Machine Reading, Models and Applications

Quentin Grail, Arnaud Sors, Morgan Funtowicz, **Julien Perez**
Machine Learning and Optimization group

6th September, 2018

LABS
NAVER LABS

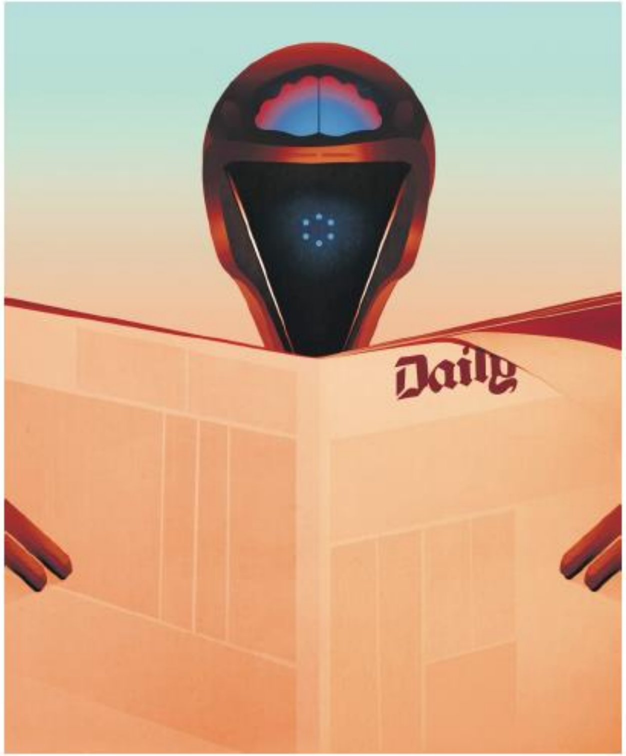
Content

1. Machine reading tasks

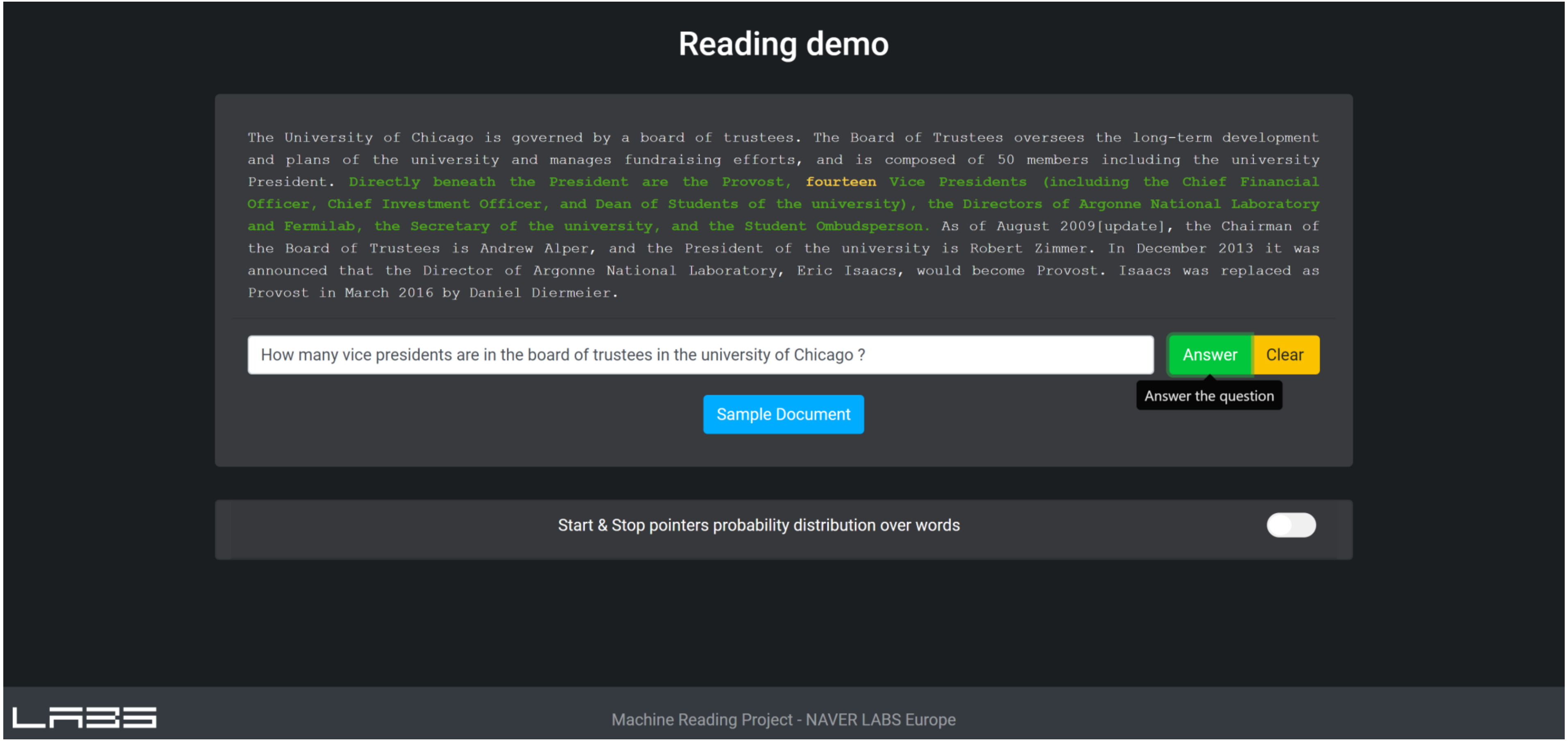
2. Models of reading

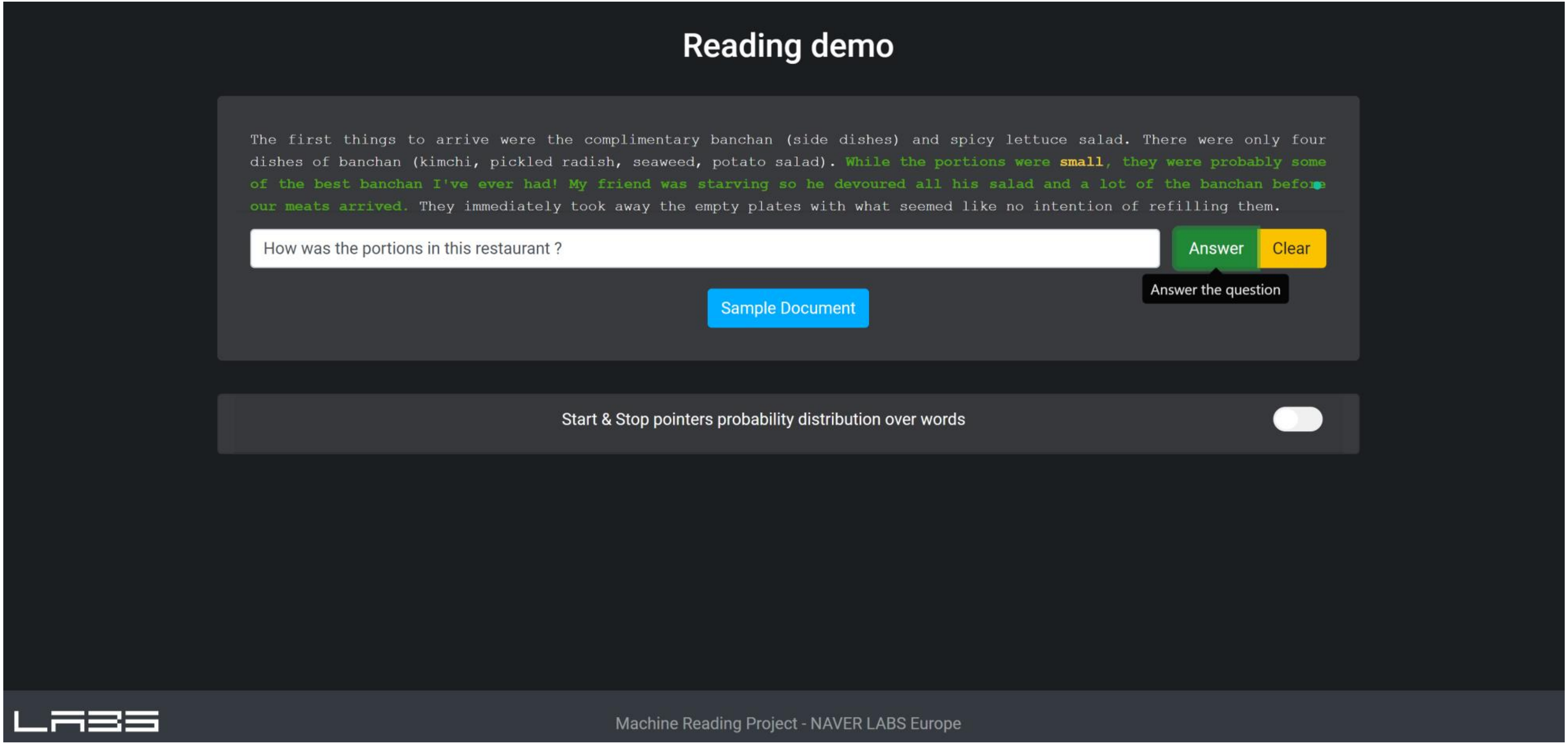
3. Applications

4. Open Questions



Courtesy of Phil Blunsom


Reading demo
The University of Chicago is governed by a board of trustees. The Board of Trustees oversees the long-term development and plans of the university and manages fundraising efforts, and is composed of 50 members including the university President. Directly beneath the President are the Provost, fourteen Vice Presidents (including the Chief Financial Officer, Chief Investment Officer, and Dean of Students of the university), the Directors of Argonne National Laboratory and Fermilab, the Secretary of the university, and the Student Ombudsperson. As of August 2009[update], the Chairman of the Board of Trustees is Andrew Alper, and the President of the university is Robert Zimmer. In December 2013 it was announced that the Director of Argonne National Laboratory, Eric Isaacs, would become Provost. Isaacs was replaced as Provost in March 2016 by Daniel Diermeier.
How many vice presidents are in the board of trustees in the university of Chicago ?
Answer
Clear
Answer the question
Sample Document
Start & Stop pointers probability distribution over words
Machine Reading Project - NAVER LABS Europe


Reading demo
The first things to arrive were the complimentary banchan (side dishes) and spicy lettuce salad. There were only four dishes of banchan (kimchi, pickled radish, seaweed, potato salad). While the portions were small, they were probably some of the best banchan I've ever had! My friend was starving so he devoured all his salad and a lot of the banchan before our meats arrived. They immediately took away the empty plates with what seemed like no intention of refilling them.
How was the portions in this restaurant ?
Answer
Clear
Answer the question
Sample Document
Start & Stop pointers probability distribution over words
Machine Reading Project - NAVER LABS Europe

Multi Documents Answering

In Blade Runner, which company built the replicants ?

Wikipedia | Predict | Clear

Matched documents

Blade Runner (franchise)	657.27
Blade Runner	657.27
Blade Runner 2049	609.13
Replicant	609.13
Blade Runner (1997 video game)	543.61

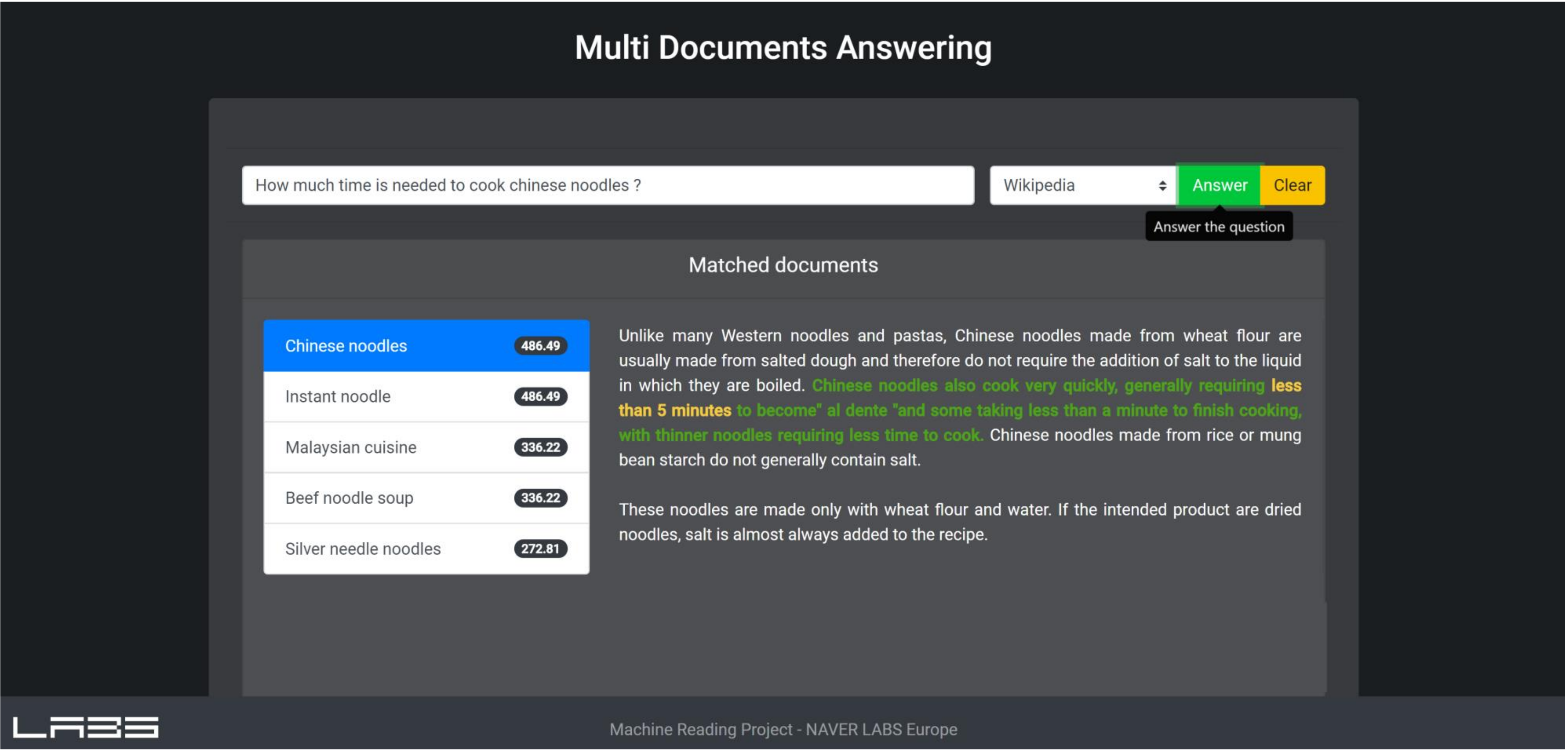
Blade Runner

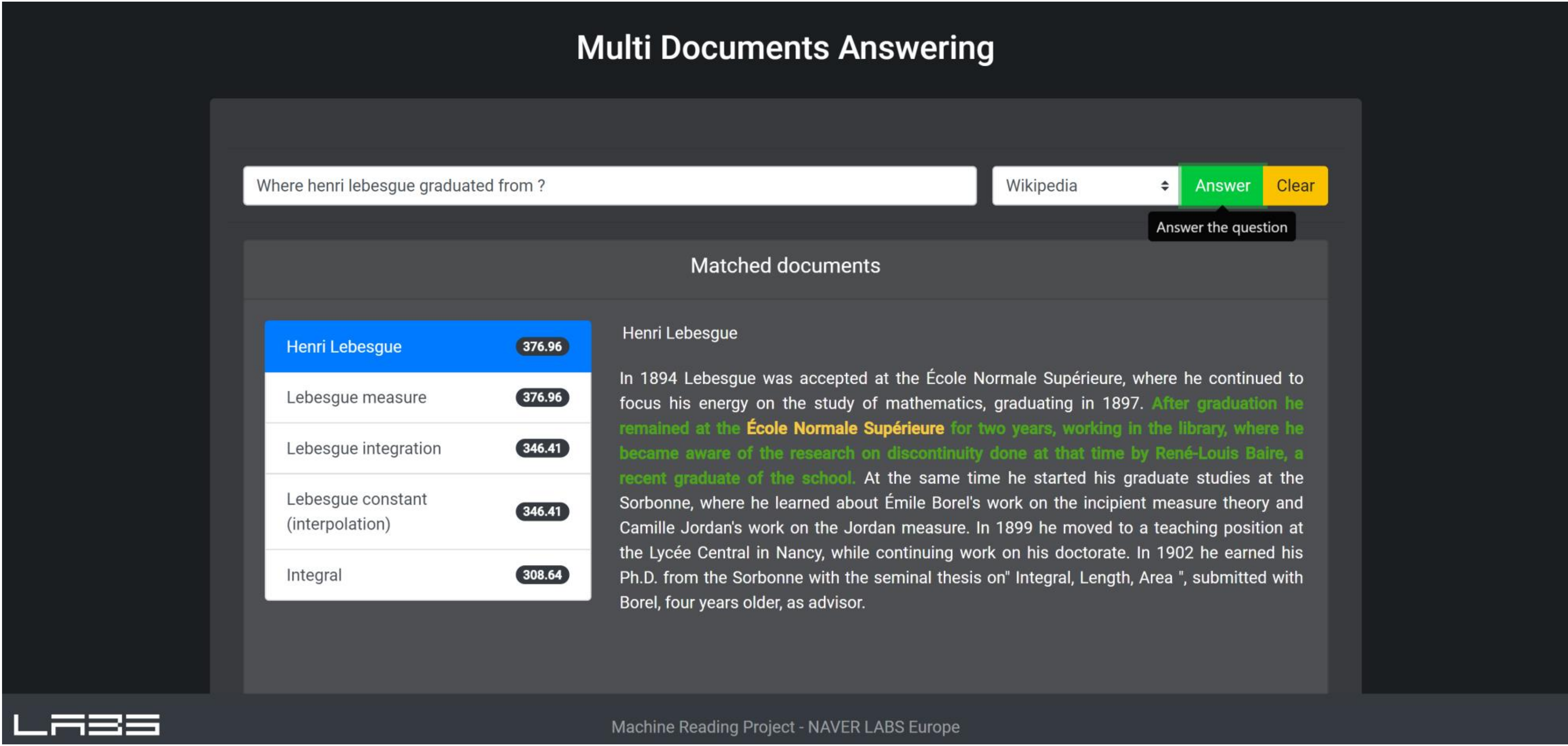
Blade Runner is a 1982 American neo-noir science fiction film directed by Ridley Scott, written by Hampton Fancher and David Peoples, and starring Harrison Ford, Rutger Hauer, Sean Young, and Edward James Olmos. It is a loose adaptation of Philip K. Dick's novel "Do Androids Dream of Electric Sheep?" (1968). **The film is set in a dystopian future Los Angeles of 2019, in which synthetic humans known as replicants are bioengineered by the powerful Tyrell Corporation to work on off-world colonies.** When a fugitive group of replicants led by Roy Batty (Hauer) escapes back to Earth, burnt-out cop Rick Deckard (Ford) reluctantly agrees to hunt them down.

" Blade Runner "initially underperformed in North American theaters and polarized critics; some praised its thematic complexity and visuals, while others were displeased with its unconventional pacing and plot. It later became an acclaimed cult film regarded as one of the all-time best science fiction movies. Hailed for its production design depicting a" retrofitted "future," Blade Runner "is a leading example of neo-noir cinema. The soundtrack, composed by Vangelis, was nominated in 1983 for a BAFTA and a Golden Globe as best original score.

The film has influenced many science fiction films, video games, anime, and television series. It brought the work of Philip K. Dick to the attention of Hollywood, and several later big-budget films were based on his work. In the year after its release," Blade Runner "won the Hugo Award for Best Dramatic Presentation, and in 1993 it was selected for preservation in

Machine Reading Project - NAVER LABS Europe


Multi Documents Answering
How much time is needed to cook chinese noodles ?
Wikipedia
Answer
Clear
Answer the question
Matched documents
Chinese noodles 486.49
Instant noodle 486.49
Malaysian cuisine 336.22
Beef noodle soup 336.22
Silver needle noodles 272.81
Unlike many Western noodles and pastas, Chinese noodles made from wheat flour are usually made from salted dough and therefore do not require the addition of salt to the liquid in which they are boiled. Chinese noodles also cook very quickly, generally requiring less than 5 minutes to become" al dente "and some taking less than a minute to finish cooking, with thinner noodles requiring less time to cook. Chinese noodles made from rice or mung bean starch do not generally contain salt.
These noodles are made only with wheat flour and water. If the intended product are dried noodles, salt is almost always added to the recipe.
Machine Reading Project - NAVER LABS Europe


Multi Documents Answering
Where henri lebesgue graduated from ?
Wikipedia
Answer
Clear
Answer the question
Matched documents
Henri Lebesgue 376.96
Lebesgue measure 376.96
Lebesgue integration 346.41
Lebesgue constant (interpolation) 346.41
Integral 308.64
Henri Lebesgue
In 1894 Lebesgue was accepted at the École Normale Supérieure, where he continued to focus his energy on the study of mathematics, graduating in 1897. After graduation he remained at the École Normale Supérieure for two years, working in the library, where he became aware of the research on discontinuity done at that time by René-Louis Baire, a recent graduate of the school. At the same time he started his graduate studies at the Sorbonne, where he learned about Émile Borel's work on the incipient measure theory and Camille Jordan's work on the Jordan measure. In 1899 he moved to a teaching position at the Lycée Central in Nancy, while continuing work on his doctorate. In 1902 he earned his Ph.D. from the Sorbonne with the seminal thesis on" Integral, Length, Area ", submitted with Borel, four years older, as advisor.
Machine Reading Project - NAVER LABS Europe

Machine Reading

motivations

Human knowledge is (**mainly**) stored in natural language

Natural Language is an **efficient** support of knowledge transcription

Languages assume **apriori knowledge** of the world a.k.a **common sense**

Language is efficient because of its **contextuallity** that leads to **ambiguity**

The Library of Trinity College Dublin

Machine Reading

Definition

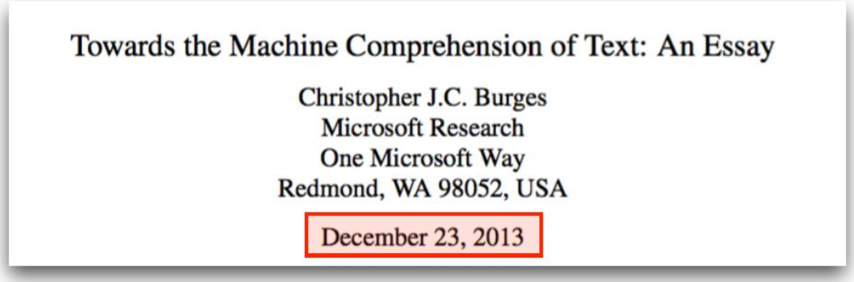
"A machine comprehends **a passage of text** if, for any **question** regarding that text, it can be **answered** correctly by a majority of native speakers.

The machine needs to provide a string which human readers would agree both

1. Answers that question
2. Does not contain information irrelevant to that question." *(Burges, 2013)*

Applications

- Collection of documents as KB
- Social media mining
- Dialog understanding
- Fact checking – Fake news detection

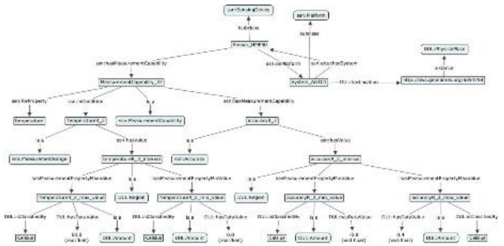


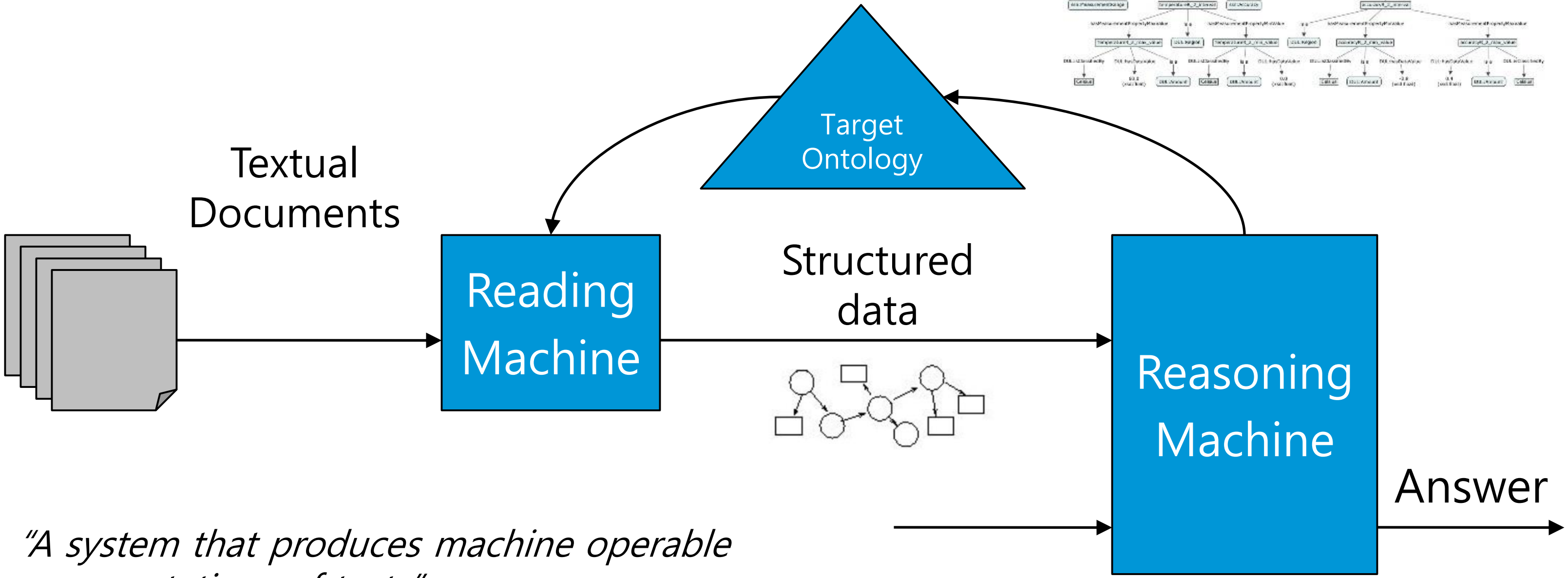
Towards the Machine Comprehension of Text: An Essay

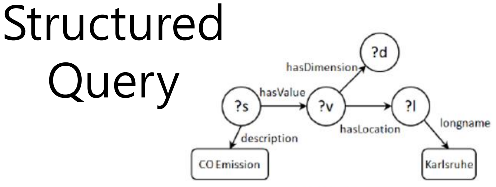
Christopher J.C. Burges
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA

December 23, 2013

Knowledge Base approach







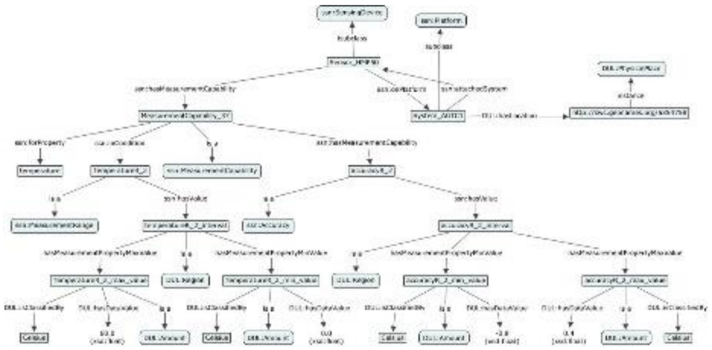
(a) Query Graph

"A system that produces machine operable representations of texts"

[1] **Machine Reading**, Peñas et al, 2011

Knowledge Base approach

" A system that produces machine operable representations of texts "

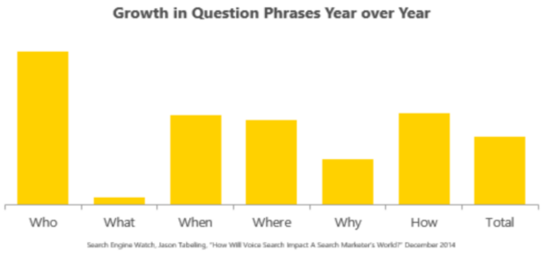


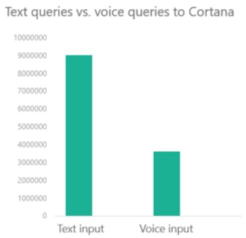
... but we have 3 problems here

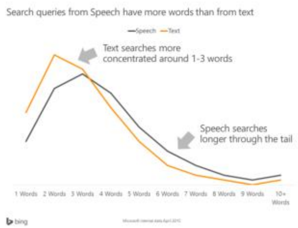
1. *Fixed/Predefined ontologies*
2. *Fixed/Predefined lexical domain*
3. ***Data duplication by structuration***

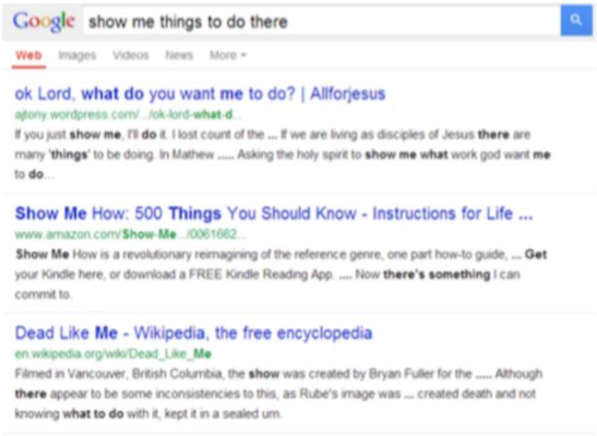
Information retrieval approach

"Information Retrieval (IR) is finding material, usually **documents**, of an unstructured nature, usually text, that s**atisfies an information need** from within large collections usually stored on computers." Manning, Introduction to IR.









[2] **Introduction to information Retrieval**, Manning et al, 2008

Classic Deep NLP approach

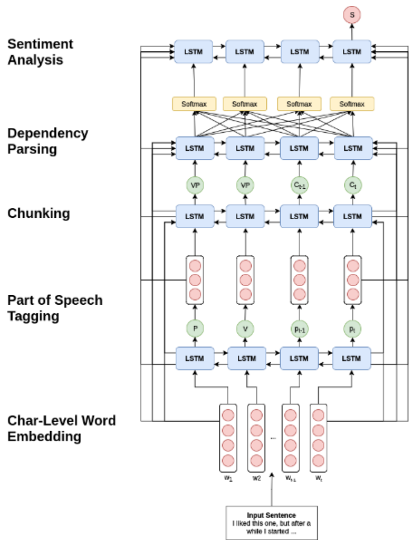
"Machine reading, yet another (Deep) NLP task ? "

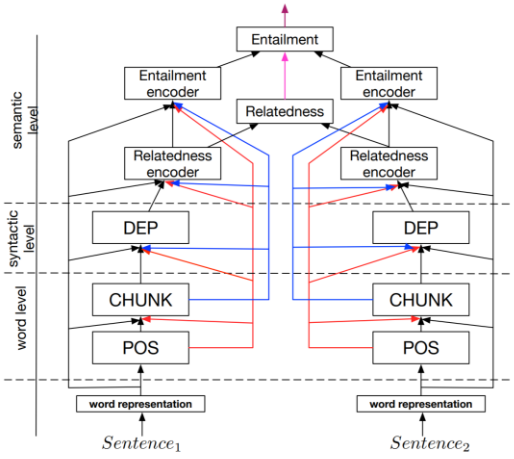
… but we have 3 problems here

1. Is (Language dependant) syntax a requirement to semantics ?

2. Additional (unnecessary) requirement
   - Annotations
   - Priors

3. Not end-to-end machine comprehension





Machine Reading

as Multi-choice question task

MCTest

- 500 passages
- 2000 questions about simple stories

RACE

- 28,000 passages
- 100,000 questions from English comprehension tests

$$\mathcal{L}(b;\theta) = \frac{1}{|b|}\sum_{i=1}^{|b|}(\mathcal{S}(\{q,d\}_i;\theta) - s_{\{q,d\}_i})^2 \qquad \mathcal{L}(b;\theta) = \frac{1}{|b|}\sum_{i=1}^{|b|}\max\left\{0, \varepsilon - s_{\{q,d_1\}_i} - s_{\{q,d_2\}_i}\right\}$$

[5] MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text,
Richardson et al, 2013
[6] RACE: Large-scale ReAding Comprehension Dataset From Examinations,
Lai et al, 2017

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.
One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.
His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.
After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
A) Fries
B) Pudding
C) James
D) Jane

2) What did James pull off of the shelves in the grocery store?
A) pudding
B) fries
C) food
D) splinters

3) Where did James go after he went to the grocery store?
A) his deck
B) his freezer
C) a fast food restaurant
D) his room

4) What did James do after he ordered the fries?
A) went to the grocery store
B) went home without paying
C) ate them
D) made up his mind to be a better turtle

Machine Reading

as Span selection

SQuAD

- 500 passages
- 100,000 questions on Wikipedia text
- Human annotated

TriviaQA

- 95k questions
- 650k evidence documents
- distant supervision

$$CCE = -\frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j \cdot log(\hat{y}_j) + (1 - y_j) \cdot log(1 - \hat{y}_j)$$

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

[6] **SQuAD: 100,000+ Questions for Machine Comprehension of Text,** Liang et al, 2016
[7] **TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension**, Zottlemoyer et al, 2017

Machine reading

Reasoning over knowledge extraction

- Textual data can specify reasoning capabilities
- **Goal**: build machines that can "understand " textual information, *i.e.* converting it into interpretable structured knowledge to be leveraged by humans and other machines alike.
- Optimized with categorical cross-entropy loss

$$CCE = -\frac{1}{N}\sum_{i=0}^{N}\sum_{j=0}^{J} y_j \cdot log(\hat{y}_j) + (1-y_j)\cdot log(1-\hat{y}_j)$$

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions
John moved to the playground.
Daniel went to the bathroom.
John went back to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets
Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

Task 9: Simple Negation
Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge
John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

[9] **Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks,** Weston and al

Machine Reading

Datasets

Before 2015:

- MCTest (Richardson et al, 2013): 2600 questions
- ProcessBank (Berant et al, 2014): 500 questions

More than 100k questions!

After 2015:

- **CNN/Daily Mail**
- Children Book Test
- WikiReading
- LAMBADA
- **SQuAD**
- Who did What
- Maluuba NewsQA
- MS MARCO
- NAVER DSTC6-T1

Building blocks

Recurrent Neural Network

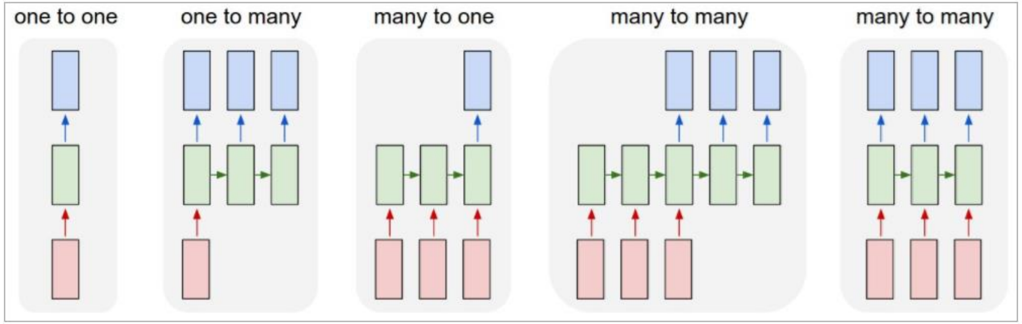
LSTM with a forget gate

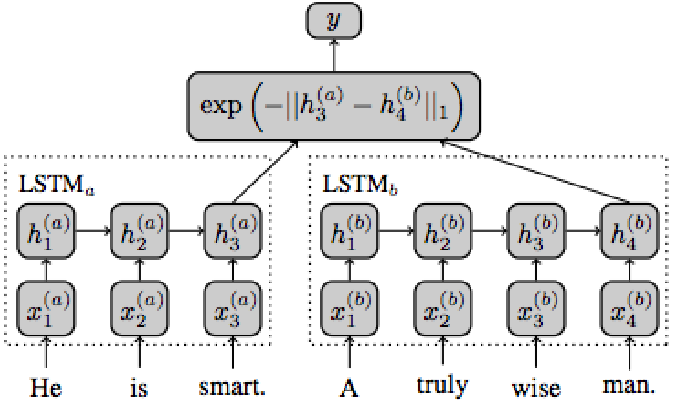
$$
\begin{aligned}
f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
h_t &= o_t \circ \sigma_h(c_t)
\end{aligned}
$$

where the initial values are $c_0 = 0$ and $h_0 = 0$
and the operator $\circ$ denotes the Hadamard product (entry-wise product).
The subscripts t refer to the time step.

Variables

- $x_t \in \mathbb{R}^d$: input vector to the LSTM unit
- $f_t \in \mathbb{R}^h$: forget gate's activation vector
- $i_t \in \mathbb{R}^h$: input gate's activation vector
- $o_t \in \mathbb{R}^h$: output gate's activation vector
- $h_t \in \mathbb{R}^h$: output vector of the LSTM unit
- $c_t \in \mathbb{R}^h$: cell state vector
- $W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$ and $b \in \mathbb{R}^h$: weight matrices and bias vector parameters





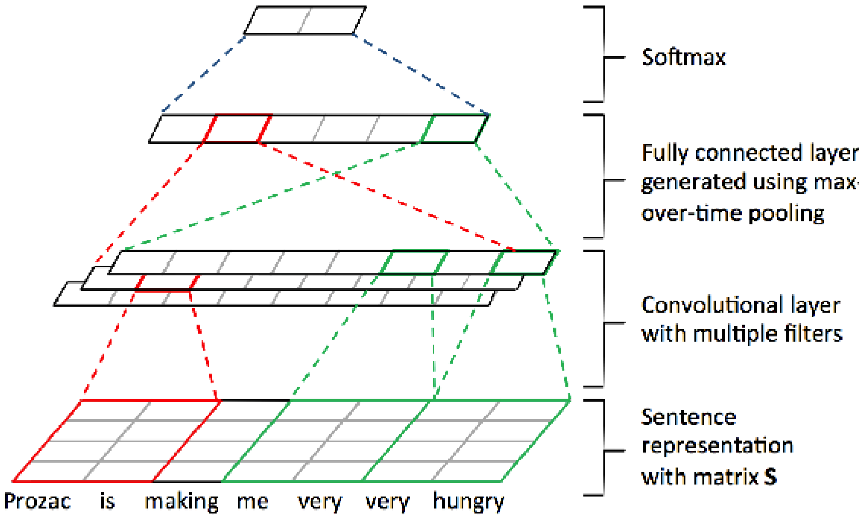
[10] **Long Short Term Memory**, Schmidhuber et al, 1997

Building blocks

Convolutional Network

Elements:

- Input sentence: $\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \ldots \oplus \mathbf{x}_n$
- Output local feature: $c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b)$
- Feature map: $\mathbf{c} = [c_1, c_2, \ldots, c_{n-h+1}]$
- Max-pooling layer
- Fully connected layer with softmax output for classification tasks

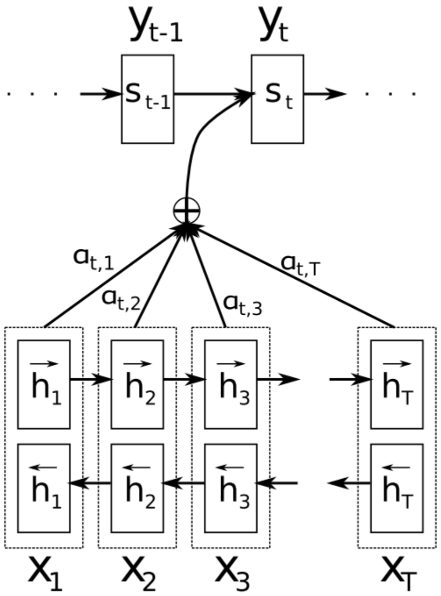


... Trivial to parallelize

[11] **Convolutional Neural Networks for Sentence Classification, Kim et al, 2017**

Building blocks

Attention mechanism



In Neural Machine Translation

- Encode each work in the input and output sentence into a vector
- Perform a linear combination of these vectors, weighted by « **attention score** »
- Use this combination as support to pick the next word

$$\alpha_{ts} = \frac{\exp\left(\mathrm{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s)\right)}{\sum_{s'=1}^{S} \exp\left(\mathrm{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_{s'})\right)} \qquad \text{[Attention weights]} \qquad (1)$$

$$\boldsymbol{c}_t = \sum_s \alpha_{ts} \bar{\boldsymbol{h}}_s \qquad \text{[Context vector]} \qquad (2)$$

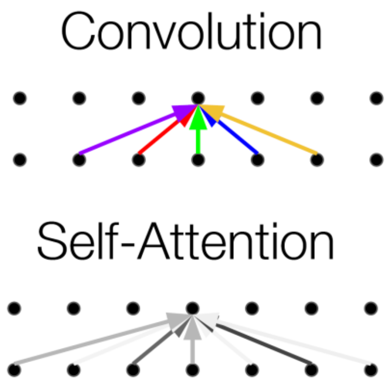
$$\boldsymbol{a}_t = f(\boldsymbol{c}_t, \boldsymbol{h}_t) = \tanh(\boldsymbol{W}_c[\boldsymbol{c}_t; \boldsymbol{h}_t]) \qquad \text{[Attention vector]} \qquad (3)$$

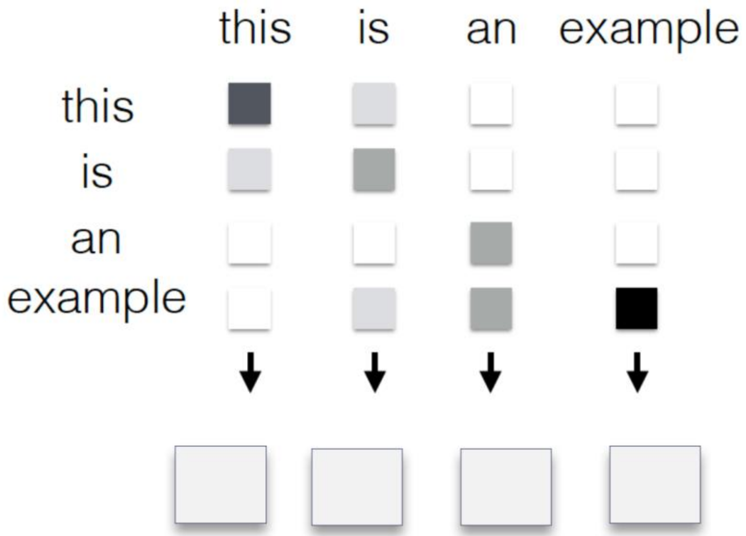
[12] **Neural Machine Translation by Jointly Learning to Align and Translate**, Badhanau et al, 2015

Building blocks

Self-Attention mechanism

Each element in the sentence attends to other elements from the SAME sentence → context sensitive encodings!



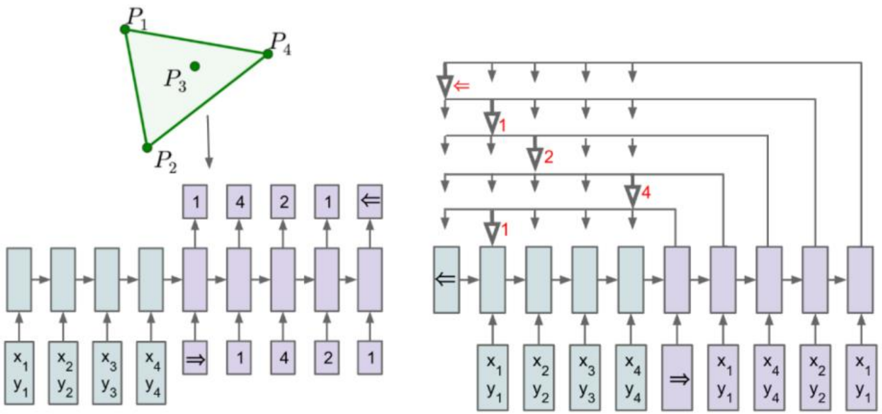


[13] **Attention Is All You Need**, Polosukhin et al, 2017

Building blocks

Pointer Networks

- Pointer networks are a variation of the seq-to-seq models.
- Instead of translating one sequence into another, the output is a sequence of pointers to the elements of the input series (i.e a permutation of the input sequence)



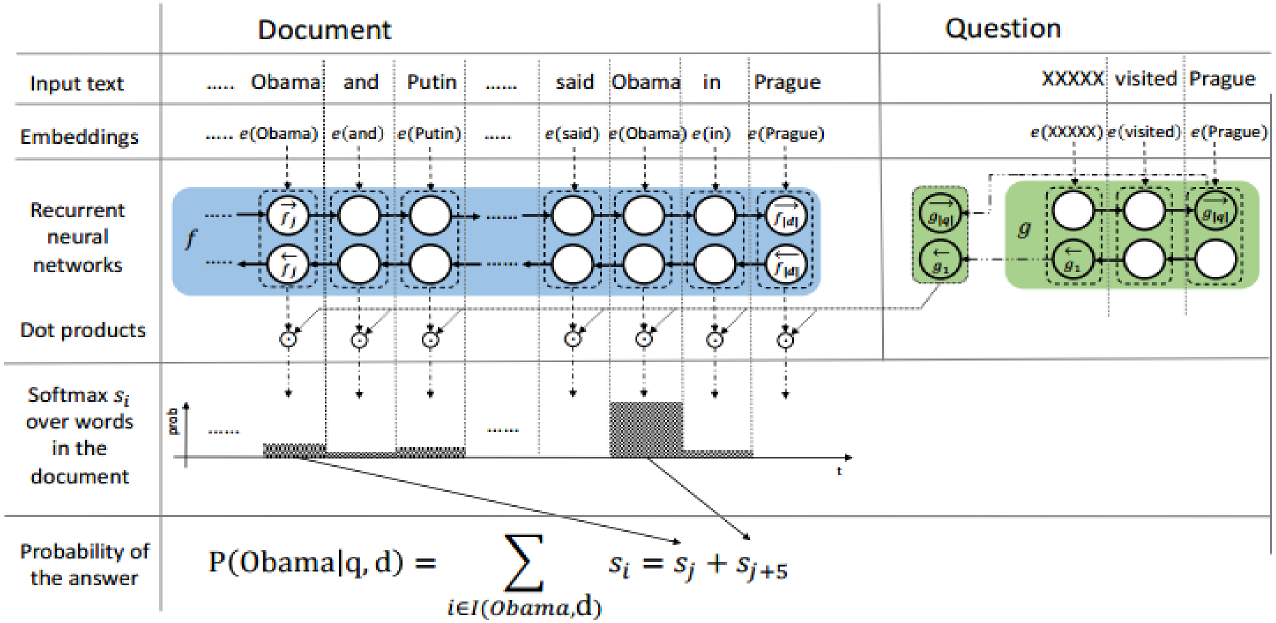
(a) Sequence-to-Sequence

(b) Ptr-Net

[14] **Pointer Networks**, Vinyals et al, 2015

Extractive models

Attention Sum Reader Network



$$s_i \propto \exp\left(f_i(\mathbf{d}) \cdot g(\mathbf{q})\right) \quad (1)$$

$$P(w|\mathbf{q},\mathbf{d}) \propto \sum_{i \in I(w,\mathbf{d})} s_i \quad (2)$$

where $I(w, \mathbf{d})$ is a set of positions where w appears in the document $\mathbf{d}$.

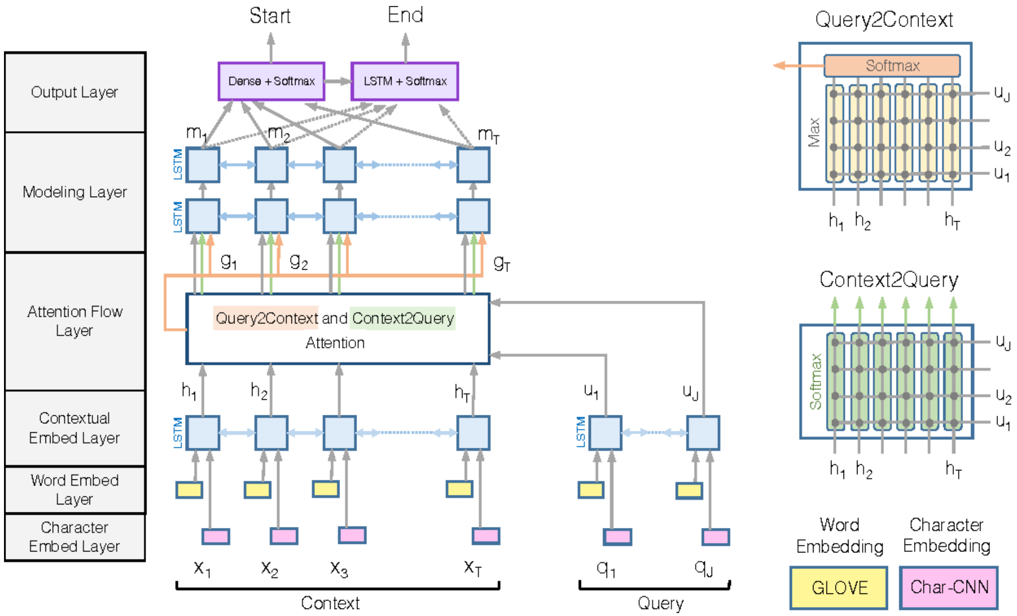
$$f_i(\mathbf{d}) = \overrightarrow{f_i}(\mathbf{d}) \;||\; \overleftarrow{f_i}(\mathbf{d}),$$

$$g(\mathbf{q}) = \overrightarrow{g_{|\mathbf{q}|}}(\mathbf{q}) \;||\; \overleftarrow{g_1}(\mathbf{q}).$$

[15] **Text Understanding with the Attention Sum Reader Network**, Kadlec et al, 2016

Extractive models

Bidirectional Attention Flow for Machine Comprehension



	CNN		DailyMail	
	val	test	val	test
Attentive Reader (Hermann et al., 2015)	61.6	63.0	70.5	69.0
MemNN (Hill et al., 2016)	63.4	6.8	-	-
AS Reader (Kadlec et al., 2016)	68.6	69.5	75.0	73.9
DER Network (Kobayashi et al., 2016)	71.3	72.9	-	-
Iterative Attention (Sordoni et al., 2016)	72.6	73.3	-	-
EpiReader (Trischler et al., 2016)	73.4	74.0	-	-
Stanford AR (Chen et al., 2016)	73.8	73.6	77.6	76.6
GAReader (Dhingra et al., 2016)	73.0	73.8	76.7	75.7
AoA Reader (Cui et al., 2016)	73.1	74.4	-	-
ReasoNet (Shen et al., 2016)	72.9	74.7	77.6	76.6
BIDAF (Ours)	**76.3**	**76.9**	**80.3**	**79.6**
MemNN* (Hill et al., 2016)	66.2	69.4	-	-
ASReader* (Kadlec et al., 2016)	73.9	75.4	78.7	77.7
Iterative Attention* (Sordoni et al., 2016)	74.5	75.7	-	-
GA Reader* (Dhingra et al., 2016)	76.4	77.4	79.1	78.1
Stanford AR* (Chen et al., 2016)	77.2	77.6	80.2	79.2

[16] **Bidirectional Attention Flow for Machine Comprehension**, Seo et al, 2016

Extractive models

R-Net

- Extractive model
- Fully differentiable
- Based on 4 stacked layers
- Language independent

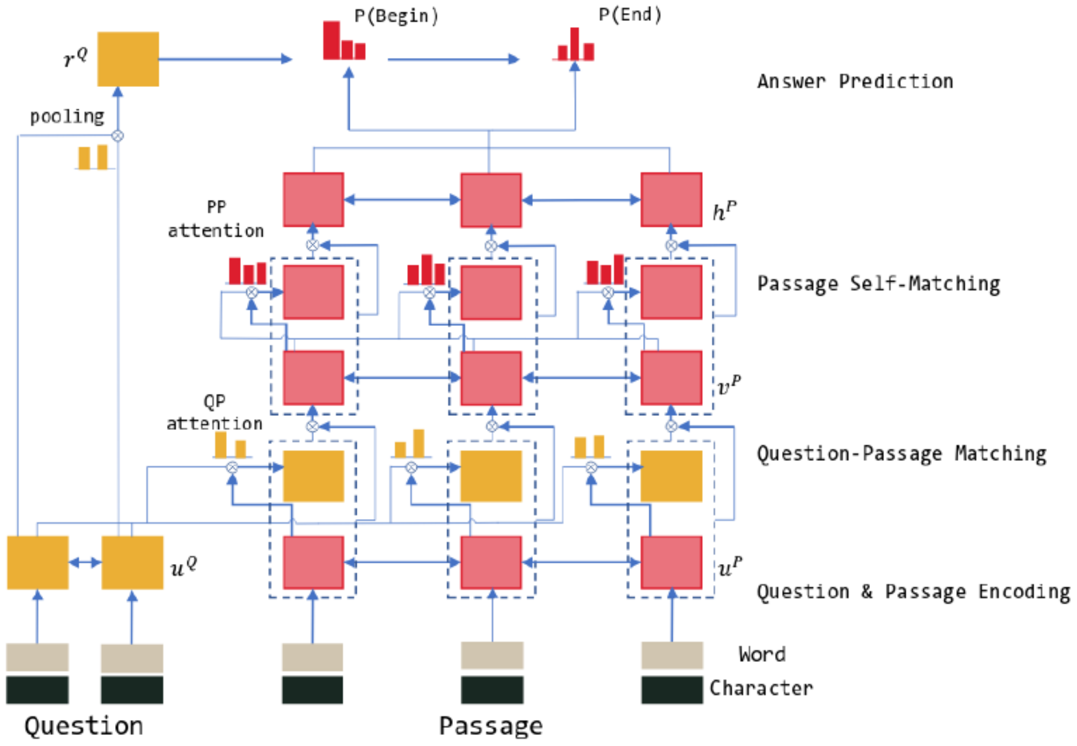


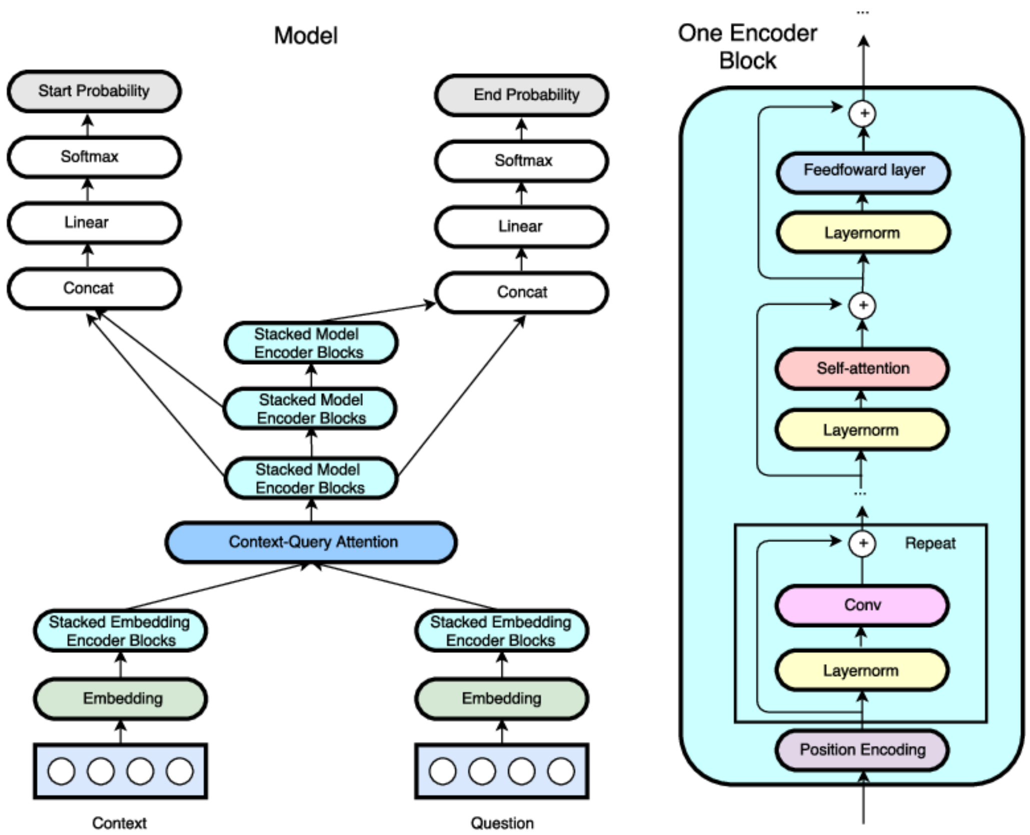
Figure 1: R-NET structure overview.

[17] **R-Net, technical report**, Microsoft Asia, 2017

Extractive models

Google QANet

- Extractive model
- Fully differentiable
- Non-autoregressive model
- Language independant
- « Attention is All you Need »



[18] **Combining Local Convolution with Global Self-Attention for Reading Comprehension**, Google Research, 2017

Extractive models

Error analysis

Error type	Ratio (%)	Example
Imprecise answer boundaries	50	**Context**: "The Free Movement of Workers Regulation articles 1 to 7 set out the main provisions on equal treatment of workers." **Question**: "Which articles of the Free Movement of Workers Regulation set out the primary provisions on equal treatment of workers?" **Prediction**: "1 to 7", **Answer**: "articles 1 to 7"
Syntactic complications and ambiguities	28	**Context**: "A piece of paper was later found on which Luther had written his last statement. " **Question**: "What was later discovered written by Luther?" **Prediction**: "A piece of paper", **Answer**: "his last statement"
Paraphrase problems	14	**Context**: "Generally, education in Australia follows the three-tier model which includes primary education (primary schools), followed by secondary education (secondary schools/high schools) and tertiary education (universities and/or TAFE colleges)." **Question**: "What is the first model of education, in the Australian system?" **Prediction**: "three-tier", **Answer**: "primary education"
External knowledge	4	**Context**: "On June 4, 2014, the NFL announced that the practice of branding Super Bowl games with Roman numerals, a practice established at Super Bowl V, would be temporarily suspended, and that the game would be named using Arabic numerals as Super Bowl 50 as opposed to Super Bowl L." **Question**: "If Roman numerals were used in the naming of the 50th Super Bowl, which one would have been used?' **Prediction**: "Super Bowl 50", **Answer**: "L"
Multi-sentence	2	**Context**: "Over the next several years in addition to host to host interactive connections the network was enhanced to support terminal to host connections, host to host batch connections (remote job submission, remote printing, batch file transfer), interactive file transfer, gateways to the Tymnet and Telenet public data networks, X.25 host attachments, gateways to X.25 data networks, Ethernet attached hosts, and eventually TCP/IP and additional public universities in Michigan join the network. All of this set the stage for Merit's role in the NSFNET project starting in the mid-1980s." **Question**: "What set the stage for Merits role in NSFNET" **Prediction**: "All of this set the stage for Merit 's role in the NSFNET project starting in the mid-1980s", **Answer**: "Ethernet attached hosts, and eventually TCP/IP and additional public universities in Michigan join the network"
Incorrect preprocessing	2	**Context**: "English chemist John Mayow (1641-1679) refined this work by showing that fire requires only a part of air that he called spiritus nitroaereus or just nitroaereus." **Question**: "John Mayow died in what year?" **Prediction**: "1641-1679", **Answer**: "1679"

Reasoning models

Gated End-to-end memory networks

$$\boldsymbol{m}_i = \boldsymbol{A}\Phi(x_i) \qquad \boldsymbol{u} = \boldsymbol{B}\Phi(q)$$

$$\boldsymbol{c}_i = \boldsymbol{C}\Phi(x_i)$$

$$p_i = \text{softmax}(\boldsymbol{u}^\top \boldsymbol{m}_i)$$

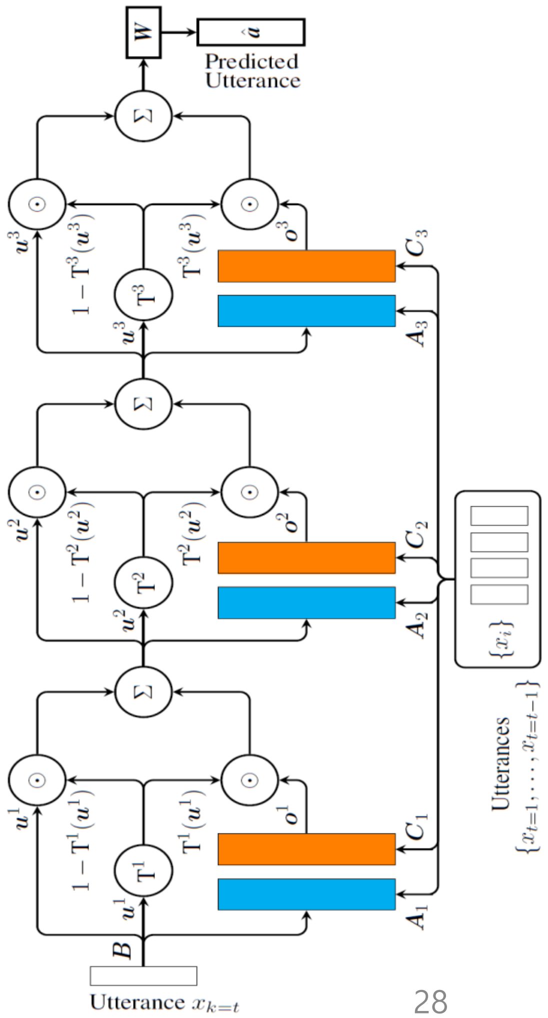
$$\boldsymbol{o} = \sum_i p_i \boldsymbol{c}_i$$

$$\mathrm{T}^k(\boldsymbol{u}^k) = \sigma(\boldsymbol{W}_T^k \boldsymbol{u}^k + \boldsymbol{b}_T^k)$$

$$\boldsymbol{u}^{k+1} = \boldsymbol{o}^k \odot \mathrm{T}^k(\boldsymbol{u}^k) + \boldsymbol{u}^k \odot (1 - \mathrm{T}^k(\boldsymbol{u}^k))$$

gated controller update

$$\hat{\boldsymbol{a}} = \text{softmax}(\boldsymbol{u}^\top \boldsymbol{W}' \Phi(\boldsymbol{y}_1), \ldots, \boldsymbol{u}^\top \boldsymbol{W}' \Phi(\boldsymbol{y}_{|C|}))$$



Properties

- End-to-End memory access regulation
- Close to Highway Network and Residual Network

[19] **Gated End-to-End Memory Network**, Liu and Perez, EACL 2017

20 bAbi tasks: Benchmark results

	Baseline			MemN2N								
Task	Strongly Supervised MemNN [22]	LSTM [22]	MemNN WSH	BoW	PE	PE LS	PE LS RN	1 hop PE LS joint	2 hops PE LS joint	3 hops PE LS joint	PE LS RN joint	PE LS LW joint
1: 1 supporting fact	0.0	50.0	0.1	0.6	0.1	0.2	0.0	0.8	0.0	0.1	0.0	0.1
2: 2 supporting facts	0.0	80.0	42.8	17.6	21.6	12.8	8.3	62.0	15.6	14.0	11.4	18.8
3: 3 supporting facts	0.0	80.0	76.4	71.0	64.2	58.8	40.3	76.9	31.6	33.1	21.9	31.7
4: 2 argument relations	0.0	39.0	40.3	32.0	3.8	11.6	2.8	22.8	2.2	5.7	13.4	17.5
5: 3 argument relations	2.0	30.0	16.3	18.3	14.1	15.7	13.1	11.0	13.4	14.8	14.4	12.9
6: yes/no questions	0.0	52.0	51.0	8.7	7.9	8.7	7.6	7.2	2.3	3.3	2.8	2.0
7: counting	15.0	51.0	36.1	23.5	21.6	20.3	17.3	15.9	25.4	17.9	18.3	10.1
8: lists/sets	9.0	55.0	37.8	11.4	12.6	12.7	10.0	13.2	11.7	10.1	9.3	6.1
9: simple negation	0.0	36.0	35.9	21.1	23.3	17.0	13.2	5.1	2.0	3.1	1.9	1.5
10: indefinite knowledge	2.0	56.0	68.7	22.8	17.4	18.6	15.1	10.6	5.0	6.6	6.5	2.6
11: basic coreference	0.0	38.0	30.0	4.1	4.3	0.0	0.9	8.4	1.2	0.9	0.3	3.3
12: conjunction	0.0	26.0	10.1	0.3	0.3	0.1	0.2	0.4	0.0	0.3	0.1	0.0
13: compound coreference	0.0	6.0	19.7	10.5	9.9	0.3	0.4	6.3	0.2	1.4	0.2	0.5
14: time reasoning	1.0	73.0	18.3	1.3	1.8	2.0	1.7	36.9	8.1	8.2	6.9	2.0
15: basic deduction	0.0	79.0	64.8	24.3	0.0	0.0	0.0	46.4	0.5	0.0	0.0	1.8
16: basic induction	0.0	77.0	50.5	52.0	52.1	1.6	1.3	47.4	51.3	3.5	2.7	51.0
17: positional reasoning	35.0	49.0	50.9	45.4	50.1	49.0	51.0	44.4	41.2	44.5	40.4	42.6
18: size reasoning	5.0	48.0	51.3	48.1	13.6	10.1	11.1	9.6	10.3	9.2	9.4	9.2
19: path finding	64.0	92.0	100.0	89.7	87.4	85.6	82.8	90.7	89.9	90.2	88.0	90.6
20: agent's motivation	0.0	9.0	3.6	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2
Mean error (%)	6.7	51.3	40.2	25.1	20.3	16.3	13.9	25.8	15.6	13.3	12.4	15.2
Failed tasks (err. > 5%)	4	20	18	15	13	12	11	17	11	11	11	10
On 10k training data												
Mean error (%)	3.2	36.4	39.2	15.4	9.4	7.2	6.6	24.5	10.9	7.9	7.5	11.0
Failed tasks (err. > 5%)	2	16	17	9	6	4	4	16	7	6	6	6

Table 1: Test error rates (%) on the 20 QA tasks for models using 1k training examples (mean test errors for 10k training examples are shown at the bottom). Key: BoW = bag-of-words representation; PE = position encoding representation; LS = linear start training; RN = random injection of time index noise; LW = RNN-style layer-wise weight tying (if not stated, adjacent weight tying is used); joint = joint training on all tasks (as opposed to per-task training).

Dialog State tracking

Examples & Definition

	Utterance	Food
S	Hello, How may I help you?	
U	I need a Persian restaurant in the south part of town.	0.2 Persian
S	What kind of food would you like?	
U	Persian.	0.8 Persian
S	I'm sorry but there is no restaurant serving persian food	
U	How about Portuguese food?	0.4 Persian 0.6 Portuguese
S	Are you looking for Portuguese food?	
U	Yes.	0.1 Persian 0.9 Portuguese
S	Nandos is a nice place in the south of town serving tasty Portuguese food.	

Slot	User may give as a constraint?
area	Yes, 15 possible values
children allowed	Yes, 2 possible values
food	Yes, 28 possible values
has internet	Yes, 2 possible values
has tv	Yes, 2 possible values
name	Yes, 163 possible values
near	Yes, 52 possible values
pricerange	Yes, 4 possible values
type	Yes, 3 possible values (restaurant, pub, coffeeshop)
addr	No
phone	No
postcode	No
price	No

Informable slots in DSTC3 (Tourist Information Domain)

Slot	User may give as a constraint?
area	Yes, 5 possible values
food	Yes, 91 possible values
name	Yes, 113 possible values
pricerange	Yes, 3 possible values
addr	No
phone	No
postcode	No
signature	No

Informable slots in DSTC2 (Restaurant Information Domain)

[20] **The third Dialog State Tracking Challenge**, Henderson et al, 2016

Dialogue State Tracking

State of the art

Generative

- {Factorial} HMM
- Particle Filter

Discriminative

- Rule-based
- CRF/Max Entropy
- Deep Neural Network

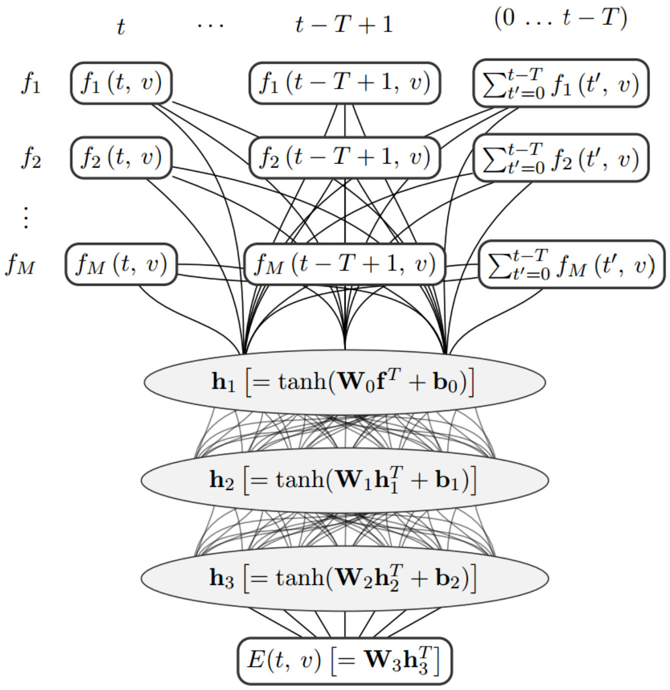


Figure 1: The Neural Network structure for computing $E(t, v) \in \mathbb{R}$ for each possible value v in the set $S_{t,s}$. The vector $\mathbf{f}$ is a concatenation of all the input nodes.

[21] **A generalized rule based tracker for dialogue state tracking**, Yu et al, 2014

[22] **Deep Neural Network Approach for the Dialog State Tracking Challenge**, Henderson et al, 2014

Dialog State Tracking

Open Challenges

1. Longer context
2. Looser supervision schema
3. Reasoning capability
4. Minimize intermediary reps
   - Fixed Ontology
   - Fixed KB

Good Morning, how can I help you

I need a car for March 10th to go to Paris

Ok, I'm checking this

and find me a cheap hotel for *the day after*

(-_-) "

[23] **Dialog State Tracking, a machine reading approach using deep memory networks**, Perez and Liu, EACL 2017

Dialog State Tracking

Machine reading approach

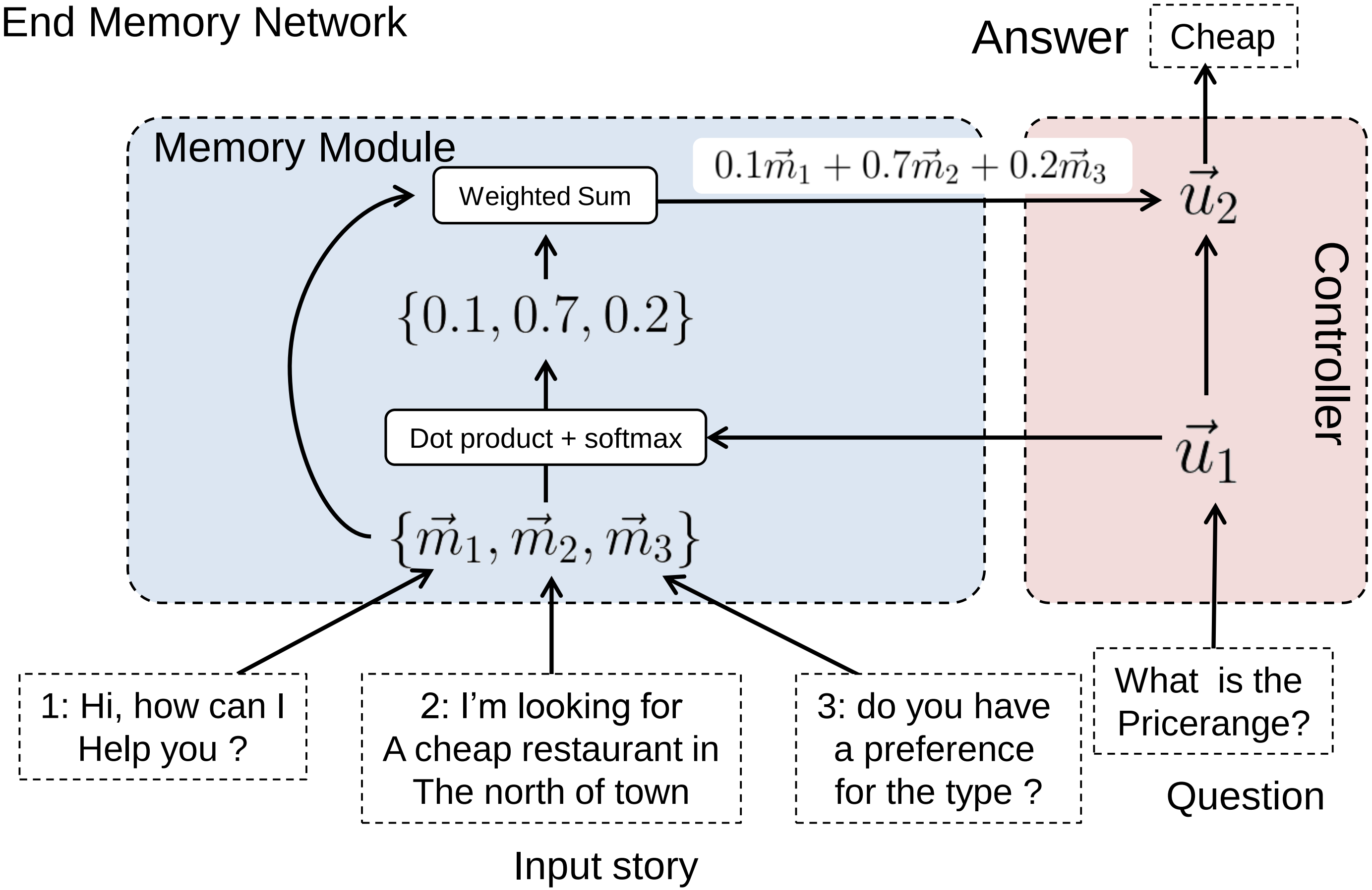
Index	Actor	Utterance
1	Cust	Im looking for a cheap restaurant in the west or east part of town.
2	Agent	Thanh Binh is a nice restaurant in the west of town in the cheap price range.
3	Cust	What is the address and post code.
4	Agent	Thanh Binh is on magdalene street city centre.
5	Cust	Thank you goodbye.
6	**Factoid Question** What is the pricerange ? Answer: {Cheap}	
7	**Yes/No Question** Is the Pricerange Expensive ? Answer: {No}	
8	**Indefinite Knowledge** Is the FoodType chinese ? Answer: {Maybe}	
8	**Listing task** What are the areas ? Answer: {West,East}	

Table 1. State tracking as machine reading task

[23] **Dialog State Tracking, a machine reading approach using deep memory networks**, Perez and Liu, EACL 2017

Dialog State tracking

with End-to-End Memory Network



[23] **Dialog State Tracking, a machine reading approach using deep memory networks**, Perez and Liu, EACL 2017

End-to-End Memory Network

Results on DSTC-2 – Goal Tracking and Reasoning

[24] Dialog State Tracking, a machine reading approach using deep memory networks, Perez et Liu, EACL 2017

Variable	**d**	**Yes-No**	**I.K.**	**Count.**	**List.**
Food	20	**0.85**	0.79	0.89	0.41
	40	0.83	**0.84**	0.88	**0.42**
	60	0.82	0.82	**0.90**	0.39
Area	20	0.86	0.83	0.94	**0.79**
	40	**0.90**	0.89	**0.96**	0.75
	60	0.88	**0.90**	0.95	0.78
PriceRange	20	**0.93**	**0.86**	**0.93**	**0.83**
	40	0.92	0.85	0.90	0.80
	60	0.91	0.85	0.91	0.81

Model	Area	Food	Price	Joint
RNN - no dict.	0.92	0.86	0.86	0.69
RNN + sem. dict.	0.91	0.86	0.93	0.73
NBT-DNN	0.90	0.84	0.94	0.72
NBT-CNN	0.90	0.83	0.93	0.72
MemN2N($d = 40$)	**0.89**	**0.88**	**0.95**	**0.74**

Dialog state tracking

Machine reading approach

On “one supporting fact” task (DSTC-2 dataset): 83% acc vs 79% for the sota.

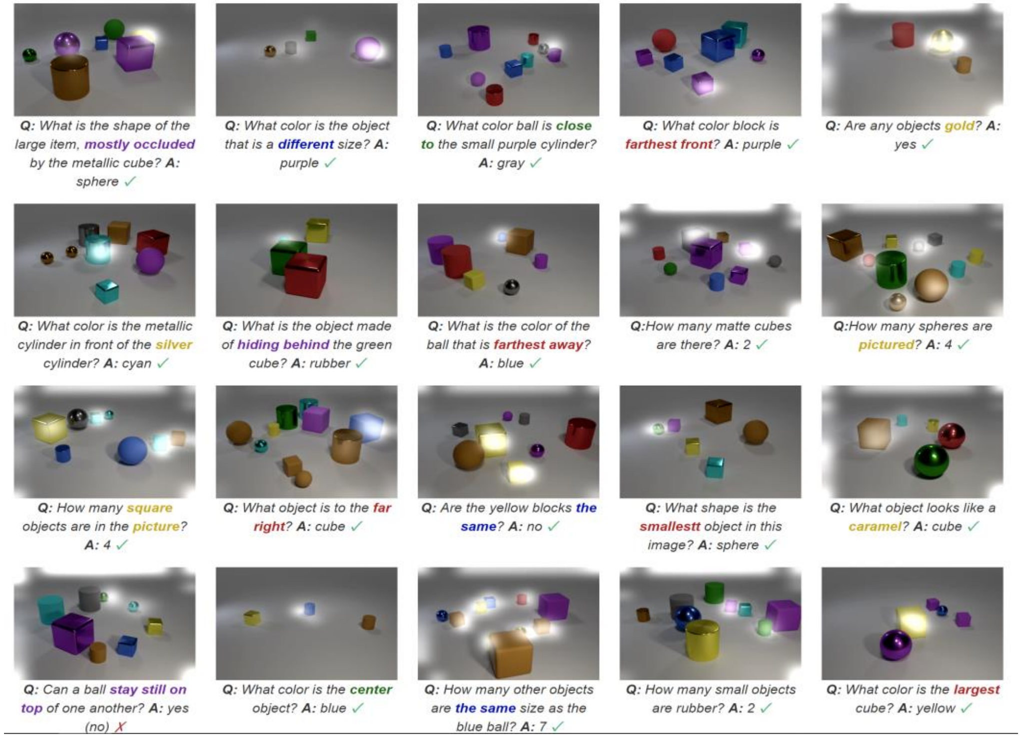
Table 11: Attention shifting example for the *PriceRange* slot from *DSTC2* dataset

Actor	Utterance	Hop 1	Hop 2	Hop 3	Hop 4	Hop 5
Cust	Im looking for a cheap restaurant that serves chinese food	0.00	0.14	0.01	0.00	0.00
Agent	What part of town do you have in mind	0.02	0.17	0.05	0.00	0.00
Cust	I dont care	0.00	0.00	0.14	0.00	0.00
Agent	Rice house serves chinese food in the cheap price range	0.00	0.02	0.03	0.98	1.00
Cust	What is the address and telephone number	0.57	0.07	0.15	0.00	0.00
Agent	Sure rice house is on mill road city centre	0.03	0.01	0.13	0.02	0.00
Cust	Phone number	0.00	0.01	0.03	0.00	0.00
Agent	The phone number of rice house is 765-239-09	0.37	0.58	0.45	0.00	0.00
Cust	Thank you good bye	0.00	0.00	0.00	0.00	0.00
What is the pricerange ? Answer: cheap						

[24] **Dialog State Tracking, a machine reading approach**, Perez and Liu, 2017

Review reading

Inspiration from relational visual question answering [Johnson et al, 2017]



p.s. Here are some more examples of the model's predictions. See how the model correctly handle questions that involve **obstructions**, **object uniqueness**, **relative distances**, **superlatives**, **varied vocabulary**.

Review reading

ReviewQA: a relational aspect-based opinion reading dataset

Hotel: BEST WESTERN Corona
Title: Convenient Location. Helpful Staff.
Overall rating: ★★★★☆

Comment: I just needed a place to sleep and this place was ideally located for my meetings. Plimlico tube is only a few minutes walk. Room was small but clean. Staff very helpful. Breakfast OK.

Ratings

Service	★★★★☆	Location	★★★★★
Rooms	★★★☆☆	Cleanliness	★★★★☆

Task	**Natural Language Questions**	
5	What is the rating of service?	3
3	Is the client satisfied with the location?	Yes
7	Does the customer prefer the service or the room?	Service

	# documents	# queries
Train	90.000	528.665
Test	10.000	58.827
Total	100.000	587.492

Task id	Description/Comment	Example	Expected answer
1	**Detection of an aspect in a review.**	Is sleep quality mentioned in this review?	Yes/No
2	**Prediction of the customer general satisfaction.**	Is the client satisfy by this hotel?	Yes/No
3	**Prediction of the global trend of an aspect in a given review.**	Is the client satisfied with the cleanliness of the hotel?	Yes/No
4	**Prediction of whether the rating of a given aspect is above or under a given value.**	Is the rating of location under 4?	Yes/No
5	**Prediction of the exact rating of an aspect in a review.**	What is the rating of the aspect Value in this review?	A rating between 1 and 5
6	**Prediction of the list of all the positive/negative aspects mentioned in the review.**	Can you give me a list of all the positive aspects in this review?	a list of aspects
7.0	**Comparison between aspects.**	Is the sleep quality better than the service in this hotel?	Yes/No
7.1		Which one of these two aspects, service, location has the best rating?	an aspect
8	**Prediction of the strengths and weaknesses in a review.**	What is the best aspect rated in this comment?	an aspect

[25] **ReviewQA: a relational aspect-based opinion reading dataset**, Grail and Perez, 2018

Fact checking

- Given a claim, retrieve evidence documents for and against it
- Given evidence documents, find relevant paragraphs and sentences in it
- For claim and each evidence paragraph and sentence: detect stance of paragraph sentence towards a claim/target

Stance detection:
Tweet: Be prepared - if we continue the policies of the liberal left, we will be #Greece
Target: Donald Trump
Label: favor

Fake news detection:
Document: Dino Ferrari hooked the whopper wels catfish, (...), which could be the biggest in the world.
Headline: Fisherman lands 19 STONE catfish which could be the biggest in the world to be hooked
Label: agree

Natural language inference:
Premise: Fun for only children
Hypothesis: Fun for adults and children
Label: contradiction

# Headline-body pairs			49972
# Headlines			1648
# Bodies			1683
# Bodies in test set			169
# Headline-body pairs in test set			5025
Average # tokens of headline			12.6
Average # tokens of body			427.5
Unrelated	*Discuss*	*Agree*	*Disagree*
73.1%	17.8%	7.4%	1.7%

Table 1: Statistics of *FNC1* dataset

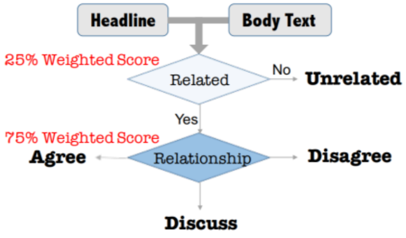


Figure 2: Score Metric for *FNC1*

[26] http://www.fakenewschallenge.org/ , 2017

Open Questions

Multi-document Open-Domain Question answering

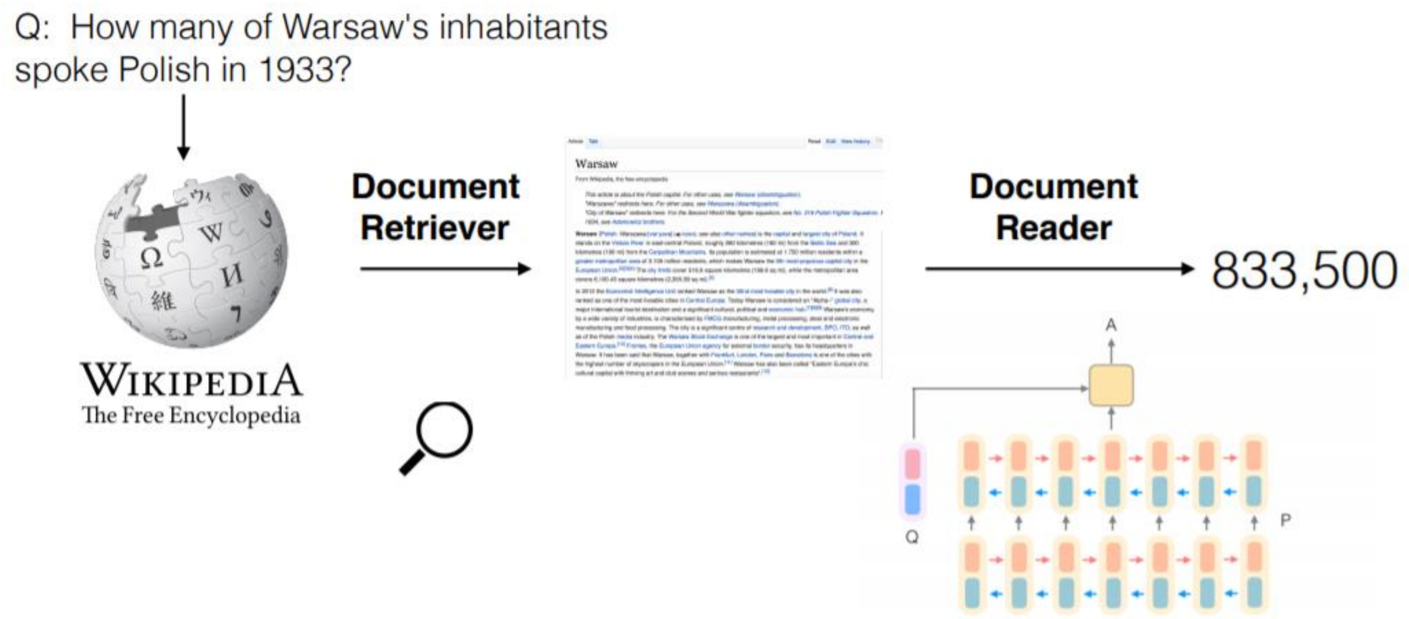


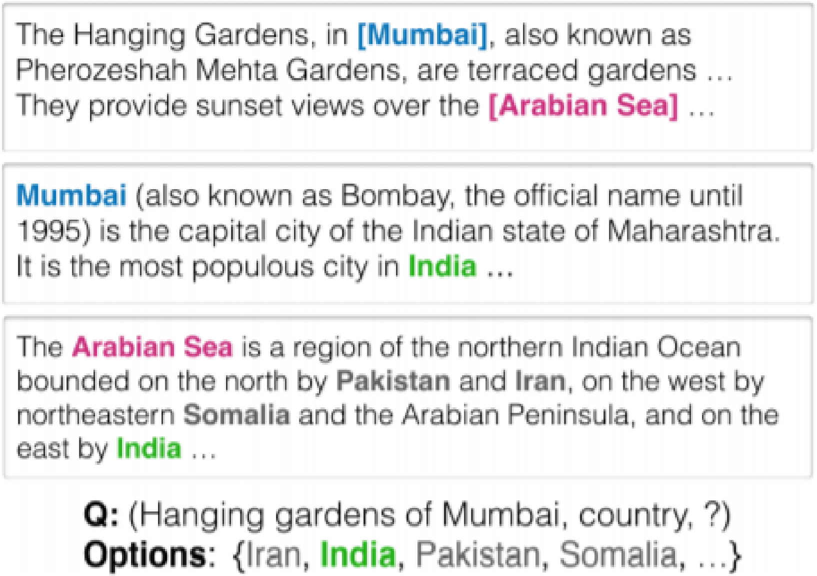
Figure 1: An overview of our question answering system DrQA.

[27] **Reading Wikipedia to Answer Open-Domain Questions**, Chen et al, 2017

Open Questions

Multi document reasoning

- Most Reading Comprehension methods limit themselves to queries which can be answered using a single sentence, paragraph, or document.
- Enabling models to combine disjoint pieces of textual evidence would extend the scope of machine comprehension
- Text understanding across multiple documents and to investigate the limits of existing methods.
- Toward ensemblist operations (union, intersection, selection …)

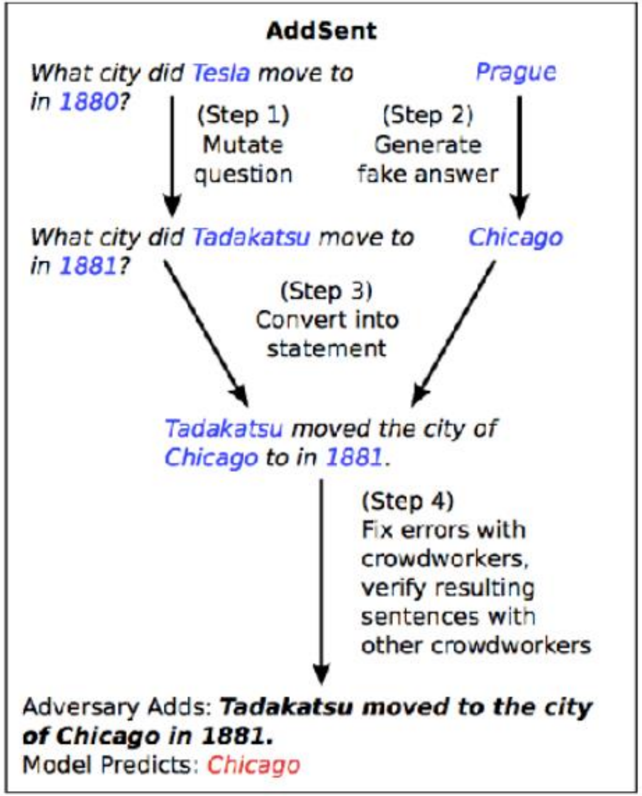


[28] **Constructing Datasets for Multi-hop Reading Comprehension Across Documents**, Riedel et al, 2017

Open Questions

Adversarial Examples

- Add a sentence or word string specifically designed to distract the model
- Drops accuracy of state-of-the-art models from 81% to 46% of Exact Match accuracy
- Current issue of deep models, already observed on image tasks



[29] **Adversarial Examples for Evaluating Reading Comprehension Systems**, Liang et al, 2017

Conclusions

- Machine reading paradigm, a next step toward natural language comprehension
- Promising results are already available
- Deep learning is (currently) a major enabler of this recent development
- Machine reading is a playground for (deep) machine learning research
- Very active community (Datasets, papers and codes)
- A lot of challenges with numerous possible impacts

Thank you

Organizing DSCT-7 workshop @AAAI'18

Naver Labs is Hiring !

europe.naverlabs.com