Towards a Unified Model for QA & Reasoning

@ KAIST

March 28, 2019

Minjoon Seo

What is Question Answering?

What is water consisted of?

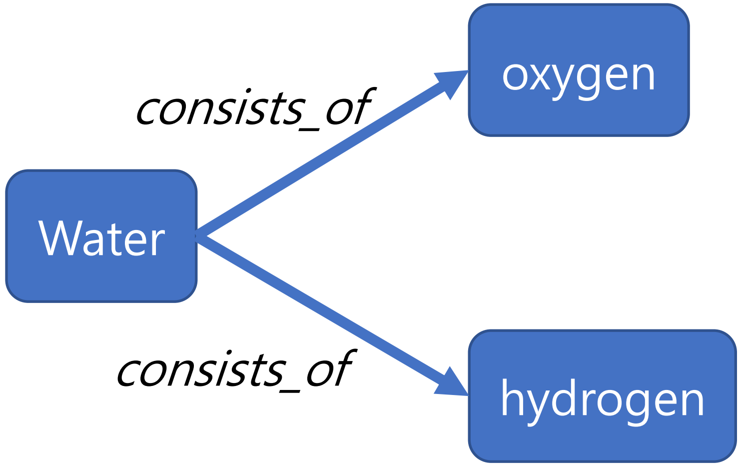
Hydrogen and oxygen!

Parsing and *Reading*

Parsing: answer from *structured* data

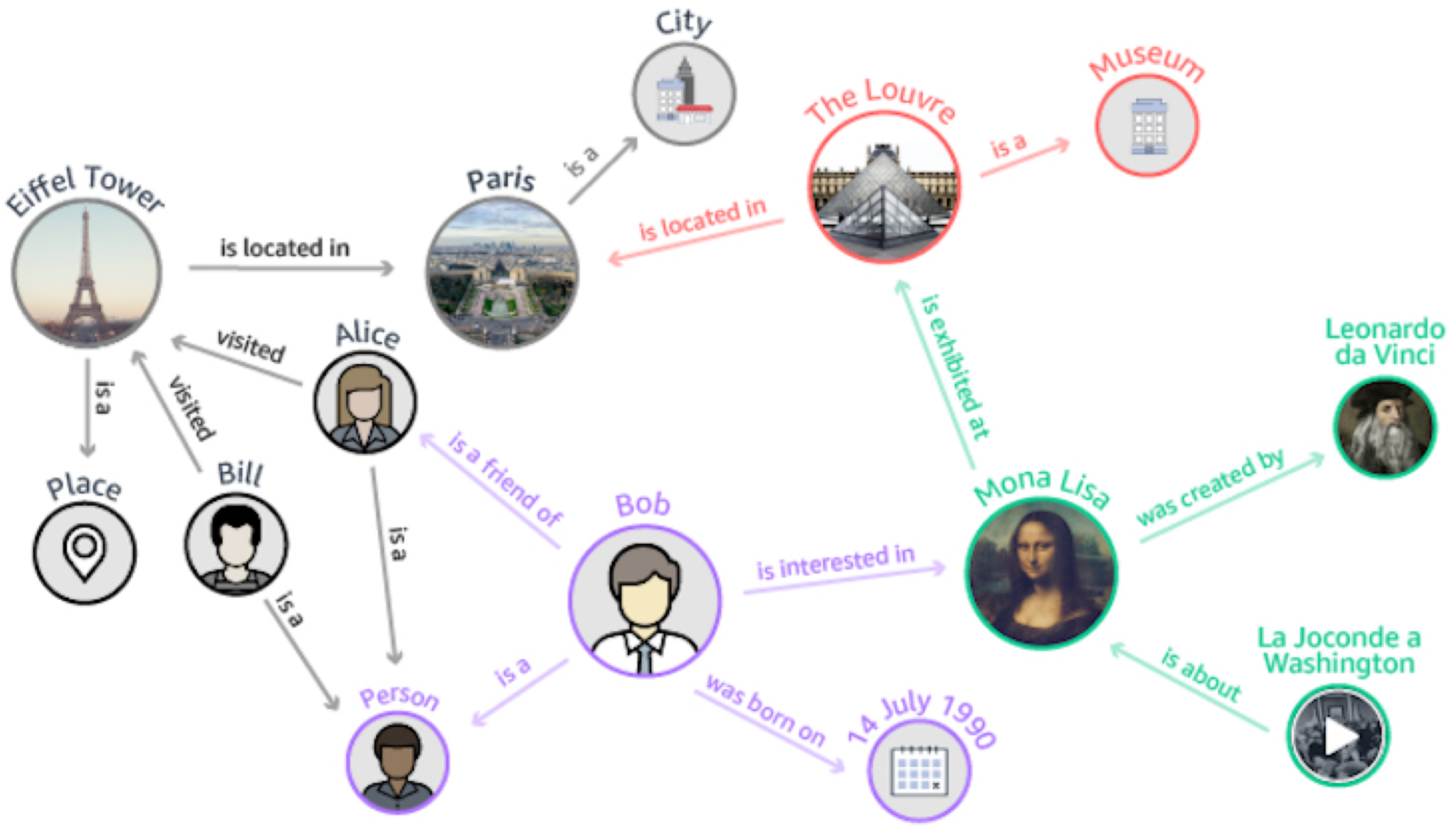
`consists_of(water, ?)`

What is water consisted of?



Hydrogen and oxygen!

Knowledge Graph



Amazon Knowledge Graph

Tables

Year	Competition	Venue	Position	Event	Notes
Representing Poland					
2001	World Youth Championships	Debrecen, Hungary	2nd	400 m	47.12
			1st	Medley relay	1:50.46
	European Junior Championships	Grosseto, Italy	1st	4x400 m relay	3:06.12
2002	World Junior Championships	Kingston, Jamaica	4th	4×400m relay	3:06.25
2003	European Junior Championships	Tampere, Finland	3rd	400 m	46.69
			2nd	4x400 m relay	3:08.62
2005	European U23 Championships	Erfurt, Germany	11th (sf)	400 m	46.62
			1st	4x400 m relay	3:04.41
	Universiade	Izmir, Turkey	7th	400 m	46.89
			1st	4x400 m relay	3:02.57
2006	World Indoor Championships	Moscow, Russia	2nd (h)	4x400 m relay	3:06.10
	European Championships	Gothenburg, Sweden	3rd	4x400 m relay	3:01.73
2007	European Indoor Championships	Birmingham, United Kingdom	3rd	4x400 m relay	3:08.14
	Universiade	Bangkok, Thailand	7th	400 m	46.85
			1st	4x400 m relay	3:02.05
2008	World Indoor Championships	Valencia, Spain	4th	4x400 m relay	3:08.76
	Olympic Games	Beijing, China	7th	4x400 m relay	3:00.32
2009	Universiade	Belgrade, Serbia	2nd	4x400 m relay	3:05.69

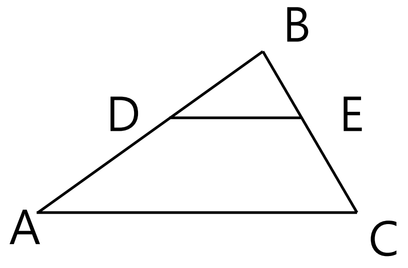
In what city did Piotr's last 1st place finish occur?

Pasupat & Liang, 2015

(First-Order) Logic

Text Input

In triangle ABC, line DE is parallel wi
th line AC, DB equals 4, AD is 8, and
DE is 5. Find AC.
(a) 9 (b) 10 (c) 12.5 (d) 15 (e) 1
7



Logical form

IsTriangle(ABC) ∧ Parallel(AC, DE) ∧

Equals(LengthOf(DB), 4) ∧ Equals(LengthOf(AD), 8) ∧ Equals(LengthOf(DE), 5) ∧ Find(LengthOf(AC))

Reading: answer from *unstructured* data

What is water consisted of?

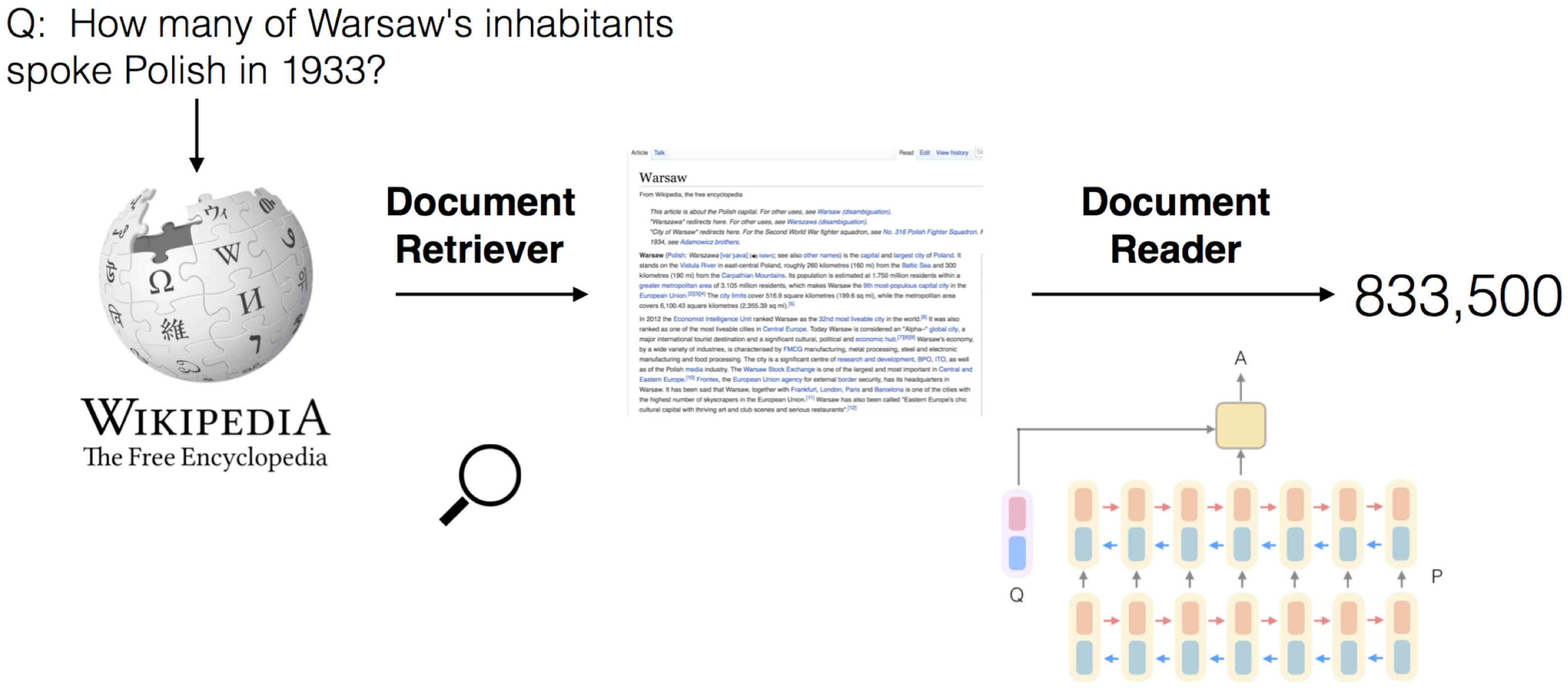
Nature

Water is a chemical substance that is composed of **hydrogen and oxygen** and is vital for all known forms of life. In typical usage, "water" refers only to its liquid form or state, but the substance also has a solid state, ice, and a gaseous state, water vapor, or steam.

Hydrogen and oxygen!

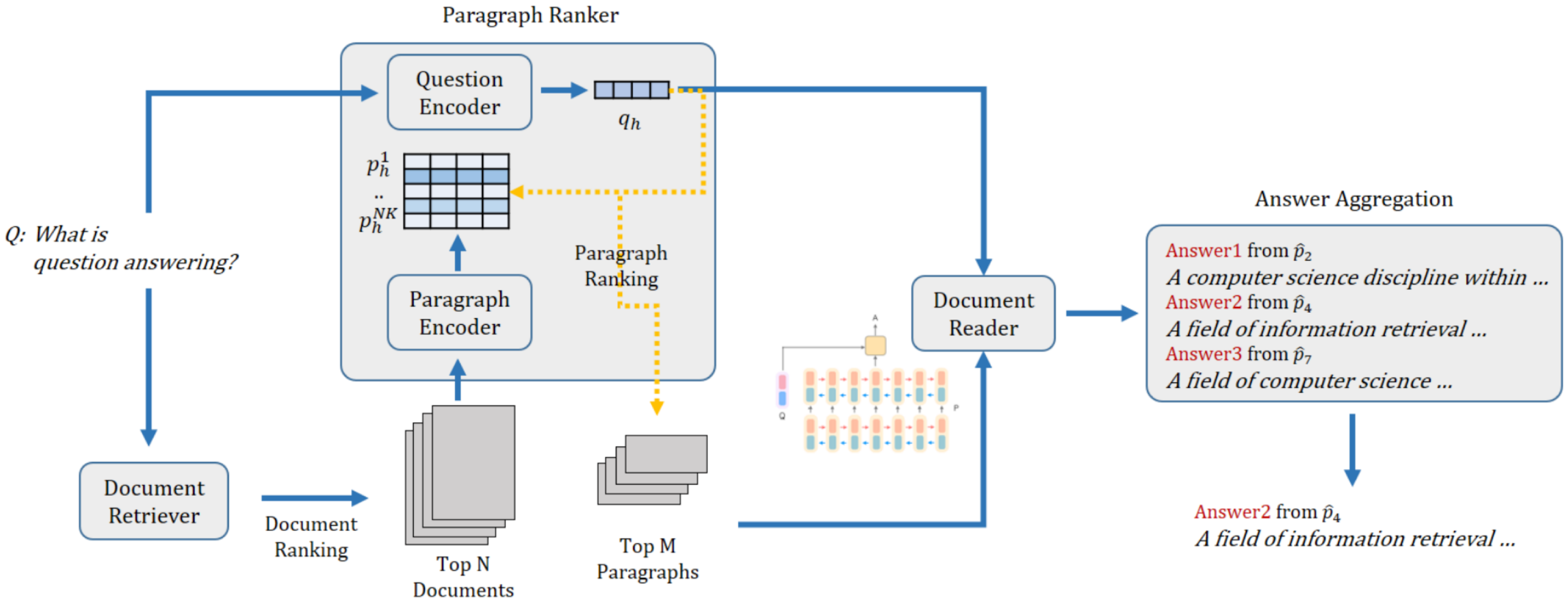
Search & Read

Open-domain QA
SQuAD, TREC, WebQuestions, WikiMovies



Chen et al., 2017

Search & Rank & Read

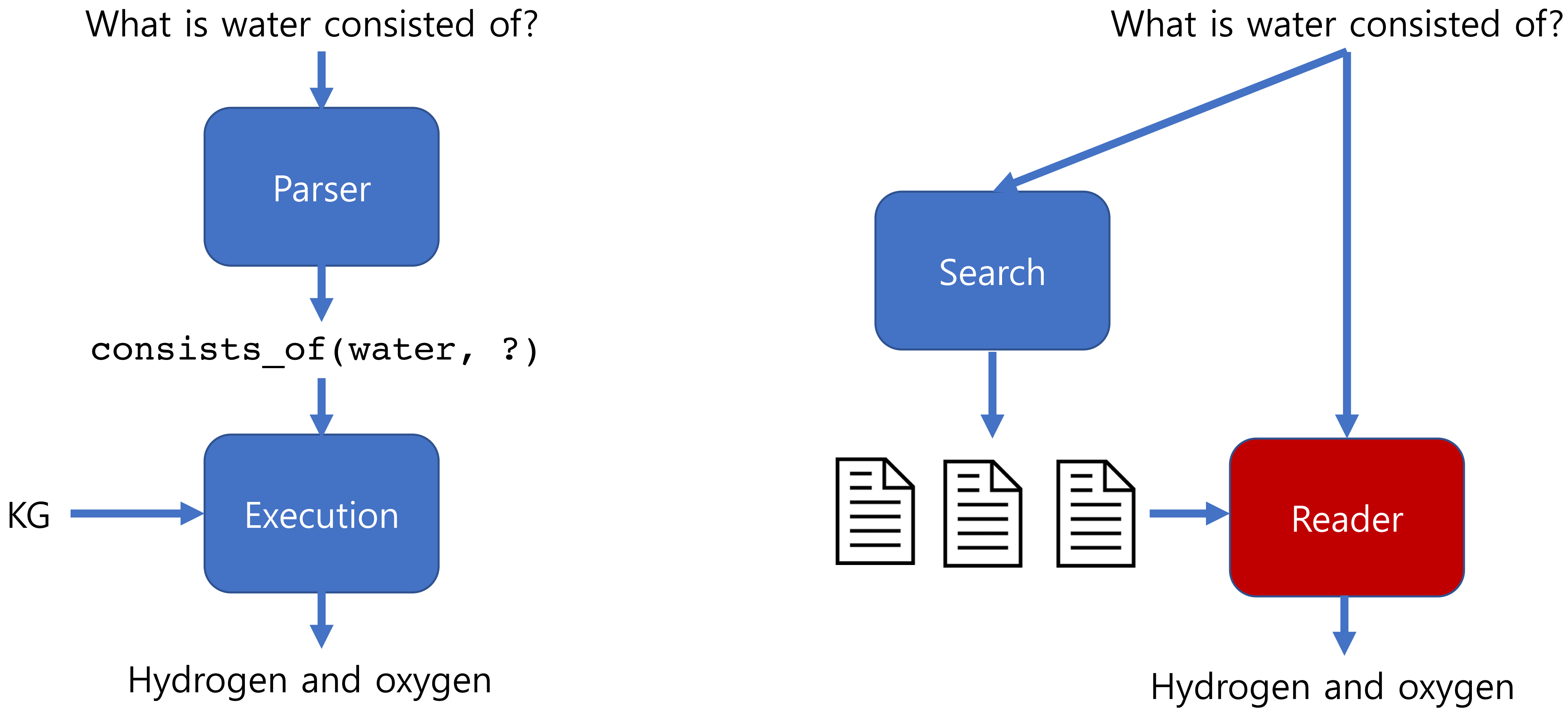


Lee et al., 2018

Parsing vs Reading

- Speed
- Domain
- Complexity

Parsing vs Reading: **Speed**



Parsing vs Reading: **Domain**

Q: "How do you become a great researcher?"

- *Parsing* is ontology- and KG-dependent
  - Designing comprehensive ontology is difficult
  - Constructing comprehensive KG is expensive

- *Reading* is ontology-free, open-domain

In an interview with Association for Psychological Science, Levine believes the key to being a great researcher is **having passion for what you do research in and working on questions that you are truly curious about**.

Parsing vs Reading: **Complexity**

- Multi-hop reasoning
- "What are the atomic numbers of the elements in water?"

Parsing

```
atomic_number(A) s.t
consists_of(water, A)
```

Reading

Need text that explicitly contains the information.

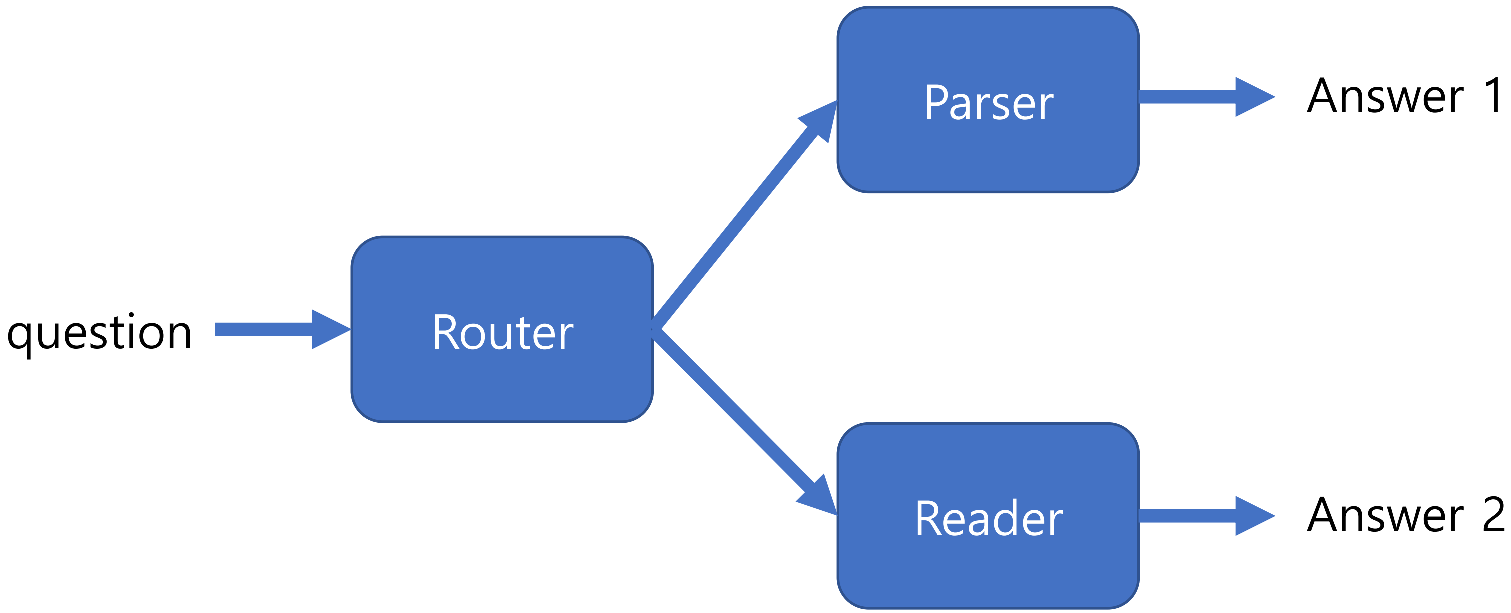
Parsing vs Reading

	Parsing	Reading
Speed	Fast	Slow
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

tl;dr:

It is clear that we need to consider **both**

Solution #1: a pipeline



Solution #2: a unified model?

Today’s Talk

- Intro: About Question Answering & Reasoning
- **Parsing**: SOTA on WikiSQL
- **Reading**: Real-time Open-domain Question Answering
- Towards a Unified Model

Today's Talk

- Intro: About Question Answering & Reasoning
- **Parsing**: SOTA on WikiSQL
- **Reading**: Real-time Open-domain Question Answering
- Towards a Unified Model

WikiSQL: NL2SQL Task

- Natural language to SQL

Table: CFLDraft

Pick #	CFL Team	Player	Position	College
27	Hamilton Tiger-Cats	Connor Healy	DB	Wilfrid Laurier
28	Calgary Stampeders	Anthony Forgone	OL	York
29	Ottawa Renegades	L.P. Ladouceur	DT	California
30	Toronto Argonauts	Frank Hoffman	DL	York
...	...	...	...	...

Question:

How many CFL teams are from York College?

SQL:

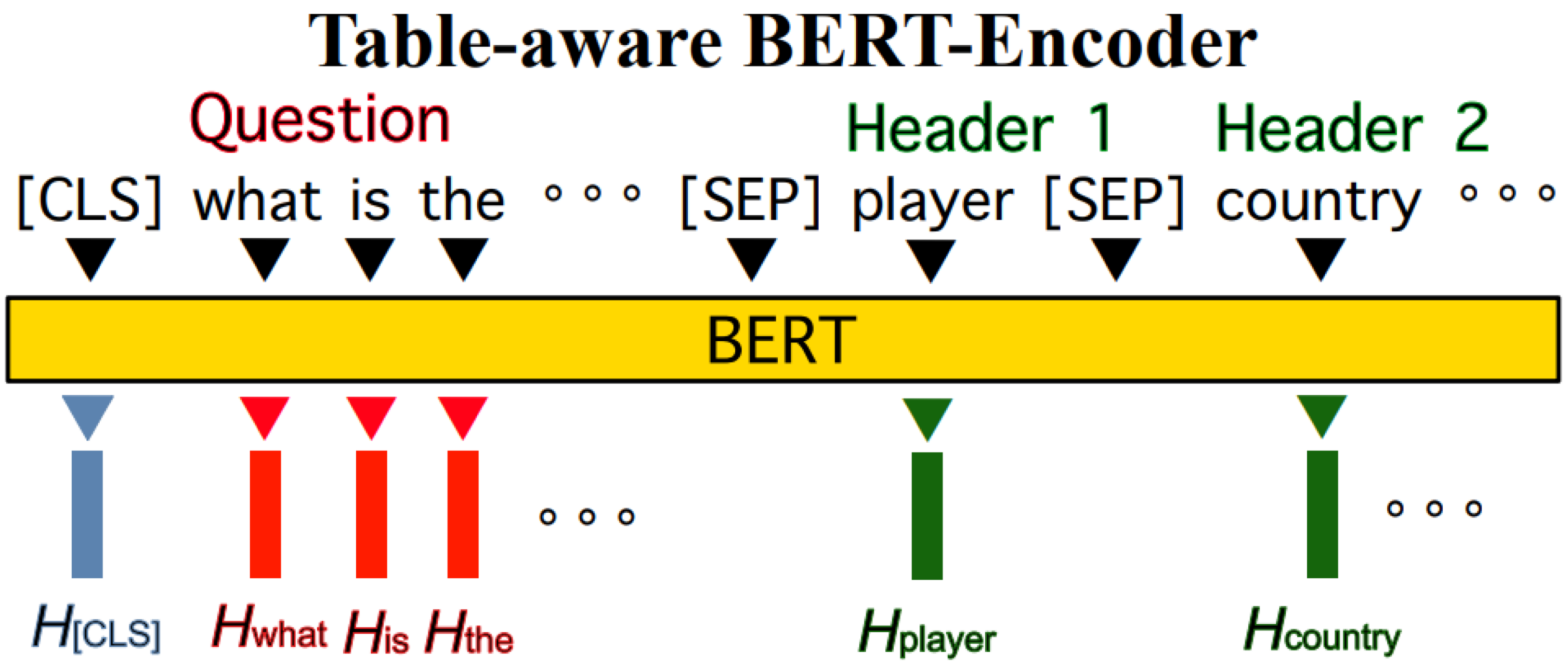
```
SELECT COUNT CFL Team FROM
CFLDraft WHERE College = "York"
```

Result:

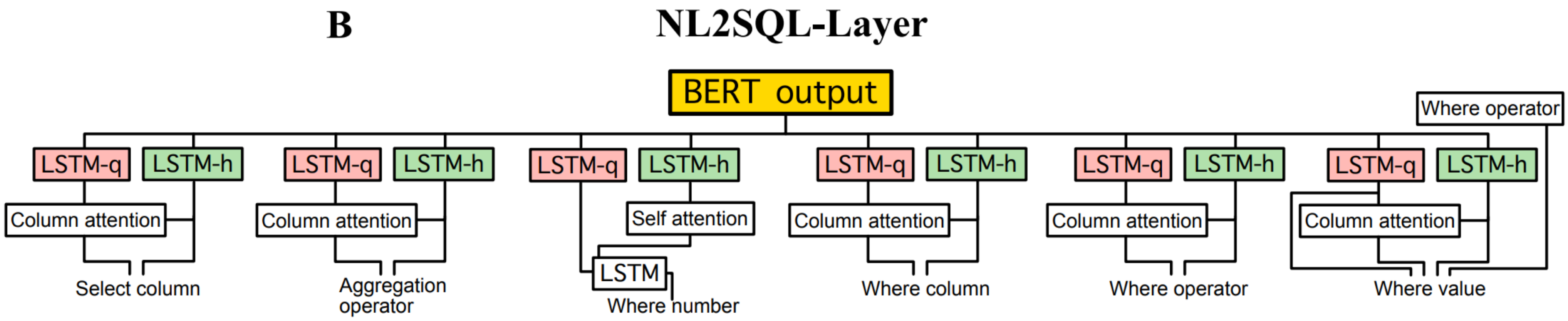
2

Zhong et al., 2017

SQLova (Hwang et al., 2019)



SQLova (Hwang et al., 2019)



SQLova Results

Model	Dev LF (%)	Dev X (%)	Test LF (%)	Test X (%)
Baseline (Zhong et al., 2017)	23.3	37.0	23.4	35.9
Seq2SQL (Zhong et al., 2017)	49.5	60.8	48.3	59.4
SQLNet (Xu et al., 2017)	63.2	69.8	61.3	68.0
PT-MAML (Huang et al., 2018)	63.1	68.3	62.8	68.0
TypeSQL (Yu et al., 2018)	68.0	74.5	66.7	73.5
Coarse2Fine (Dong and Lapata, 2018)	72.5	79.0	71.7	78.5
MQAN (McCann et al., 2018)	76.1	82.0	75.4	81.4
Annotated Seq2seq (Wang et al., 2018b) [1]	72.1	82.1	72.1	82.2
IncSQL (Shi et al., 2018) [1]	49.9	84.0	49.9	83.7
BERT-TO-SEQUENCE (ours)	57.3	-	56.4	-
SHALLOW-LAYER (ours)	**81.5** (+5.4)	**87.4** (+3.2)	**80.9** (+5.5)	**86.8** (+3.1)
NL2SQL-LAYER (SQLOVA, ours)	**81.6** (+5.5)	**87.2** (+3.2)	**80.7** (+5.3)	**86.2** (+2.5)
PointSQL-EG (Wang et al., 2018a) [1,2]	67.5	78.4	67.9	78.3
Coarse2Fine-EG (Wang et al., 2018a) [1,2]	76.0	84.0	75.4	83.8
IncSQL-EG (Shi et al., 2018) [1,2]	51.3	87.2	51.1	87.1
SHALLOW-LAYER-EG (ours) [2]	**82.3** (+6.3)	**88.1** (+0.9)	**81.8** (+6.4)	**87.5** (+0.4)
NL2SQL-LAYER-EG (SQLOVA-EG, ours) [2]	**84.2** (+8.2)	**90.2** (+3.0)	**83.6** (+8.2)	**89.6** (+2.5)
Human performance [3]	-	-	-	88.3

And it's near upper bound…

NL	What is the number of the player who went to Southern University?
TBL	"Player", "No.(s)", "Height in Ft.", "Position", "Years for Rockets", "School/Club Team/Country"
SQL (T)	`SELECT` (No.(s)) `FROM` 1-11734041-9 `WHERE` School/Club Team/Country = Southern University
SQL (P)	`SELECT` count(No.(s)) `FROM` 1-11734041-9 `WHERE` School/Club Team/Country = southern university
ANS (T)	6
ANS (P)	1
ERROR	Qestion (IV)

Parsing vs Reading

	Parsing	Reading
Speed	Fast	Slow
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

Fundamental limitation of **parsing**:
it is ontology- and KG-dependent *by definition*

Today’s Talk

- Intro: About Question Answering & Reasoning
- **Parsing**: SOTA on WikiSQL
- **Reading**: Real-time Open-domain Question Answering
- Towards a Unified Model

Parsing vs Reading

	Parsing	Reading
Speed	Fast	Slow
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

Reader: answer from *unstructured* data

Second Epistle to the Corinthians The Second Epistle to the Corinthians, often referred to as Second Corinthians (and written as 2 Corinthians), is the eighth book of the New Testament of the Bible. Paul the Apostle and “Timothy our brother” wrote this epistle to “the church of God which is at Corinth, with all the saints which are in all Achaia”.

Who wrote second Corinthians?

Rajpurkar et al., 2016

Reader: answer from *unstructured* data

Second Epistle to the Corinthians The Second Epistle to the Corinthians, often referred to as Second Corinthians (and written as 2 Corinthians), is the eighth book of the New Testament of the Bible. **Paul the Apostle and "Timothy our brother"** wrote this epistle to "the church of God which is at Corinth, with all the saints which are in all Achaia".

Who wrote second Corinthians?

Rajpurkar et al., 2016

SQuAD (Rajpurkar et al., 2016)

- 100,000+ paragraph-question-answer pairs
- First phrase-level answers
- First massive **and** manually annotated QA dataset
- Easy to play with, but has a direct useful application

63 Feb 19, 2017	Attentive CNN context with LSTM *NLPR, CASIA*	63.306	73.463
63 Sep 21, 2017	OTF dict+spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286	64.083	73.056
64 Sep 21, 2017	OTF spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286	62.897	72.016
64 Nov 02, 2016	Fine-Grained Gating *Carnegie Mellon University* https://arxiv.org/abs/1611.01724	62.446	73.327
64 Sep 21, 2017	OTF spelling+lemma (single) *University of Montreal* https://arxiv.org/abs/1706.00286	62.604	71.968
65 Sep 28, 2016	Dynamic Chunk Reader *IBM* https://arxiv.org/abs/1610.09996	62.499	70.956
66 Aug 27, 2016	Match-LSTM with Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905	60.474	70.695
67 Sep 11, 2018	Unnamed submission by Will_Wu	59.058	69.436
68 Jan 05, 2018	PivRet (single model) *anonymous*	58.764	69.276
69 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) *Singapore Management University* https://arxiv.org/abs/1608.07905	54.505	67.748

100+ models in two years!

“중국 인공지능 전쟁에서 미국에 승리”…인간보다 독해 능력 뛰어난 인공지능 첫 개발

주영재 기자 jyj@kyunghyang.com

입력 : 2018.01.16 15:54:00 | 수정 : 2018.01.16 16:39:37

중국이 인간보다 독해 능력이 뛰어난 인공지능을 개발한 첫 국가가 됐다. 중국이 2030년 미국을 제치고 인공지능 경쟁에서 1위를 차지하게 될 것이라는 전망에 힘이 실린다.

블룸버그·파이낸셜타임스 등 외신은 15일(현지시간) 알리바바가 개발한 인공지능이 미국 스탠퍼드대가 주최한 인공지능 대회에서 82.44의 정확도로 인간(82.3)보다 뛰어난 독해능력을 보였다고 보도했다.

■"인공지능 경쟁에서 중국이 미국을 제치다"

이번 인공지능 대회는 10만개 이상의 질문에 정확한 답을 내야 하는 것으로 머신러닝의 수준을 측정하는 가장 권위있는 대회라는 평가를 받고 있다. 질문들은 "비가 왜 내리는가" "아마존 열대 우림은 얼마나 큰가" "니콜라 테슬라의 출신 국가는 어디인가" "슈퍼볼 50 하프 타임 쇼의 첫 공연을 어떤 그룹이 맡았나" 등이었다. 500개 이상의 위키피디아 문서를 바탕으로 한 이번 시험은 인공지능이 거대한 양의 정보를 처리해 질문에 정확한 답을 낼 수 있는지를 측정하는데 목적이 있었다.

5% higher than humans

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) *Google A.I.*	**87.433**	**93.160**
2 Oct 05, 2018	BERT (single model) *Google A.I.*	85.083	91.835
2 Sep 09, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.356	91.202
2 Sep 26, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.954	91.677
3 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	84.454	90.490
4 Jul 08, 2018	r-net (ensemble) *Microsoft Research Asia*	84.003	90.147

Great, but how fast is it?

Hmm... 1s per document?

Barack Obama

From Wikipedia, the free encyclopedia

"Barack" and "Obama" redirect here. For other uses, see Barack (disambiguation) and Obama (disambiguation).

Barack Obama

44th President of the United States

In office
January 20, 2009 - January 20, 2017

Vice President Joe Biden

Preceded by George W. Bush

Succeeded by Donald Trump

United States Senator from Illinois

Barack Hussein Obama II (/bəˈrɑː k huː ˈseɪn oʊˈbɑː mə/ (listen);[1] born August 4, 1961) is an American politician who served as the 44th President of the United States from 2009 to 2017. The first African American to assume the presidency, he was previously the junior United States Senator from Illinois from 2005 to 2008. He served in the Illinois State Senate from 1997 until 2004.

Obama was born in 1961 in Honolulu, Hawaii, two years after the territory was admitted to the Union as the 50th state. Raised largely in Hawaii, Obama also spent one year of his childhood in Washington State and four years in Indonesia. After graduating from Columbia University in New York City in 1983, he worked as a community organizer in Chicago. In 1988 Obama enrolled in Harvard Law School, where he was the first black president of the *Harvard Law Review*. After graduation, he became a civil rights attorney and professor, and taught constitutional law at the University of Chicago Law School from 1992 to 2004. Obama represented the 13th District for three terms in the Illinois Senate from 1997 to 2004, when he ran for the U.S. Senate. Obama received national attention in 2004 with his unexpected March primary win, his well-received July Democratic National Convention keynote address, and his landslide November election to the Senate. In 2008, Obama was nominated for president a year after his campaign began and after a close primary campaign against Hillary Clinton. He was elected over Republican John McCain and was inaugurated on January 20, 2009. Nine months later, Obama was named the 2009 Nobel Peace Prize laureate, accepting the award with the caveat that he felt there were others "far more deserving of this honor than I."

During his first two years in office, Obama signed many landmark bills into law. The main reforms were the Patient Protection and Affordable Care Act (often referred to as "Obamacare", shortened as the "Affordable Care Act"), the Dodd-Frank Wall Street Reform and Consumer Protection Act, and the Don't Ask, Don't Tell Repeal Act of 2010. The American Recovery and Reinvestment Act of 2009 and Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010 served as economic stimulus amidst the Great Recession. After a lengthy debate over the national debt limit, Obama signed the Budget Control and the American Taxpayer Relief Acts. In foreign policy, Obama increased U.S. troop levels in Afghanistan, reduced nuclear weapons with the United States-Russia New START treaty, and ended military involvement in the Iraq War. He ordered military involvement in Libya in opposition to Muammar Gaddafi; Gaddafi was killed by NATO-assisted forces, and he also ordered the military operation that resulted in the death of Osama bin Laden.

After winning re-election by defeating Republican opponent Mitt Romney, Obama was sworn in for a second term in 2013. During his second term, Obama promoted inclusiveness for LGBT Americans. His administration filed briefs that urged the Supreme Court to strike down same-sex marriage bans as

1s

WIKIPEDIA
La enciclopedia libre

5.6 Million Documents

3 Billion Words

Great, but how long does it take?

Hmm… 1s per document?

So… 6 days.

One week?

!#$@*%(@*@

Actually, I will just retrieve a few documents, and just **read** them!

WIKIPEDIA
La enciclopedia libre

두산백과

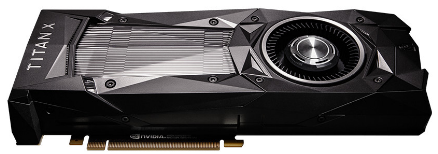
버락 오바마

[Barack Hussein Obama]

요약 미국의 정치가. 인권변호사 출신으로 일리노이주 상원의원(3선)을 거쳐 연방 상원의원을 지냈으며, 2008년 민주당 대통령 후보로 출마하여 공화당의 존 매케인 후보에 압승하고 제44대 미국 대통령에 당선됨으로써 미국 최초의 흑인(정확하게는 혼혈 흑인) 대통령이 되었다. 취임 후 핵무기 감축, 중동평화회담 재개 등에 힘써 2009년 노벨 평화상을 수상하였다.

출생-사망	1961.8.4 ~
본명	버락 후세인 오바마(Barack Hussein Obama)
국적	미국
활동분야	정치
출생지	미국 하와이주 호놀룰루
주요수상	노벨 평화상(2009)

1961년 8월 4일 미국 하와이주의 호놀룰루에서 케냐 출신의 하와이대학 유학생인 흑인 아버지와 미국인 백인 어머니 사이에서 태어났다. 2세 때 하와이대의 첫 아프리카 유학생이었던 아버지는 이혼하고 케냐로 돌아갔다. 1966년 재혼한 어머니를 따라 인도네시아로 이주하였다가 어머니가 다시 이혼하여 하와이로 돌아왔다. 다인종·다민족·다문화 가정에서 자라면서 술과 담배와 마약에도 손을 대는 등 불우한 청소년 시절을 보냈으나, 이러한 경험을 통하여 관용과 화합을 배우게 되었다.



1961

Chen et al., 2017

Still *slow*, and
error *propagates*…

?

WIKIPEDIA
La enciclopedia libre

1961

direct **and** ***fast*** **reader?**

WIKIPEDIA
La enciclopedia libre

5.6M documents

5.8 days

5,000,000x shorter

0.1s

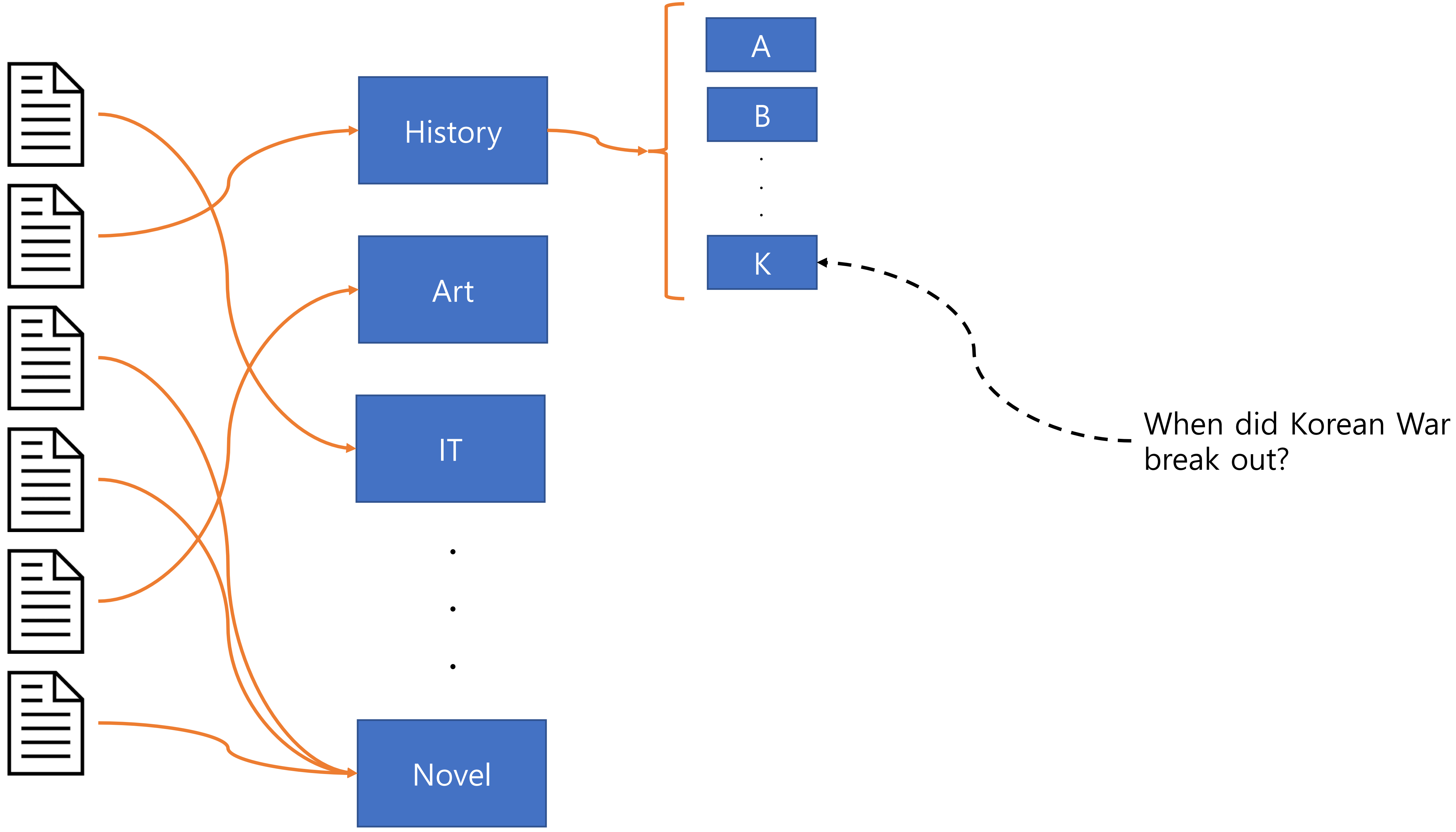
Titan Xp

4,000x slower

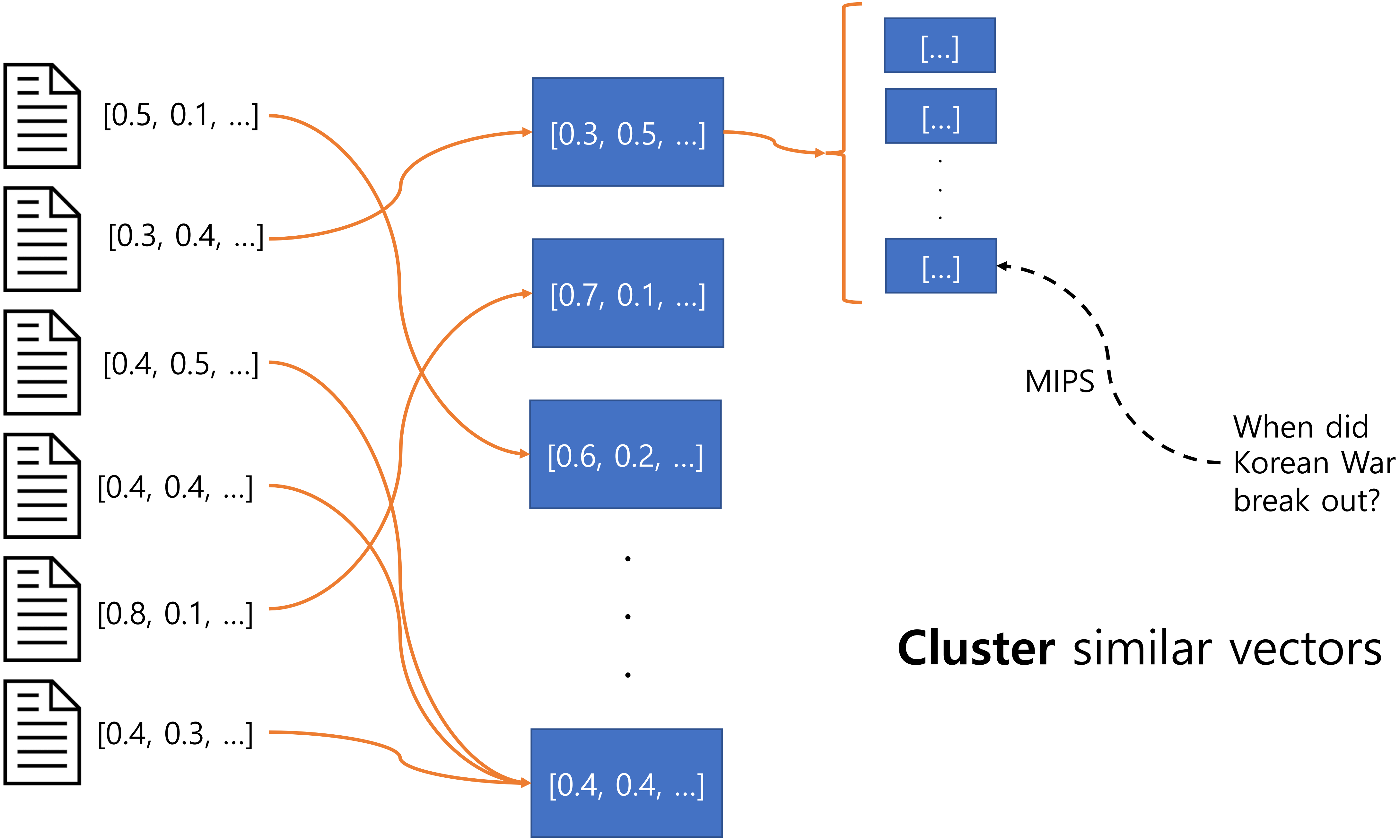
CPU

20 billion times faster?

This is a Green Factory Library.


History
Art
IT
Novel
A
B
K
When did Korean War
break out?

Precompute vectors.
Organize vectors.



[0.5, 0.1, ...]
[0.3, 0.4, ...]
[0.4, 0.5, ...]
[0.4, 0.4, ...]
[0.8, 0.1, ...]
[0.4, 0.3, ...]
[0.3, 0.5, ...]
[0.7, 0.1, ...]
[0.6, 0.2, ...]
[0.4, 0.4, ...]
[...]
[...]
[...]
MIPS
When did Korean War break out?
Cluster similar vectors

Kernel types

- **Symmetric**: proper metric functions (Nearest Neighbor Search)
  - L2
  - L1
  - Angular distance
- **Asymmetric**: inner product (MIPS)
  - Dot product (cosine distance)

$$O(kN)$$

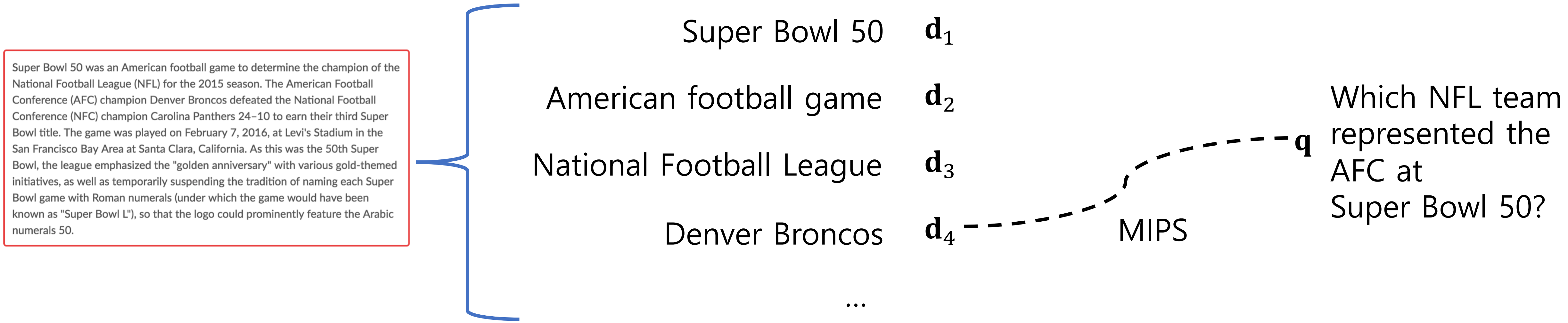
$$O(kN^{\rho}\log(N))$$

ρ= approximation factor (<1)

Shrivastava and Li, 2014

Sublinear-time approximation.
Very fast.

Document → Phrase?


Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.
Super Bowl 50 $\mathbf{d}_1$
American football game $\mathbf{d}_2$
National Football League $\mathbf{d}_3$
Denver Broncos $\mathbf{d}_4$
...
MIPS
$\mathbf{q}$
Which NFL team represented the AFC at Super Bowl 50?

Mathematically…

- document ***d*** 와 question ***q*** are given:

$$\hat{a} = \underset{a}{\operatorname{argmax}} P_\theta(a|q;d)$$

where

As-is: need to compute score for every new query

$$P_\theta(a|q;d) \propto \exp(F_\theta(a,q,d))$$

$$\Big\downarrow \ \text{Decomposition}$$

To-be: *H* can be *pre-computed* and *indexed*

$$P_\theta(a|q;d) \propto \exp(G_\theta(q) \cdot H_\theta(a,d))$$

But the decomposition is not easy.

A new research problem:
Phrase-Indexed Question Answering
(PIQA)

Seo et al., 2018

PIQA (Seo et al., 2018)

Constraint	Model	F1 (%)	EM (%)
PI	TF-IDF	15.0	3.9
	LSTM	57.2	46.8
	LSTM+SA	59.8	49.0
	LSTM+ELMo	60.9	50.9
	LSTM+SA+ELMo	62.7	52.7
None	Rajpurkar et al. (2016)	51.0	40.0
	Yu et al. (2018)	89.3	82.5

Decomposability gap

DeSPI (Seo & Lee et al., 2019)

	Model	EM	F1	W/s
Original	DrQA	69.5	78.8	4.8K
	BERT	84.1	90.9	51
Query-Agnostic	LSTM+SA	49.0	59.8	-
	LSTM+SA+ELMo	52.7	62.7	-
	DESPI (dense only)	73.6	81.7	28.7M
	–Coherency scalar	71.5	81.5	28.7M

Decomposability gap

Exact search: **6000** times faster than DrQA, **5M** times faster than BERT

DeSPI (Seo & Lee et al., 2019)

	F1	EM	s/Q	#D/Q
DrQA	-	$\boxed{29.8}$	$\boxed{35}$	5
R^3	37.5	-	-	-
Paragraph ranking	-	30.2	-	-
Multi-step reasoner	39.2	31.9	-	-
MINIMAL	42.5	34.7	-	-
BERTserini	46.1	38.6	-	-
DeSPI	42.3	$\boxed{33.4}$	$\boxed{0.51}$	532
−Sparse vector	20.5	13.3	0.22	532
+Sparse search	38.6	31.2	0.23	5

Demo

Today’s Talk

- Intro: About Question Answering & Reasoning
- **Parsing**: SOTA on WikiSQL
- **Reading**: Real-time Open-domain Question Answering
- Towards a Unified Model

Parsing vs Reading

	Parsing	Reading
Speed	Fast	Slow
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

Parsing vs Reading

	Parsing	Reading
Speed	Fast	Fast
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

Parsing vs Reading

	Parsing	Reading
Speed	Fast	Fast
Domain	Limited	Ontology-free, open-domain
Complexity	Multi-hop reasoning	Limited

Can a **reader** handle complex queries?

Have people tried it?

- **Yes**, on synthetic data: bAbI (Weston et al., 2016)
  - Syntactically simple (synthetic) text
  - Predefined types of reasoning
- And **Yes**, on state changing dataset: ProPara (Dalvi et al., 2018)
  - About single entity
  - The state of the entity changes throughout the text sequentially
- Another **Yes**, on multi-hop QA dataset: HotpotQA (Yang et al., 2018)
  - SQuAD-like, but multi-hop and open-domain

bAbI (Weston et al., 2016)

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

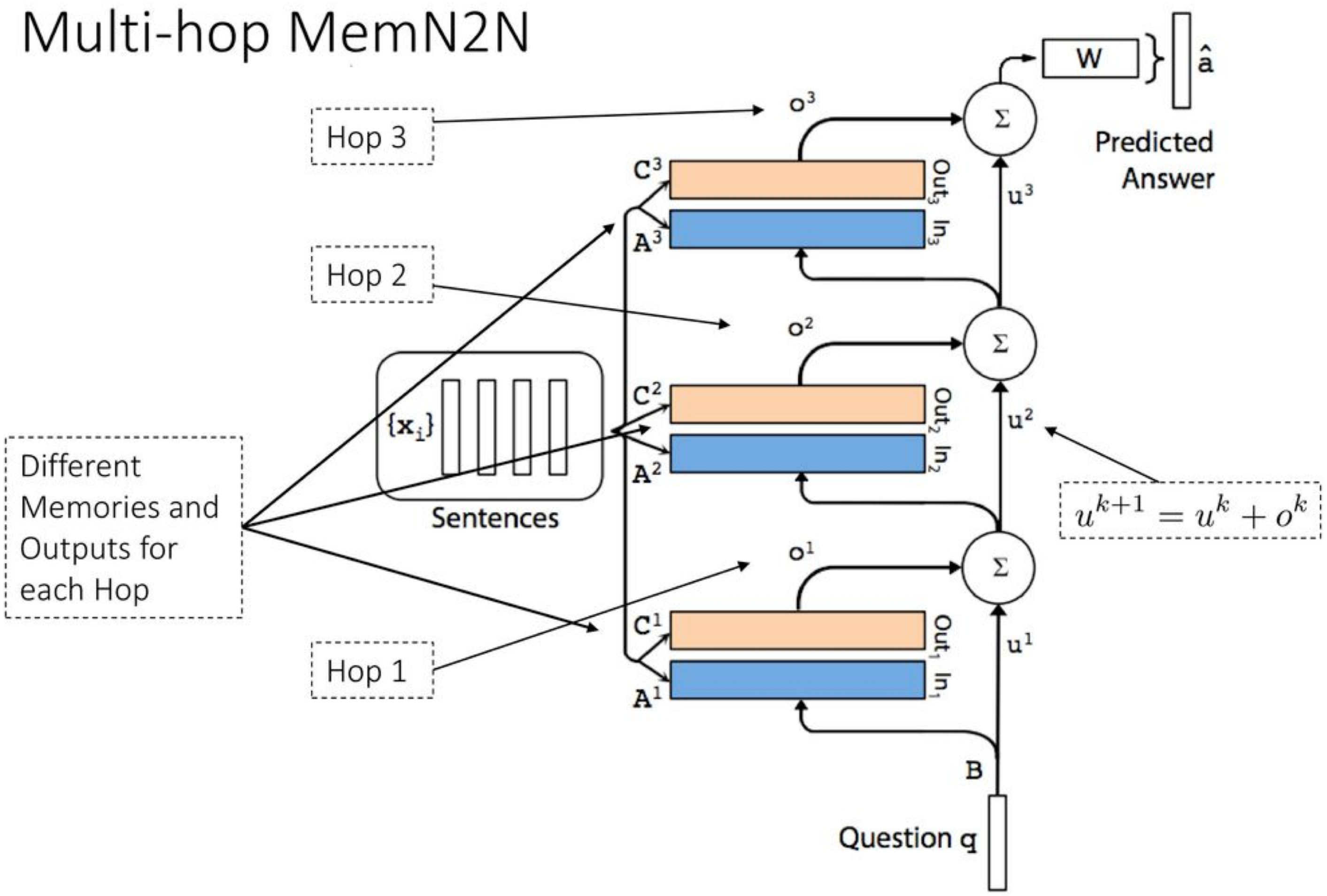
Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

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

End-to-end Memory Networks (Sukbaatar et al., 2015)



End-To-End Memory Networks, Sukhbaatar et. al., NIPS 2015

Query-Reduction Networks (Seo et al., 2017)

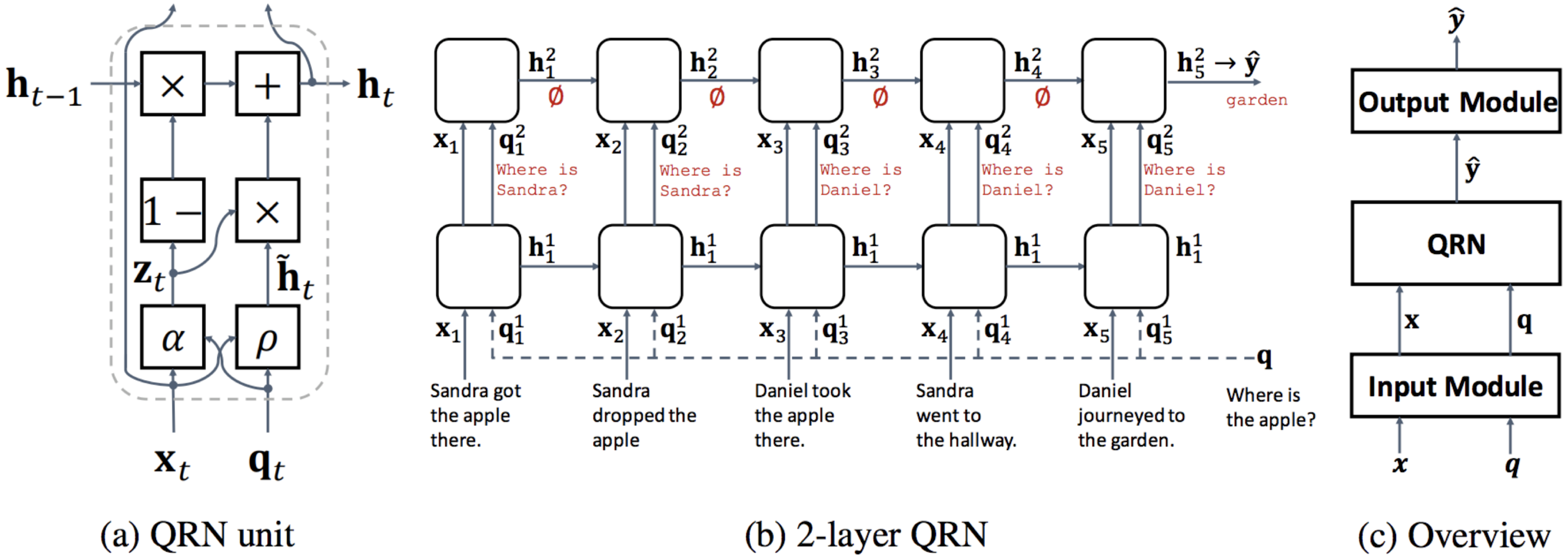
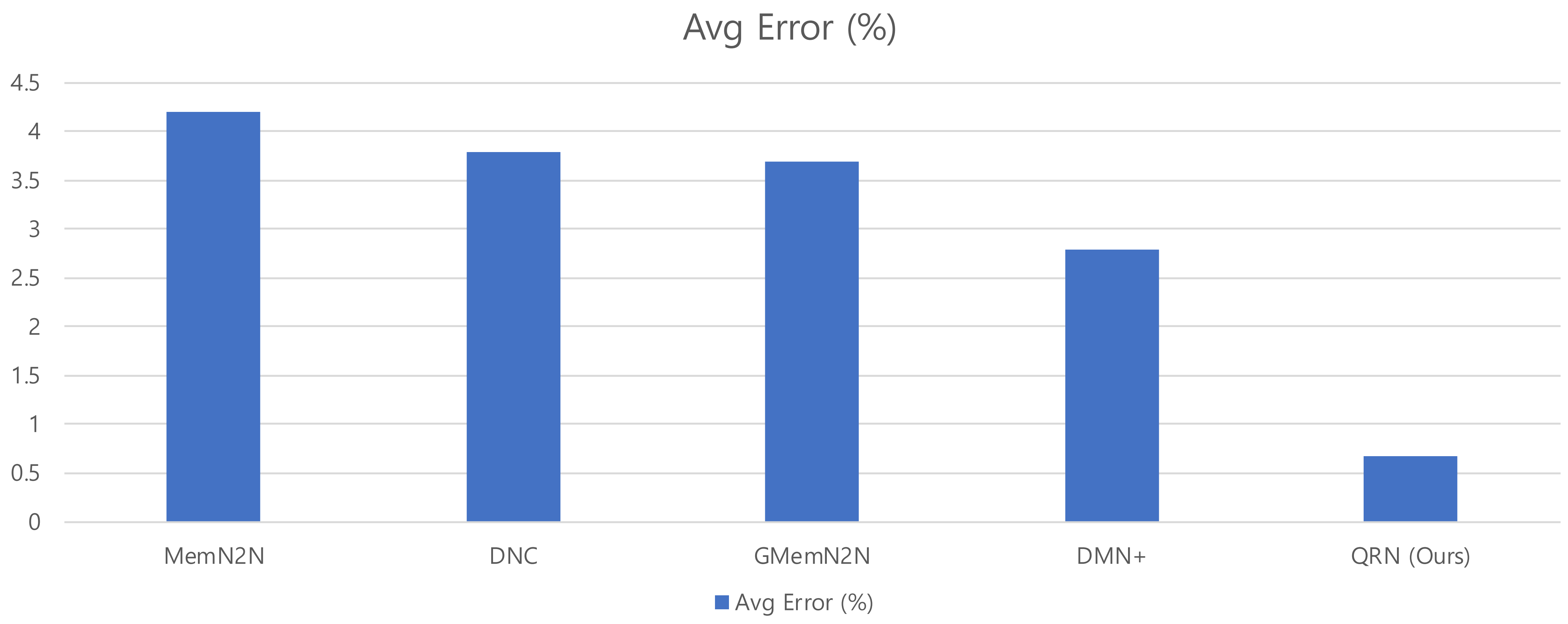


Figure 1: (1a) QRN unit, (1b) 2-layer QRN on 5-sentence story, and (1c) entire QA system (QRN and input / output modules). $\boldsymbol{x}, \boldsymbol{q}, \hat{\boldsymbol{y}}$ are the story, question and predicted answer in natural language, respectively. $\mathbf{x} = \langle \mathbf{x}_1, \ldots, \mathbf{x}_T \rangle, \mathbf{q}, \hat{\mathbf{y}}$ are their corresponding vector representations (upright font). α and ρ are update gate and reduce functions, respectively. $\hat{\mathbf{y}}$ is assigned to be $\mathbf{h}_5^2$, the local query at the last time step in the last layer. Also, red-colored text is the inferred meanings of the vectors (see 'Interpretations' of Section 5.3).

bAbI QA Results (10k)



ProPara (Dalvi et al., 2018)

- Real data
- Sequentially state-changing

Chloroplasts in the leaf of the plant trap light from the sun. The roots absorb water and minerals from the soil. This combination of water and minerals flows from the stem into the leaf. Carbon dioxide enters the **leaf**. Light, water and minerals, and the carbon dioxide all combine into a mixture. This mixture forms **sugar** (glucose) which is what the plant eats.

Q: Where is sugar produced?
A: in the leaf

QRN (Seo et al., 2017) vs ProStruct (Tandon et al., 2018)

	Precision	Recall	F1
ProLocal	77.4	22.9	35.3
QRN	55.5	31.3	40.0
EntNet	50.2	33.5	40.2
ProGlobal	46.7	52.4	49.4
ProStruct	74.2	42.1	**53.7**

Table 1: Results on the prediction task (test set).

HotpotQA (Yang et al., 2018)

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

Leaderboard (Fullwiki Setting)

In the *fullwiki* setting, a question-answering system must find the answer to a question in the scope of the entire Wikipedia. Similar to in the distractor setting, systems are evaluated on the accuracy of their answers (Ans) and the quality of the supporting facts they use to justify them (Sup).

	Model	Code	Ans		Sup		Joint	
			EM	F_1	EM	F_1	EM	F_1
1 Feb 21, 2019	Cognitive Graph (single model) *Anonymous*		**37.12**	**48.87**	**22.82**	**57.69**	**12.42**	**34.92**
2 Mar 5, 2019	MUPPET (single model) *Anonymous*		30.61	40.26	16.65	47.33	10.85	27.01
3 Mar 4, 2019	GRN (single model) *Anonymous*		27.34	36.48	12.23	48.75	7.40	23.55
4 Nov 25, 2018	QFE (single model) *NTT Media Intelligence Laboratories*		28.66	38.06	14.20	44.35	8.69	23.10
5 Oct 12, 2018	Baseline Model (single model) *Carnegie Mellon University, Stanford University, & Universite de Montreal* (Yang, Qi, Zhang, et al. 2018)		23.95	32.89	3.86	37.71	1.85	16.15
- Feb 28, 2019	DecompRC (single model) *Anonymous*		30.00	40.65	N/A	N/A	N/A	N/A
- Mar 3, 2019	MultiQA (single model) *Anonymous*		30.73	40.23	N/A	N/A	N/A	N/A

So, when is **reader** capable of end-to-end *reasoning*?

- Syntactically simple sentences (bAbI)
- Sequential reasoning (ProPara)
- Strong supervision (ProPara, HotpotQA)

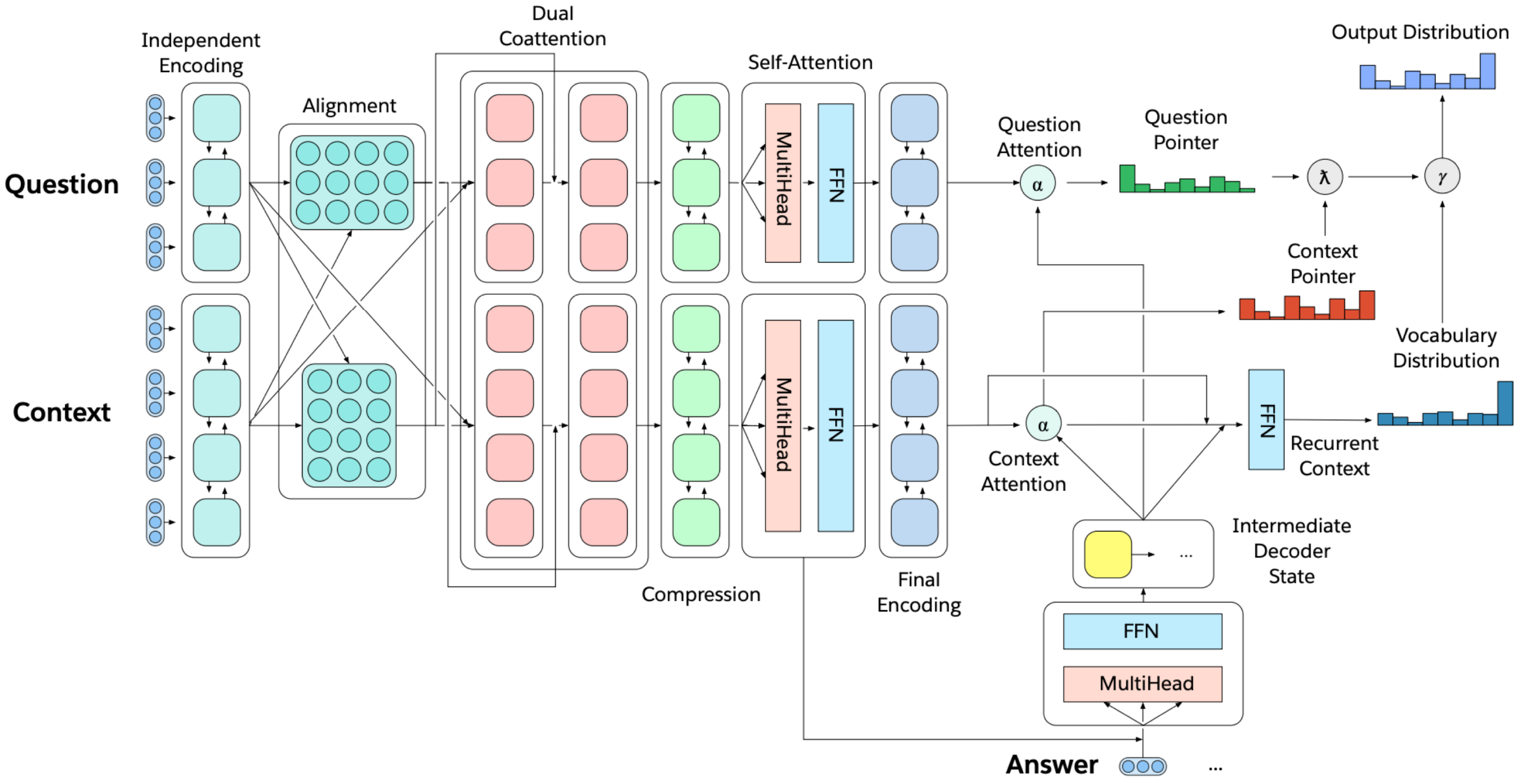
Conclusion

- **Parser** and **Reader** have different strengths
- Making **Parser** open-domain, ontology-free is hard (or doesn't make sense by definition)
- **Reader** is starting to overcome some of its bottlenecks (e.g. speed, reasoning)
- Future research will get us closer to a true *unified model* for question answering and reasoning

Thanks!

- minjoon.seo@navercorp.com
- seominjoon@gmail.com
- https://seominjoon.github.io

DecaNLP (McCann et al., 2018)



References

- Pasupat and Liang. Compositional Semantic Parsing on Semi-Structured Tables. ACL 2015.
- Minjoon Seo, Ali Farhadi, Hannaneh Hajishirzi, Oren Etzioni, Clint Malcolm. Solving Geometry Problems: Combining Text and Diagram Interpretations. EMNLP 2015.
- Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading Wikipedia to Answer Open-Domain Questions. ACL 2017
- Jinhyuk Lee, Seongjun Yun, Hyunjae Kim, Miyoung Ko and Jaewoo Kang. Ranking Paragraphs for Improving Answer Recall in Open-Domain Question Answering. EMNLP 2018.
- Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin and Tomas Mikolov. Towards AI Complete Question Answering: A Set of Prerequisite Toy Tasks. ICLR 2016
- Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus. End-to-end memory networks. NIPS 2015.
- Niket Tandon, Bhavana Dalvi Mishra, Joel Grus, Wen-tau Yih, Antoine Bosselut, and Peter Clark. Reasoning about Actions and State Changes by Injecting Commonsense Knowledge. EMNLP 2018.
- Bhavana Dalvi, Lifu Huang, Niket Tandon, Wen-tau Yih, and Peter Clark. 2018. Tracking state changes in procedural text: a challenge dataset and models for process paragraph comprehension. NAACL 2018.
- Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. HOTPOTQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. EMNLP 2018.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang. SQuAD: 100,000+ Questions for Machine Comprehension of Text. EMNLP 2016.
- Wonseok Hwang, Jinyeong Yim, Seunghyun Park, and Minjoon Seo. A Comphrensive Exploration on WikiSQL with Table-Aware Word Contextualization. arXiv 2019.
- Minjoon Seo, Jinhyuk Lee, Tom Kwiatkowski, Ankur Parikh, Ali Farhadi, and Hannaneh Hajishirzi. Real-Time Open-Domain Question Answering with Dense-Sparse Phrase Index. In submission to ACL 2019.
- Shrivastava and Li. Asymmetric lsh (alsh) for sublinear time maximum inner product search (mips). NIPS 2014.
- Minjoon Seo, Tom Kwiatkowski, Ankur Parikh, Ali Farhadi, and Hannaneh Hajishirzi. Phrase-Indexed Question Answering: A New Challenge for Scalable Document Comprehension. EMNLP 2018.
- Minjoon Seo, Sewon Min, Ali Farhadi, and Hannaneh Hajishirzi. Query-Reduction Networks for Question Answering. ICLR 2017.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019.
- Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, Richard Socher. The Natural Language Decathlon: Multitask Learning as Question Answering. arXiv 2018.