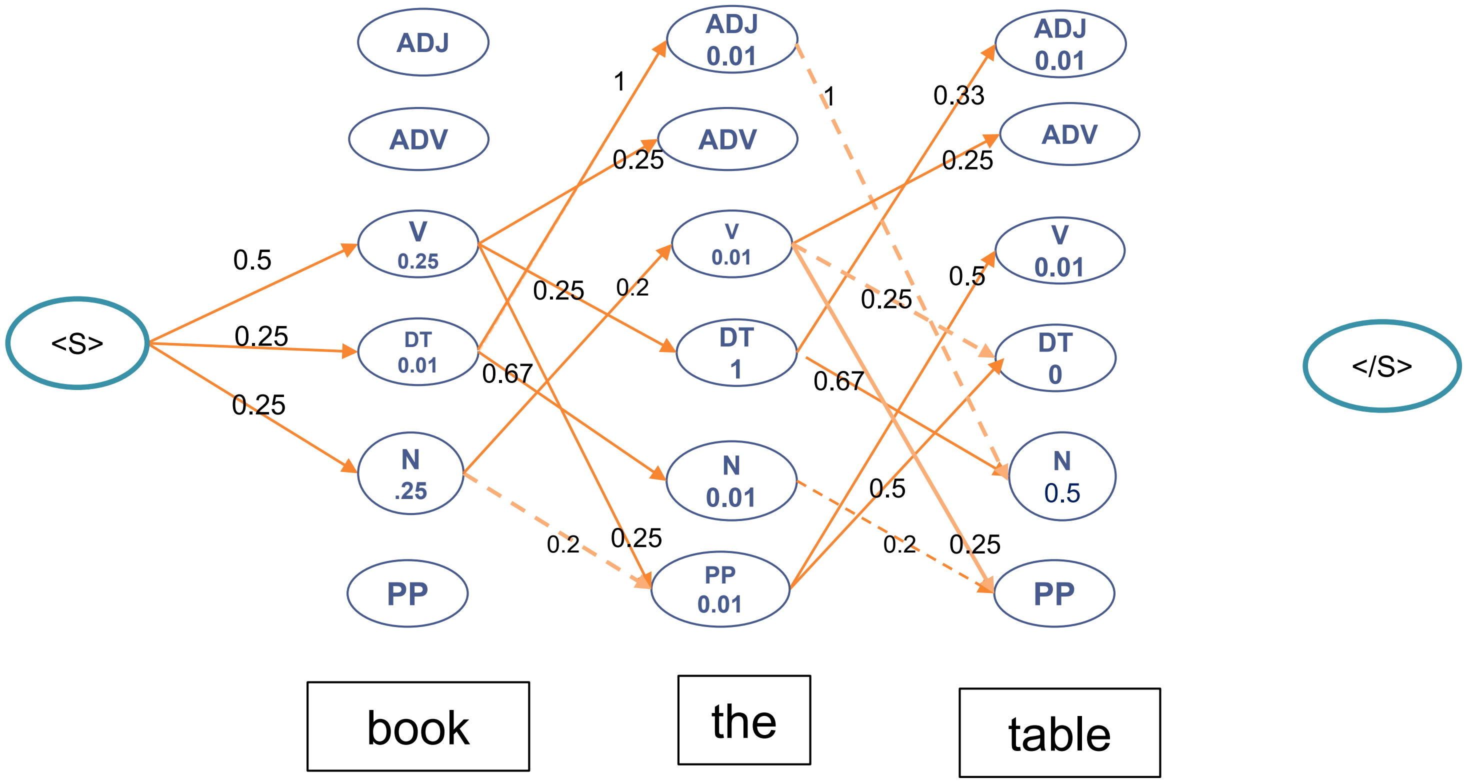
	V	DT	N	PP	ADJ	ADV	</S>
<S>	2	1	1	0	0	0	0
V	0	1	0	1	0	1	1
DT	0	0	2	0	1	0	0
N	1	0	0	1	0	0	3
PP	1	1	0	0	0	0	0
ADJ	0	0	1	0	0	0	0
ADV	0	0	0	0	0	0	1



Viterbi decoding Example

- Viterbi
  - To reach all possible destinations at $\mathbf{S_{n+1}}$
  - Keep the best starting point at $\mathbf{S_n}$

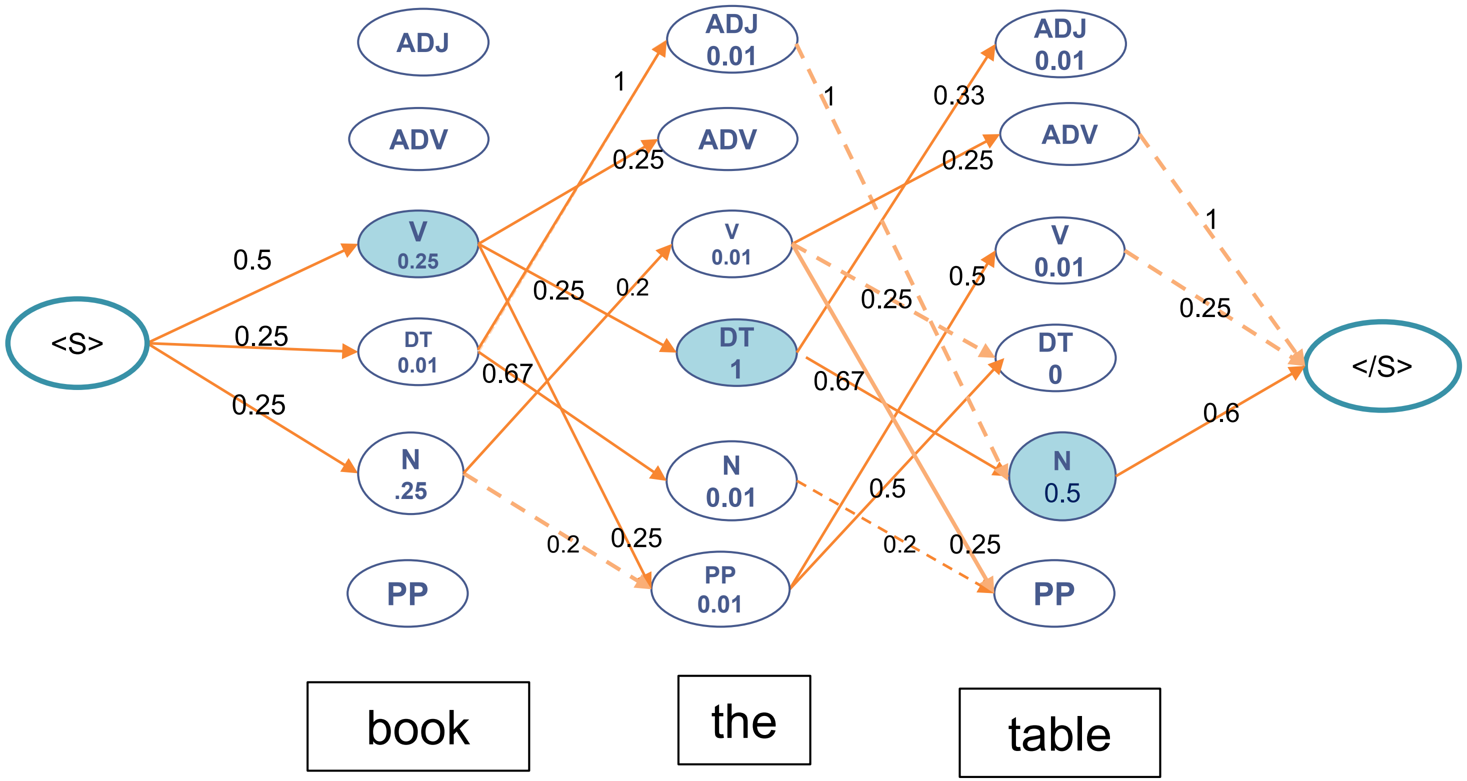
	V	DT	N	PP	ADJ	ADV	</S>
<S>	2	1	1	0	0	0	0
V	0	1	0	1	0	1	1
DT	0	0	2	0	1	0	0
N	1	0	0	1	0	0	3
PP	1	1	0	0	0	0	0
ADJ	0	0	1	0	0	0	0
ADV	0	0	0	0	0	0	1



Viterbi decoding Example

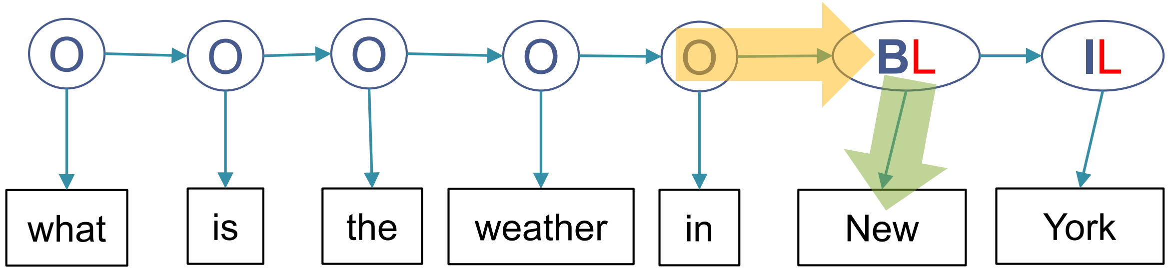
- Viterbi
  - To reach all possible destinations at $\mathbf{S_{n+1}}$
  - Keep the best starting point at $\mathbf{S_n}$
  - Trace back to get the best path

	V	DT	N	PP	ADJ	ADV	</S>
<S>	2	1	1	0	0	0	0
V	0	1	0	1	0	1	1
DT	0	0	2	0	1	0	0
N	1	0	0	1	0	0	3
PP	1	1	0	0	0	0	0
ADJ	0	0	1	0	0	0	0
ADV	0	0	0	0	0	0	1

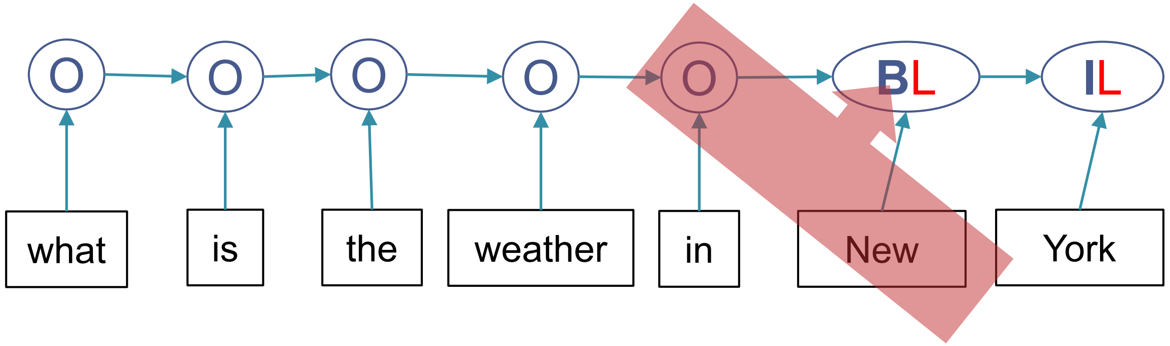


Maximum Entropy Markov Model

- HMM defines the next state under the current state conditions given



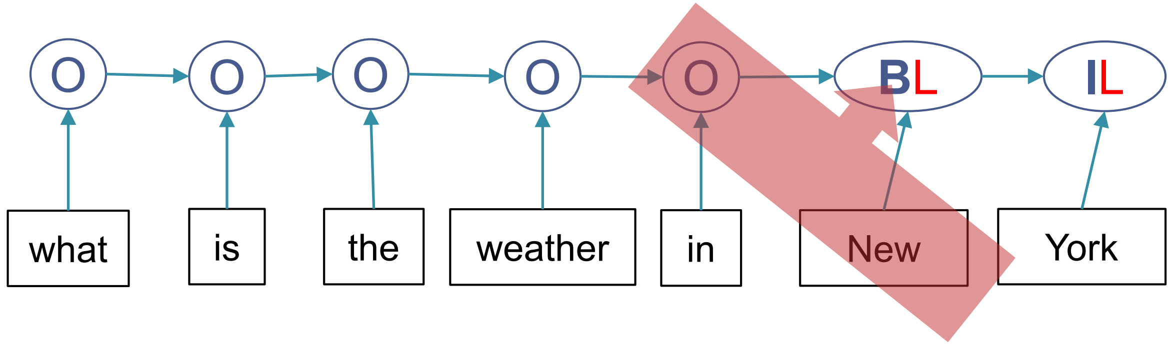
- MEMM gives one probability estimate per hidden state, which is the probability of the next tag given the previous tag and the observation.

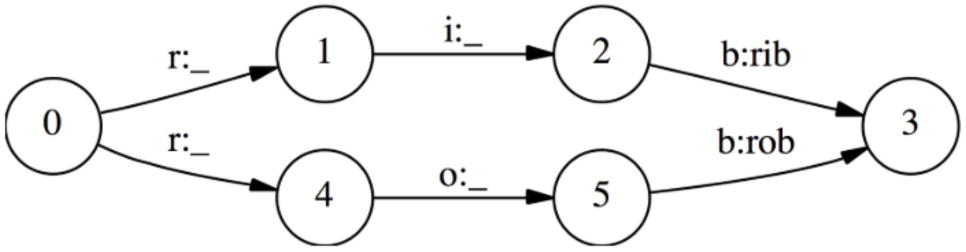


$$P(s \mid s', o) = \frac{1}{Z(o, s')} \exp\left(\sum_{i=1}^{N} w_i \cdot f_i(o, s')\right)$$

MEMM

- MEMM
  - computes the probability of the next state, given the current state and the observation
  - Transitions from a given state are competing against each other only
  - Per state normalization, i.e. sum of transition probability for any state has to sum to 1



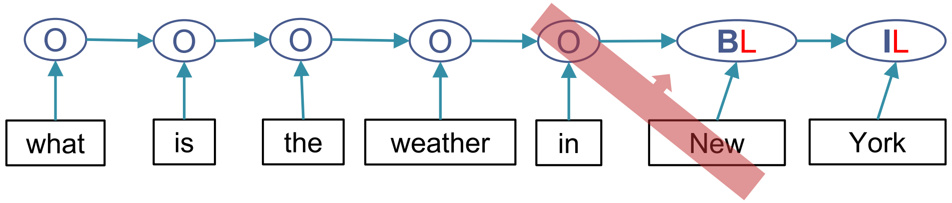


Label Bias Problem.
(taken from Lafferty et al. 2001)

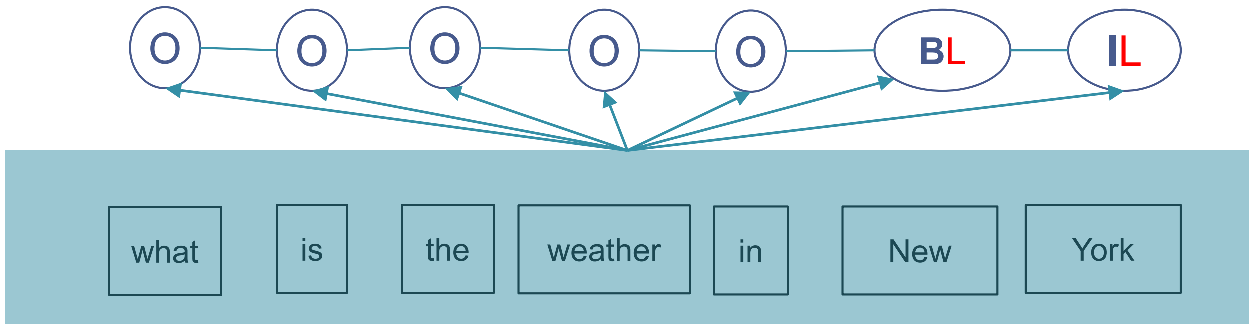
"for state (1 and 4) with single transitions, word's sequence with higher frequency will always win."

Conditional Random Field

- MEMM computes the probability of the next state, given the current state and the observation



- CRF computes the joint probability distribution of the entire label sequence



Length of sequence x

Sum over all feature function

Weight for given feature function

Feature Functions

$$P(\overline{y}|\overline{x};w) = \frac{exp(\sum_i \sum_j w_j f_j(y_{i-1}, y_i, \overline{x}, i))}{\sum_{y' \in Y} exp(\sum_i \sum_j w_j f_j(y'_{i-1}, y'_i, \overline{x}, i))}$$

Sum over all possible label sequence

Feature function can access all of observation

Linear-Chain Conditional Random Field.
(taken from Sameer Maskey slides)

CRF Important Observations

- MEMMs are normalized locally over each observation, and hence suffer from the Label Bias problem

- CRFs avoid the label bias problem as they are globally re-normalized.

- The inference algorithm is again based on Viterbi algorithm.

- Output transition dependent on the state and the observation are not modelled separately but as one conditional probability

Features for Entity Recognition

- Stanford NER by CRF

Our Features

- Word features: current word, previous word, next word, all words within a window
- Orthographic features:
  - Jenny → Xxxx
  - IL-2 → XX-#
- Prefixes and Suffixes:
  - Jenny → <J, <Je, <Jen, …, nny>, ny>, y>
- Label sequences
- Lots of feature conjunctions

Distributional Similarity Features

- Large, unannotated corpus
- Each word will appear in contexts - induce a distribution over contexts
- Cluster words based on how similar their distributions are
- Use cluster IDs as features
- Great way to combat sparsity
- We used Alexander Clark's distributional similarity code (easy to use, works great!)
- 200 clusters, used 100 million words from English gigaword corpus

CRF with Neural Models

- Bi-LSTM + CRF with Features + Word Embedding

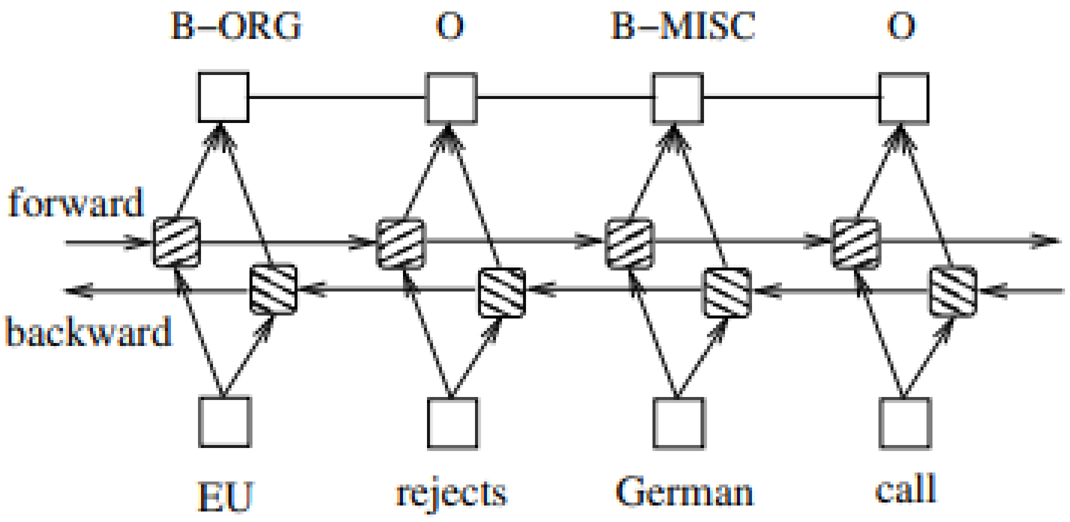


Figure 7: A BI-LSTM-CRF model.

Algorithm 1 Bidirectional LSTM CRF model training procedure

```
for each epoch do
    for each batch do
        1) bidirectional LSTM-CRF model forward pass:
            forward pass for forward state LSTM
            forward pass for backward state LSTM
        2) CRF layer forward and backward pass
        3) bidirectional LSTM-CRF model backward pass:

            backward pass for forward state LSTM
            backward pass for backward state LSTM
        4) update parameters
    end for
end for
```

CRF with Neural Models

- Bi-LSTM + CRF + CNN with Word and Character Embedding
  - truly end-to-end
  - requiring no feature engineering

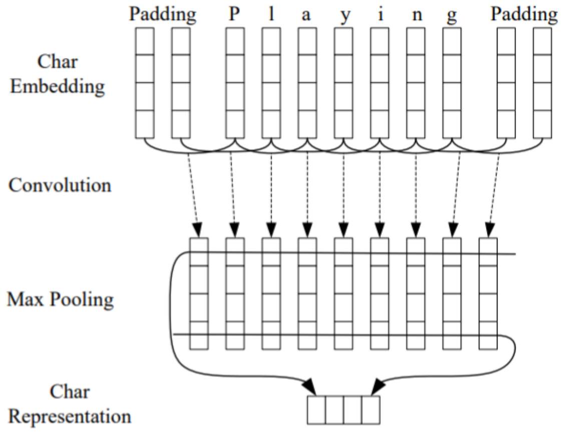
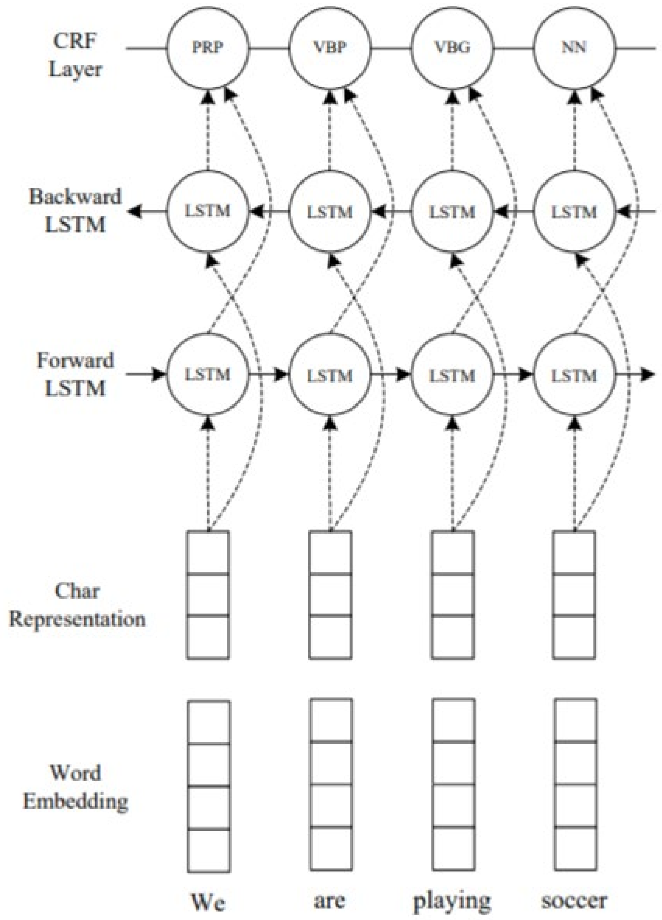


Figure 1: The convolution neural network for extracting character-level representations of words. Dashed arrows indicate a dropout layer applied before character embeddings are input to CNN.



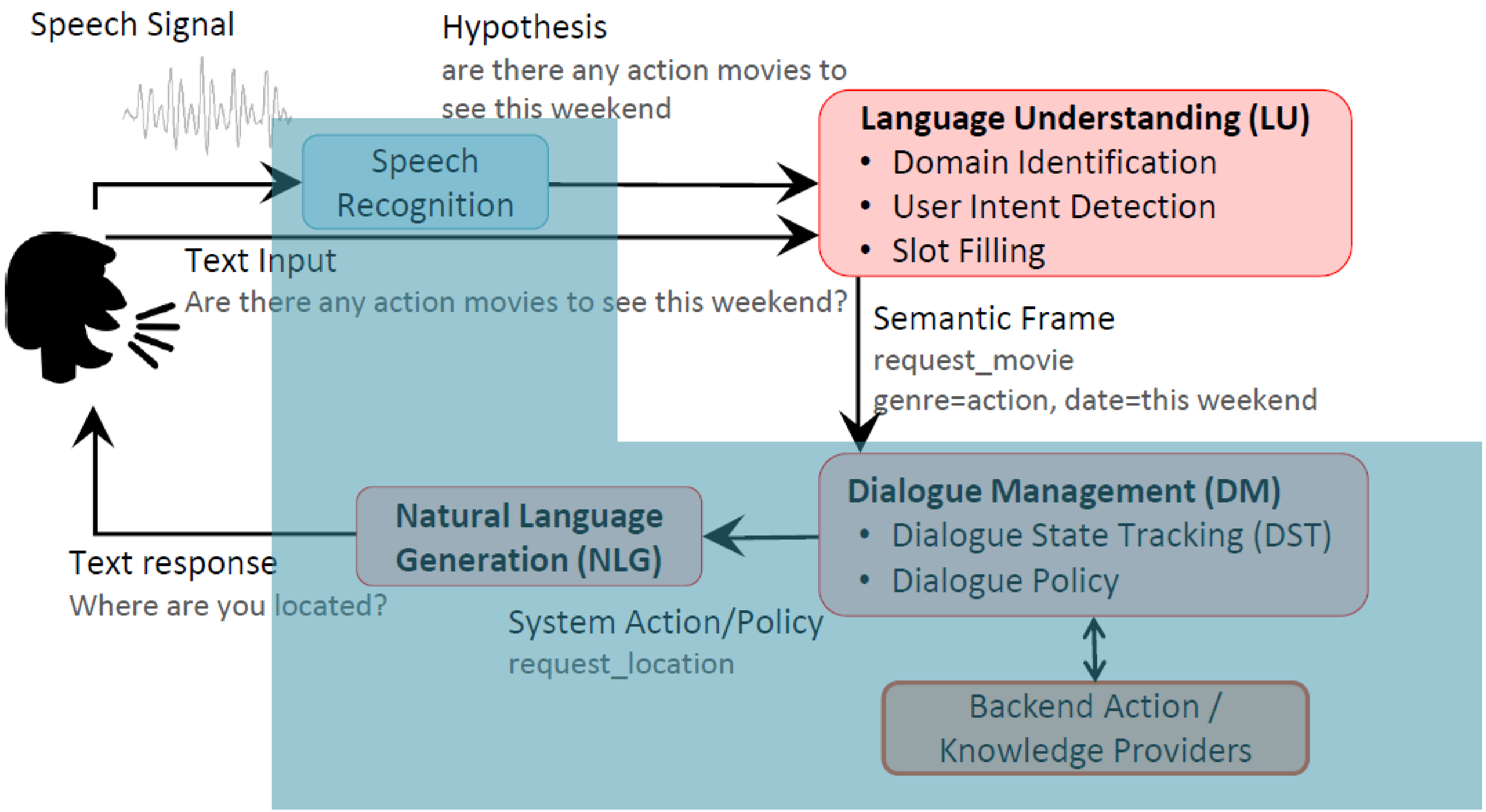
CRF with Neural Models

- Bi-LSTM + CRF + CNN with Word and Character Embedding
  - truly end-to-end
  - requiring no feature engineering
  - significant improvement

Model	F1
Chieu and Ng (2002)	88.31
Florian et al. (2003)	88.76
Ando and Zhang (2005)	89.31
Collobert et al. (2011)‡	89.59
Huang et al. (2015)‡	90.10
Chiu and Nichols (2015)‡	90.77
Ratinov and Roth (2009)	90.80
Lin and Wu (2009)	90.90
Passos et al. (2014)	90.90
Lample et al. (2016)‡	90.94
Luo et al. (2015)	91.20
This paper	**91.21**

Table 5: NER F1 score of our model on test data set from CoNLL-2003. For the purpose of comparison, we also list F1 scores of previous top-performance systems. ‡ marks the neural models.

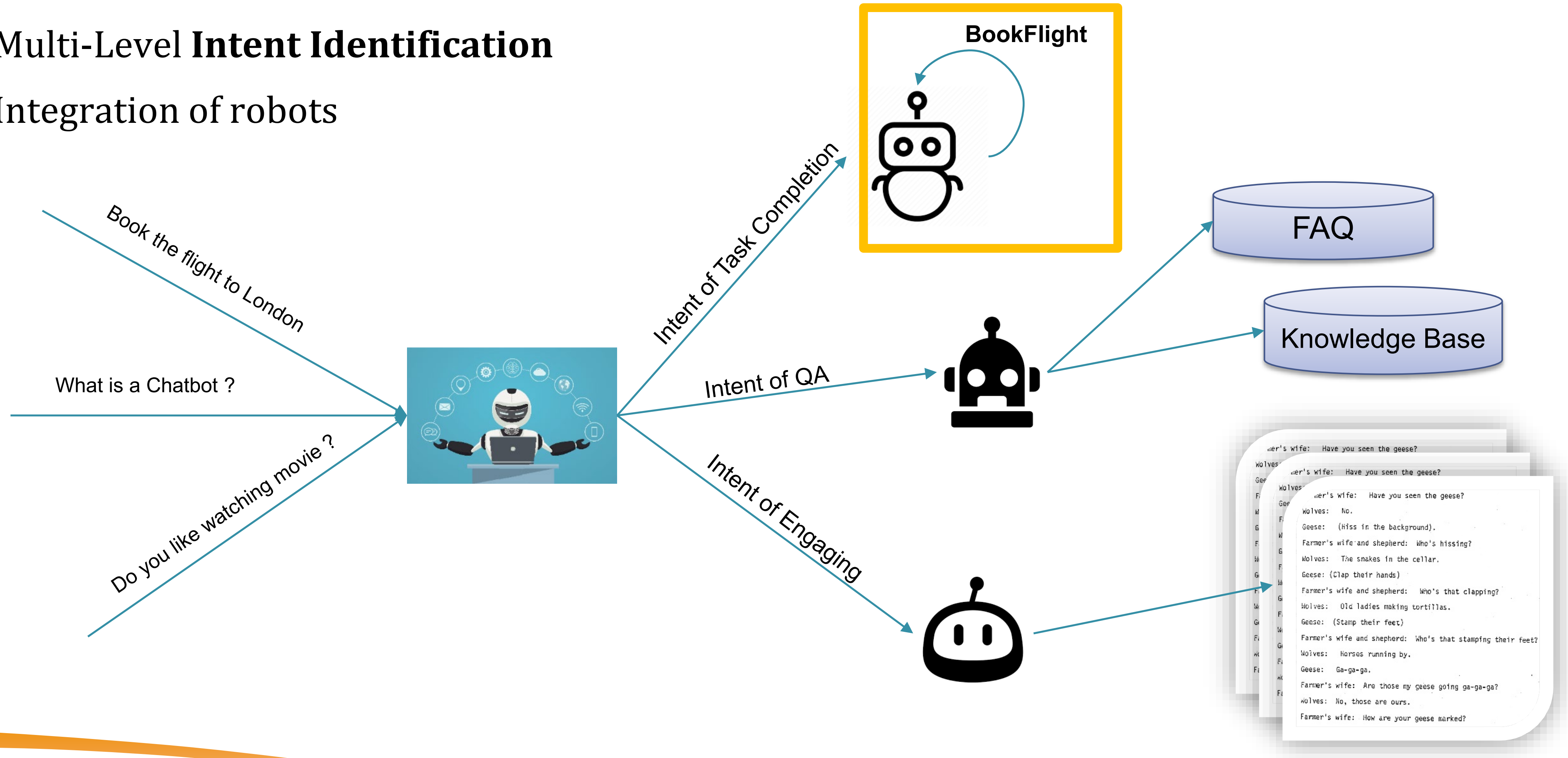
Core Architecture



https://sites.google.com/view/deepdial/

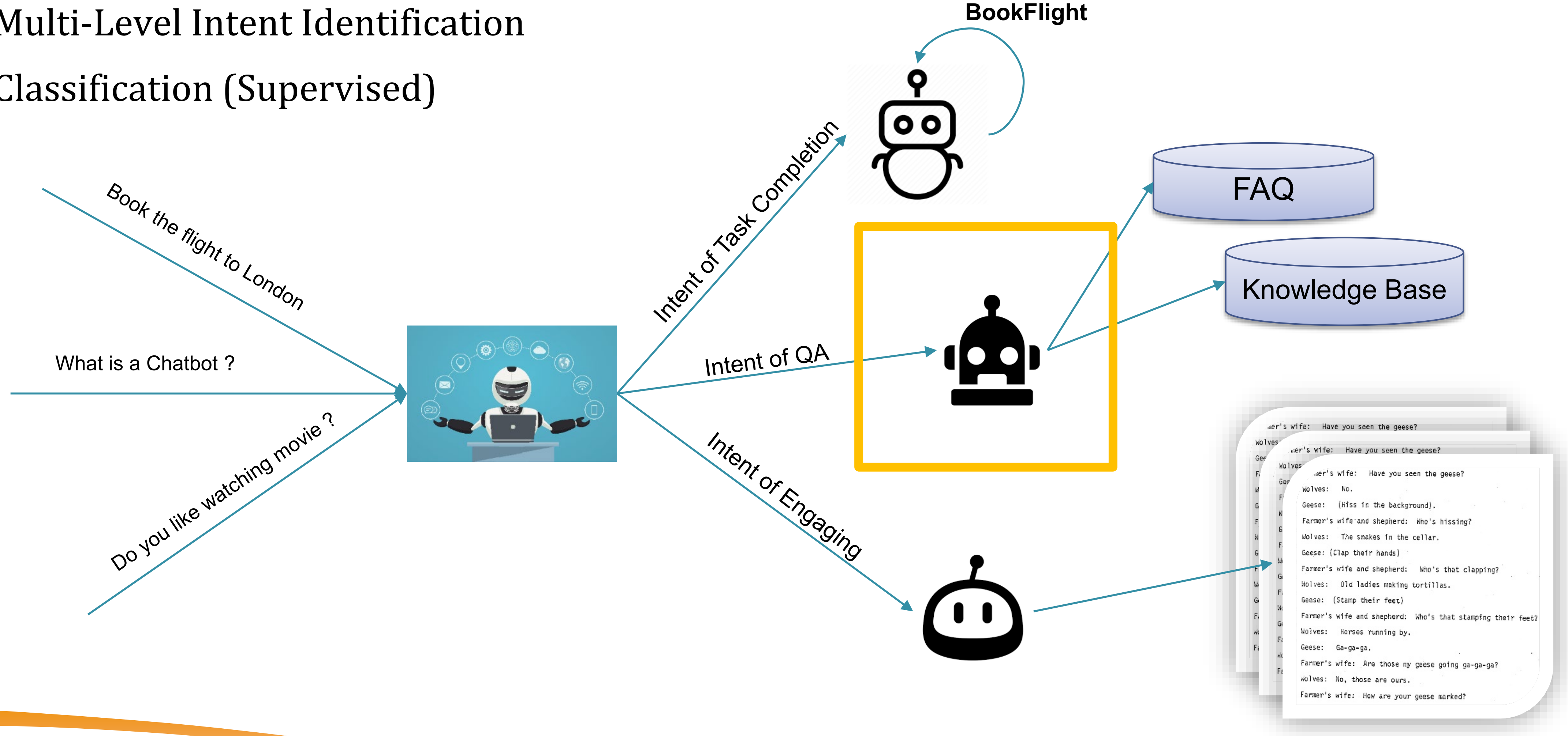
Language Understanding

- Multi-Level **Intent Identification**
- Integration of robots



Language Understanding for QA

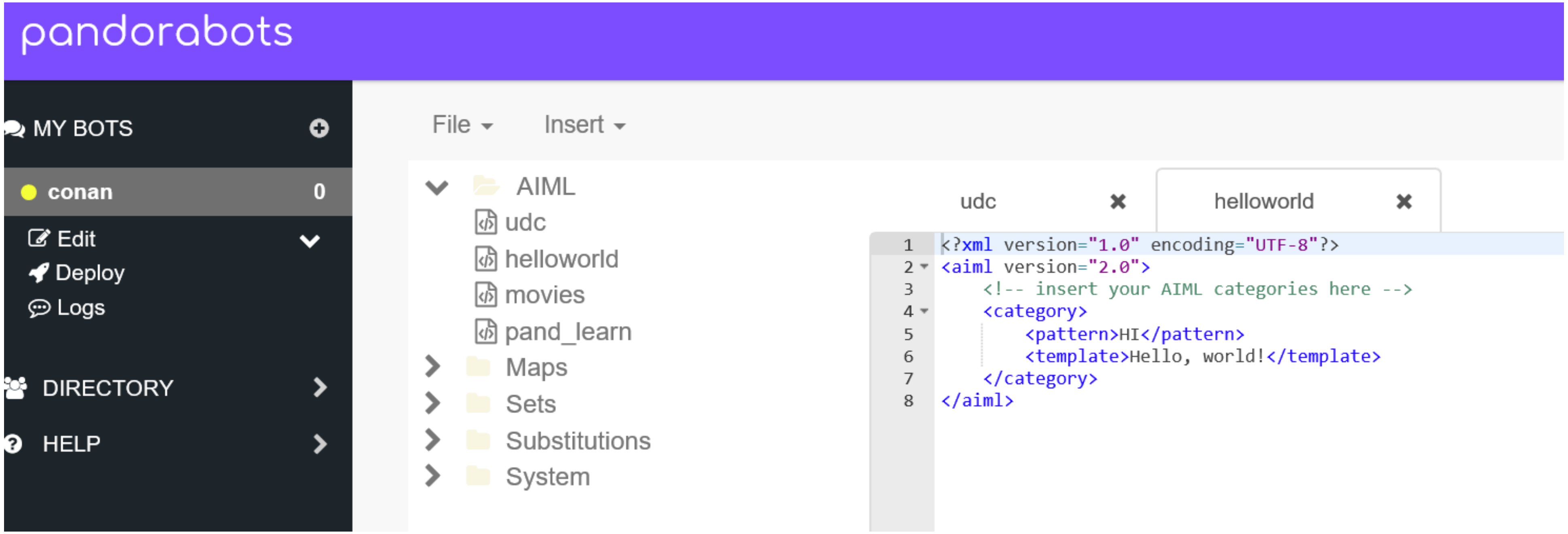
- Multi-Level Intent Identification
- Classification (Supervised)



Language Understanding for QA

Template based Approach

- *Pandorabots Platform*
  - Built with **AIML** Engine
  - Useful for extremely narrow domain
  - Creating a bot in Pandorabots Playground



https://home.pandorabots.com

Pandorabots

- An Example of FAQ Chatbots

What is type 2 diabetes?
What are the main symptoms of type 2 diabetes?
What is the main cause of type 2 diabetes?
How do you treat type 2 diabetes?
Can type 2 diabetes be cured?
Does exercise help?
Tell me about blood sugar.
What are the long term complications?

```
<category>

<pattern>what is Type 2 diabetes</pattern>

<template>type 2 diabetes is a metabolic disorder that is
characterized by high blood glucose in the context of insulin
resistance and relative insulin deficiency</template>

</category>
```

Code 7.2 AIML category for "What is type 2 diabetes?"

```
<category>

<pattern>what are the causes </pattern>

<template>

   <srai>what is the main cause of type 2 diabetes</srai>

</template>

</category>
```

Code 7.3 AIML category for "What is type 2 diabetes?" including an <srai> tag

Pandorabots

- An Example of FAQ Chatbots
  - Wildcards
  - Regular expressions

```
<pattern>what are the main symptoms ^ </pattern>
```

```
<category>
<pattern>I feel * </pattern>
<template>When do you feel <star/></template>
</category>
```

Code 7.5 AIML code using the * wildcard and the <star/> tag

```
<category>
  <pattern>I feel * and I * </pattern>
  <template>So you feel <star/> and you
     <star index = "2" /> </template>
</category>
```

Code 7.6 AIML code using the <star/> tag at several positions

Pandorabots

- An Example of FAQ Chatbots
  - Set and get variables

```
<category>

  <pattern>my main medication is * </pattern>

  <template>OK, I have noted that your main medication is

    <set name = "medication"<star/></set> </template>

</category>
```

Code 7.7 AIML code that defines a predicate called "medication"

```
<template>you said that your main medication is <get name =
"medication"/></template>
```

Code 7.8 AIML code that uses the value of a predicate

Pandorabots

- An Example of FAQ Chatbots
  - Sets and Maps

What is the capital of Alabama?
What is the capital of Arizona?
What is the capital of California?

Table 7.1 Sets in AIML

State set	State2capital set
Alabama	Alabama:Montgomery
Arizona	Arizona:Phoenix
California	California:Sacramento
...	...

```
<category>
 <pattern>What is the capital of <set>state</set></pattern>
 <template><map name="state2capital"><star/></map></template>
</category>
```

Code 7.10 An AIML map

Pandorabots

- *Pandorabots Platform*
  - Context
  - Topics

```
<?xml version = "1.0" encoding = "UTF-8"?>
<aiml version = "1.0.1" encoding = "UTF-8"?>
   <category>
      <pattern>LET DISCUSS MOVIES</pattern>
      <template>Yes <set name = "topic">movies</set></template>
   </category>

   <topic name = "movies">
      <category>
         <pattern> * </pattern>
         <template>Watching good movie refreshes our minds.</template>
      </category>

      <category>
         <pattern> I LIKE WATCHING COMEDY! </pattern>
         <template>I like comedy movies too.</template>
      </category>

   </topic>
</aiml>
```

Template based Approach

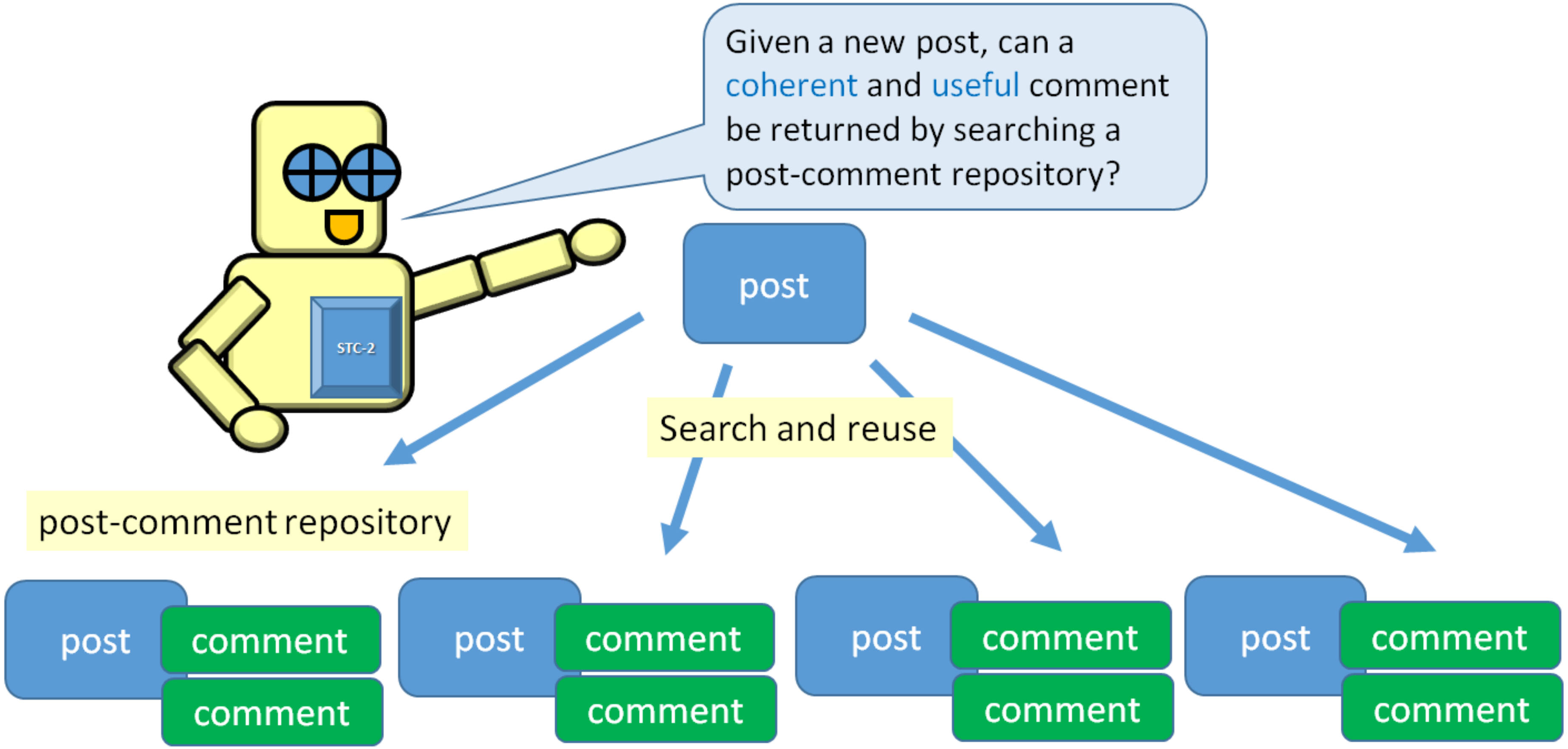
- *Pandorabots Platform*
  - Built with **AIML** Engine
  - Useful for extremely narrow domain
  - Built in Categories
  - Automatic knowledge acquisition

Pandorabots + Coca-Cola: The Future of Next Generation Vending

At VentureBeat's MobileBeat 2017, Pandorabots and Coca-Cola announced the Next Generation Vending project, one of several collaborations...

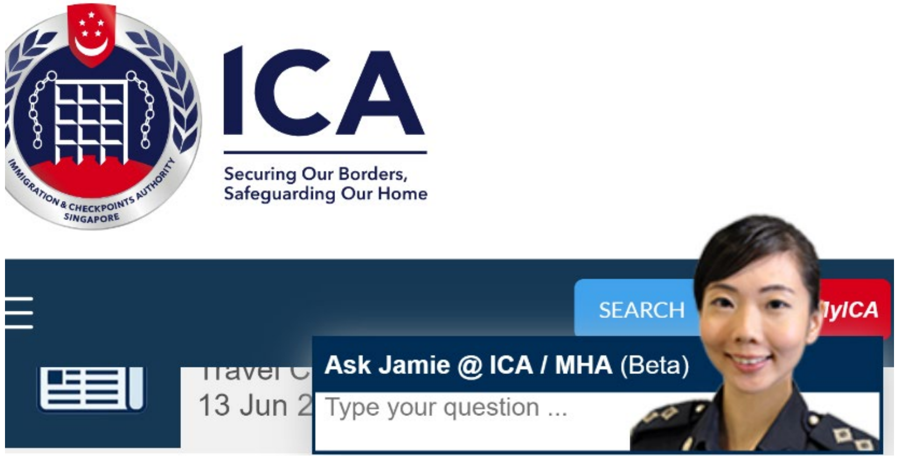
Retrieval-Based Approach

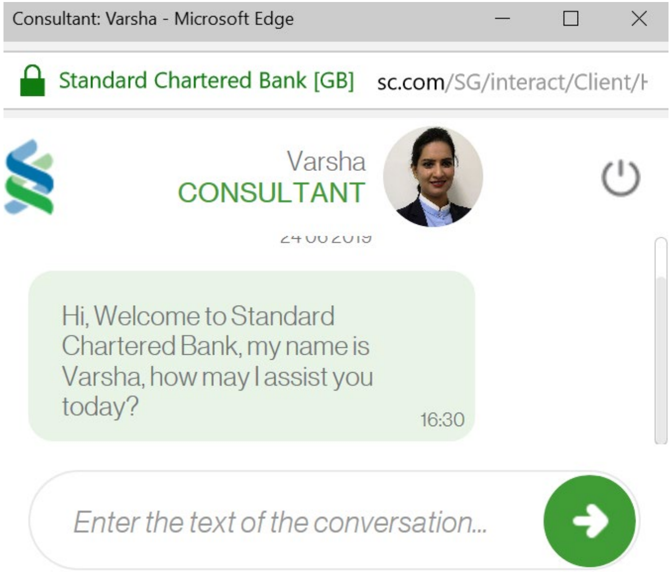
retrieval-based method



Retrieval-Based Approach

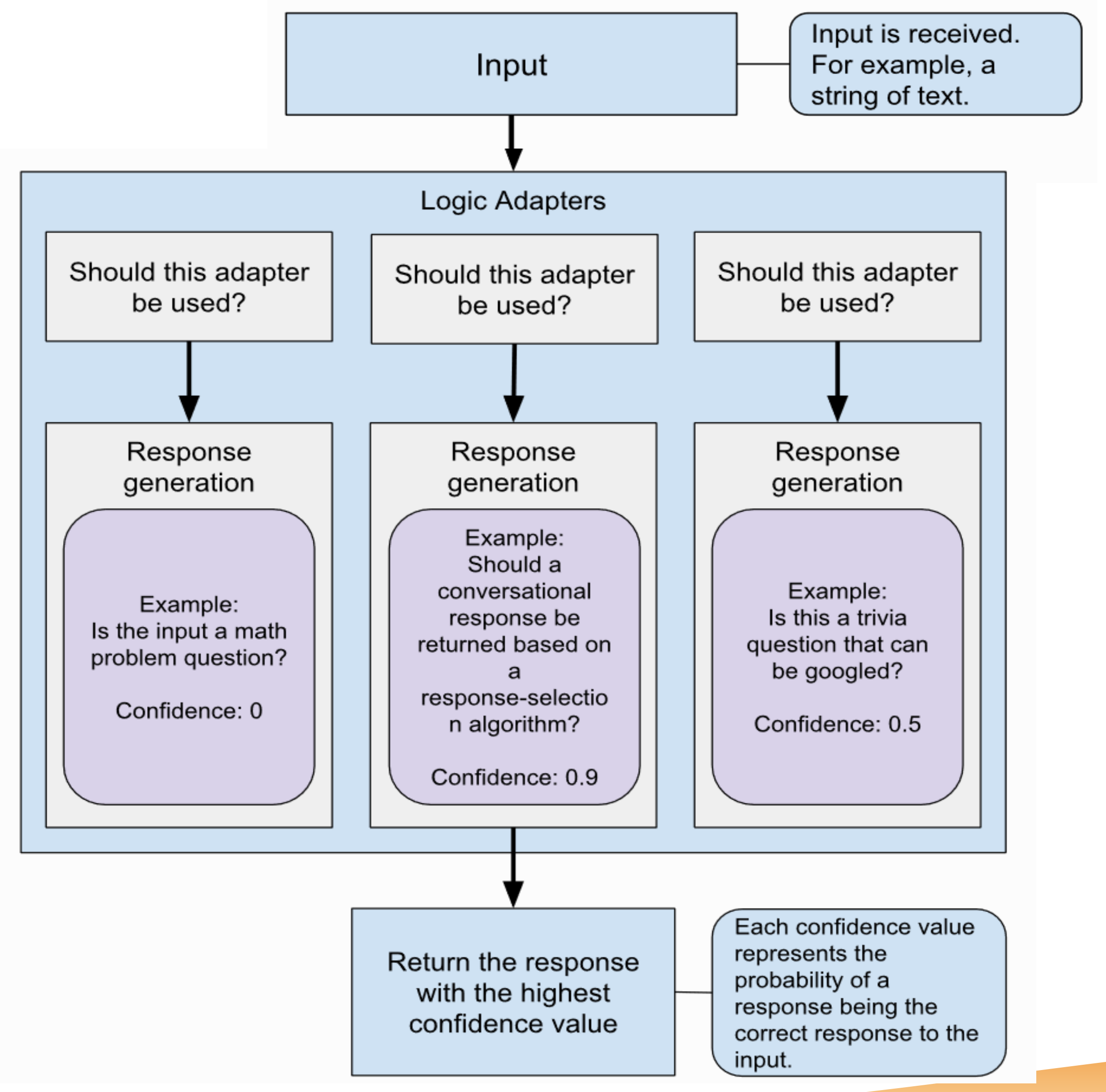
- Chatterbot
  - Python library for building Chatbots
  - Conversation data collection and Python Programing skills are needed
  - Easy to build and iterate
  - Widely used in FAQ and Website Chatbots





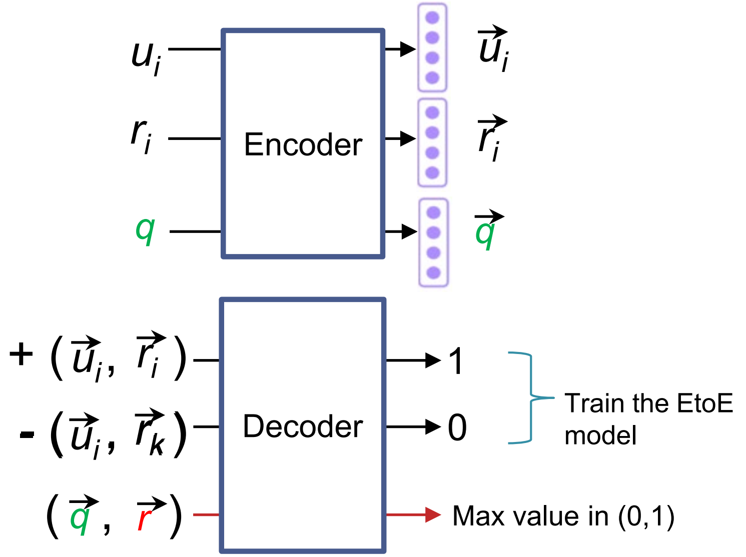
ChatterBot

- An Example of FAQ Chatbots
  - Logic Adapters
  - Best Match Adapter
    - **Bigram search**
  - Time Logic Adapter
  - Mathematical Evaluation Adapter

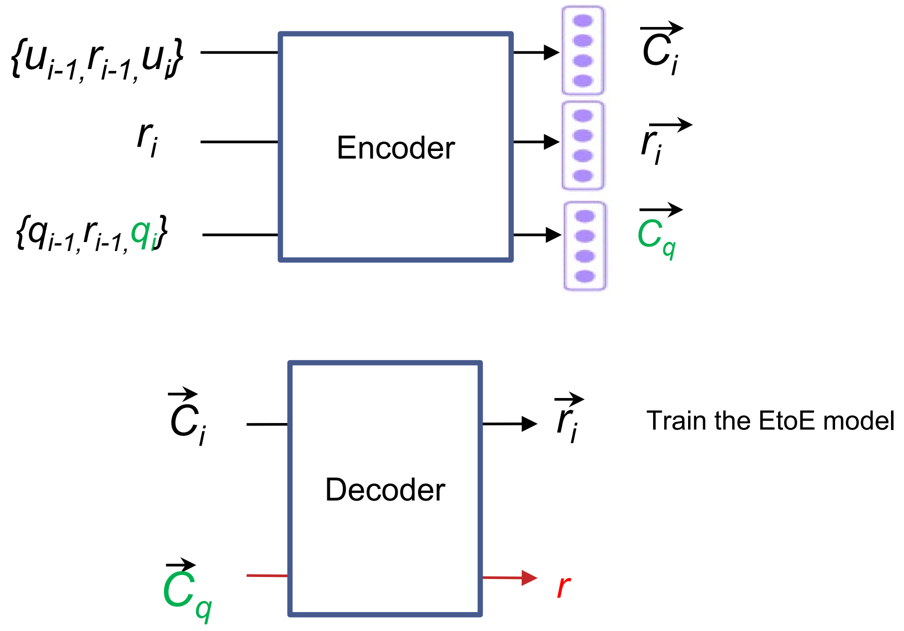


Advanced Retrieval-Based Approach

- Encode the **Utterances** and **Responses**

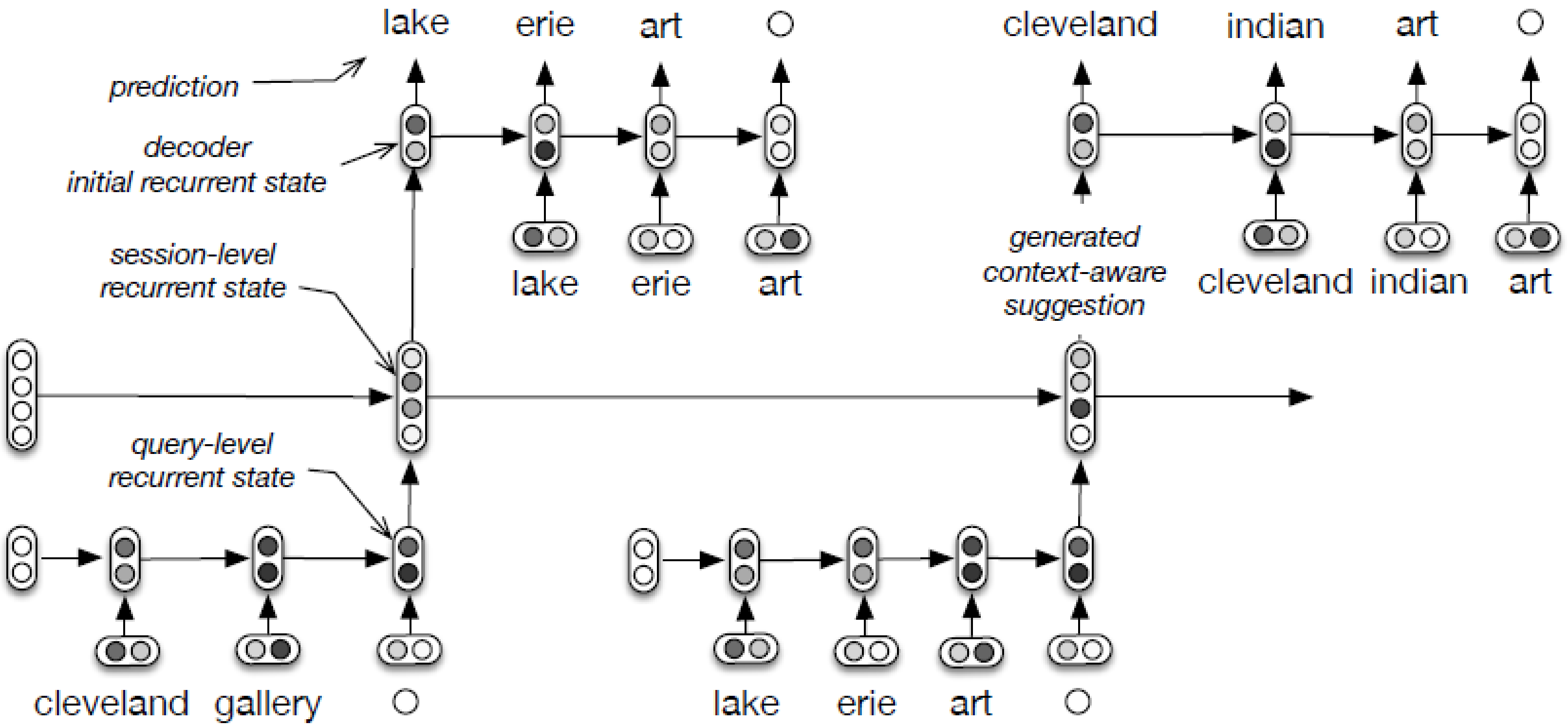


- Encode the **Context** and **Responses**



Embedding through Generative Models

- HRED
  - Inspired by Embedding through Language modelling “uncompleted sentence -> possible word”
  - “Query 1 -> Query 2” modelling for online searching or purchasing
  - Reused in “Question -> Answer ” modelling



Language Understanding for Machine Reading Comprehension

Machine Reading Comprehension

- AI-complete question answering
  - A set of prerequisite toy tasks
  - Detect Supporting facts

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

https://arxiv.org/abs/1503.08895

Machine Reading Comprehension

- AI-complete question answering
  - Memory Network

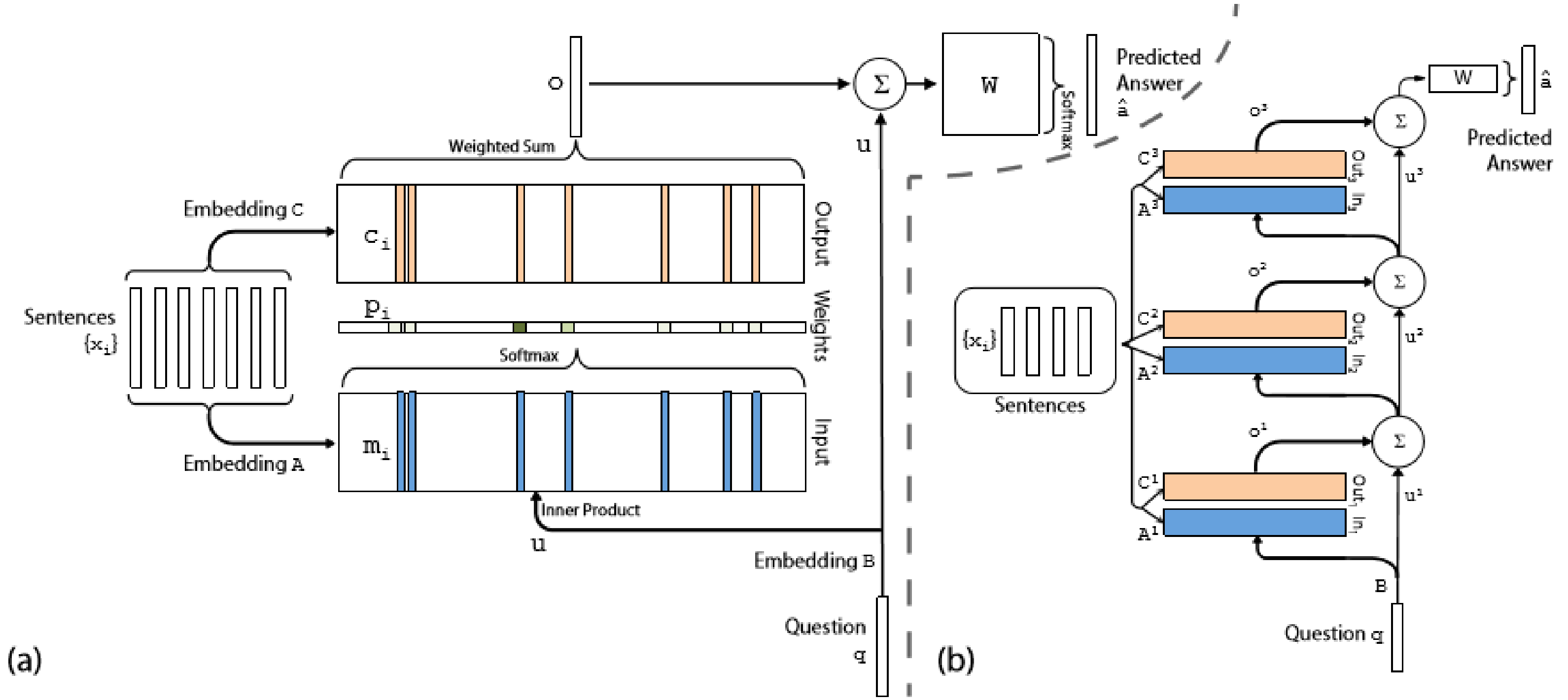


Figure 1: (a): A single layer version of our model. (b): A three layer version of our model. In practice, we can constrain several of the embedding matrices to be the same (see Section 2.2).

https://arxiv.org/abs/1503.08895

Machine Reading Comprehension

- Stanford Question Answering Dataset (SQuAD **2016**) Dataset
  - 100,000+ questions on top of Wikipedia
  - Manually ladled by crowdworkers on AMT
  - Constituency parses by Stanford CoreNLP as candidates
  - Logistic Regression as **baseline** model
  - 79 % Acc to locate the sentence
  - 51% F1 to identity the **exact span**

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

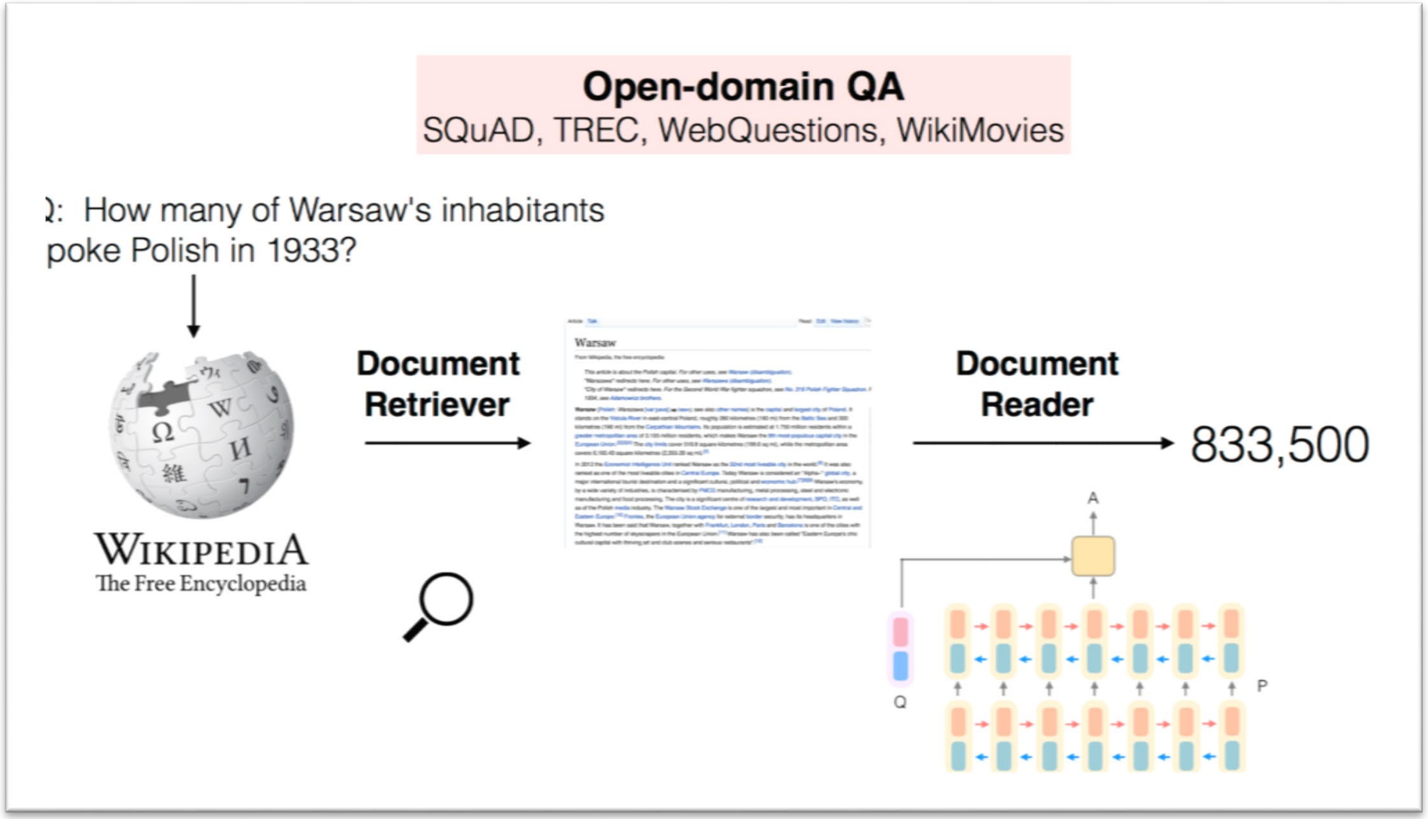
Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Figure 1: Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage.

https://rajpurkar.github.io/SQuAD-explorer/

Machine Reading Comprehension

- DrQA
  - Answering factoid questions while using Wikipedia
  - **Document Retriever**
    - Select relevant documents
    - Tfidf + Ngram +bag-of-words
  - **Document Reader**
    - Predict starting and Ending position
    - Encode question and paragraphes
    - **Multi-Layer** BI-LSTM
  - **Workshop**



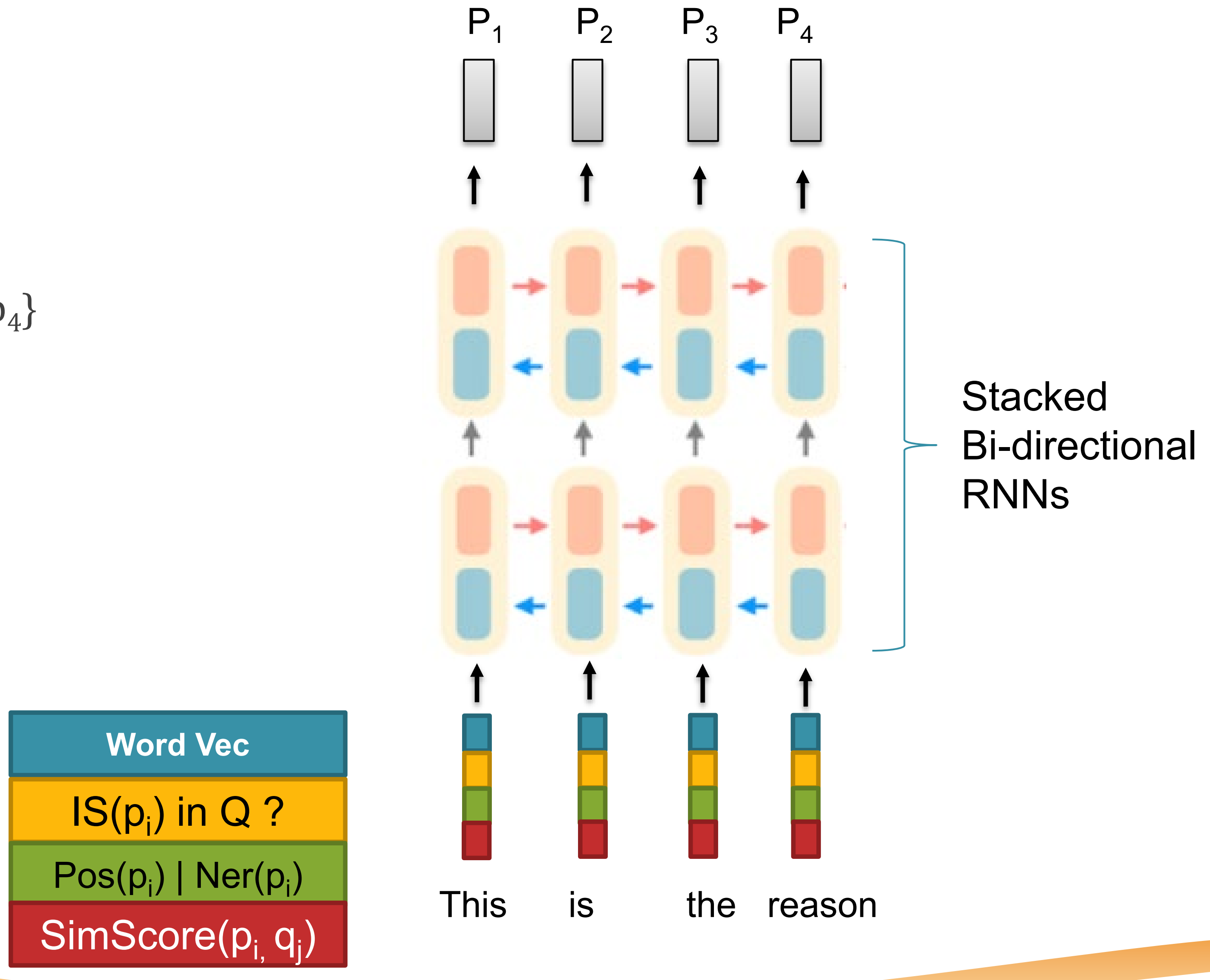
https://github.com/facebookresearch/DrQA/

Machine Reading Comprehension

- **Document Reader**
  - Q= "What is it" = $\{q_1\ q_2\ q_3\}$
  - P-1 ="It is a dog" = $\{p_1\ p_2\ p_3\ p_4\}$
  - P-2="This is the reason" = $\{p_1\ p_2\ p_3\ p_4\}$
- The task is to
  - determine P-1 is relevant to the question Q
    - TF-IDF weighted bag-of-word vectors
  - label the token **"a"** as the starting position of the answer, the token "dog" as the ending position of the answer

Machine Reading Comprehension

- **Document Reader**
  - Q= “What is it” = $\{q_1\, q_2\, q_3\}$
  - P-1 =“It is a dog” = $\{p_1\, p_2\, p_3\, p_4\}$
  - P-2=“This is the reason” = $\{p_1\, p_2\, p_3\, p_4\}$
  - **Paragraph Encoder**
    - **Multi-Layer** BI-LSTM

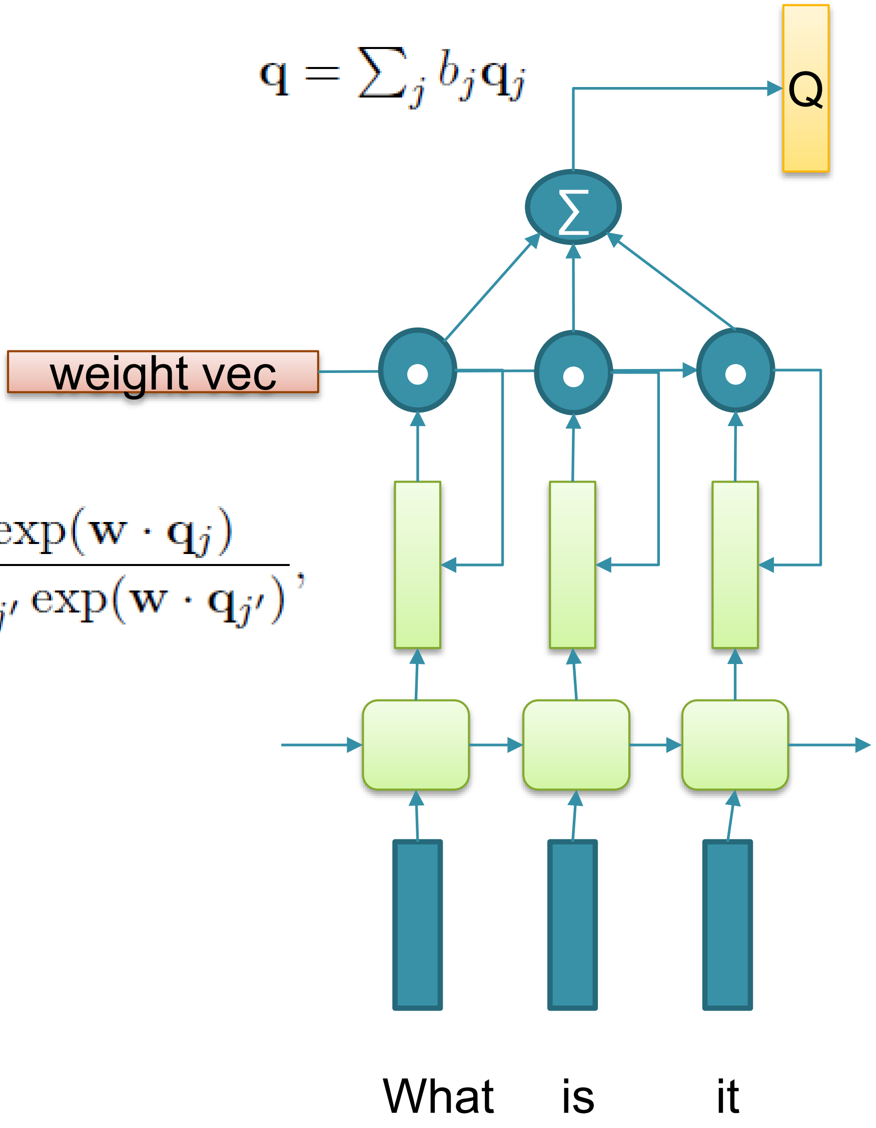


Machine Reading Comprehension

- **Document Reader**
  - Q= “What is it” = $\{q_1\, q_2\, q_3\}$
  - P-1 =“It is a dog” = $\{p_1\, p_2\, p_3\, p_4\}$
  - P-2=“This is the reason” = $\{p_1\, p_2\, p_3\, p_4\}$
- **Question Encoder**
  - **Word Embedding vectors**
  - **RNN**
  - **Weighted vector to learn**

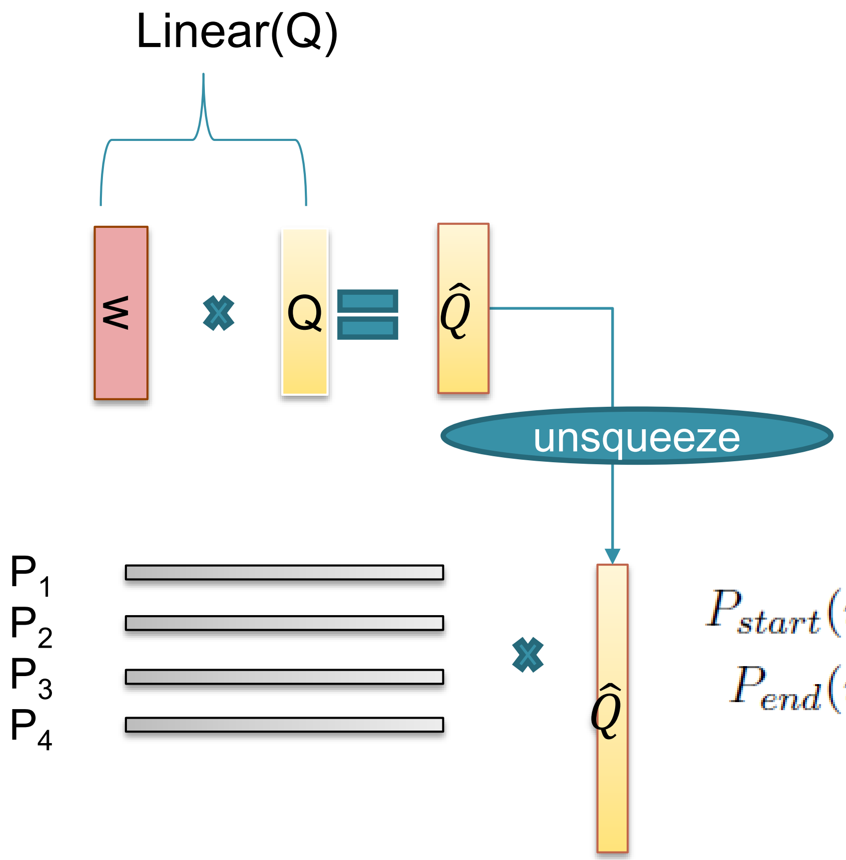
$$\mathbf{q} = \sum_j b_j \mathbf{q}_j$$

$$b_j = \frac{\exp(\mathbf{w} \cdot \mathbf{q}_j)}{\sum_{j'} \exp(\mathbf{w} \cdot \mathbf{q}_{j'})},$$



Machine Reading Comprehension

- **Document Reader**
  - Q= "What is it"   = $\{q_1\ q_2\ q_3\}$
  - P-1 ="It is a dog" = $\{p_1\ p_2\ p_3\ p_4\}$
  - P-2="This is the reason" = $\{p_1\ p_2\ p_3\ p_4\}$
  - **Prediction**
    - Two independent classifiers
    - Classifier for Start of answer
    - Classifier for End of answer
    - Take ***argmax*** for most likely span



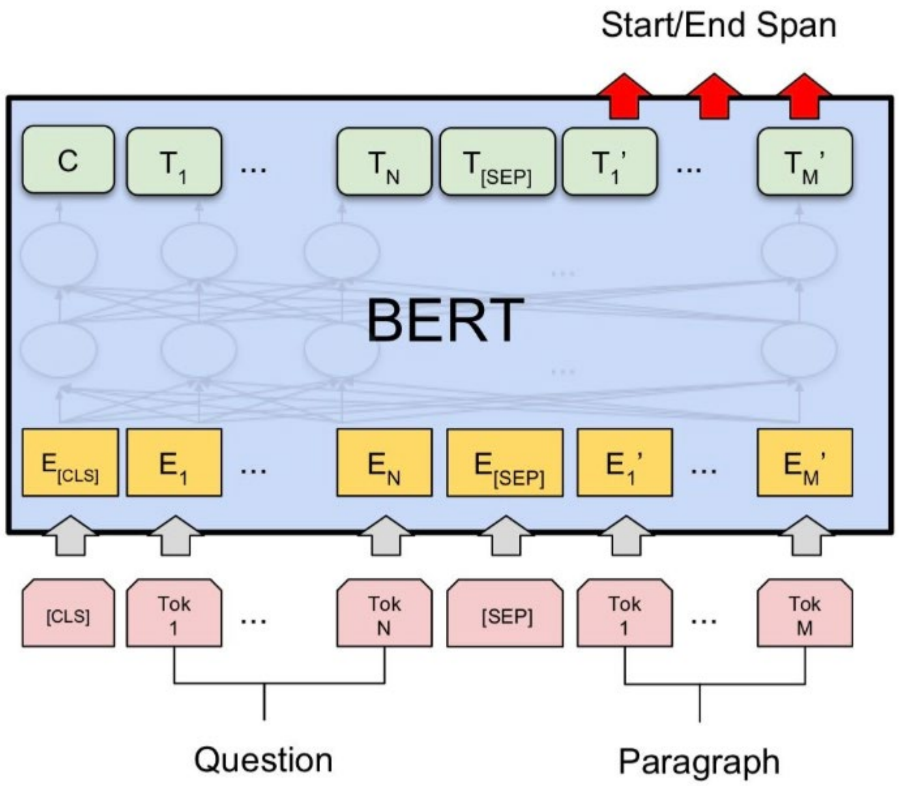
$$P_{start}(i) \propto \exp(\mathbf{p}_i \mathbf{W}_s \mathbf{q})$$
$$P_{end}(i) \propto \exp(\mathbf{p}_i \mathbf{W}_e \mathbf{q})$$

Unnormalized
Softmax
(Start/End)

Machine Reading Comprehension

- Some Other Architectures
  - BERT based model
  - BERT embedding
  - Linear layer mapping to two labels:
    - Start/End of span
    - Minimize the average loss of Start and End

- More variants can be founded
  - https://paperswithcode.com/sota/question-answering-on-squad20



Reference

- Galitsky, Boris. "Developing Enterprise Chatbots." (2019).
- McTear, Michael, Zoraida Callejas, and David Griol. *The conversational interface: Talking to smart devices*. Springer, 2016.
- Xu, Anbang, et al. "A new chatbot for customer service on social media." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017.
- https://www.invespcro.com/blog/chatbots-customer-service
- http://customerthink.com/5-reasons-why-your-chatbot-needs-natural-language-processing

Reference

- Sordoni, Alessandro, et al. “A hierarchical recurrent encoder-decoder for generative context-aware query suggestion.” (CIKM 2015)
- Sukhbaatar, Sainbayar, Jason Weston, and Rob Fergus. "End-to-end memory networks." Advances in neural information processing systems. 2015.
- Chen, Y.-N., Hakkani-T¨ur, D., Tur, G., Gao, J., and Deng, L. (2016b). End-to-end memory networks with knowledge carryover for multi-turn spoken language understanding
- Wu, C.-S., Madotto, A., Winata, G. I., and Fung, P. (2018). End-to-end dynamic query memory network for entity-value independent task-oriented dialog.
- Chen, D., Fisch, A.,Weston, J., and Bordes, A. (2017a). ReadingWikipedia to answer open-domain questions. arXiv 1704.00051
- Chen, Y.-N. and Gao, J. (2017). Open-domain neural dialogue systems. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Tutorial Abstracts), pages 6–10.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.