

Differentiable Reasoning over a Virtual Knowledge Base

NOVA IR Reading group

07/09/2020

Which companies are developing vaccines for COVID-19?

What is the size of the COVID-19 virus?

Which companies are developing vaccines for COVID-19?

What is the size of the COVID-19 virus?

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

Experiments suggested that SARS-CoV-2 has affinity to the ACE2 receptors of human cells.

ACE2 is an enzyme attached to the cells in the lungs, heart, kidney, and intestines.

Each SARS-CoV-2 virion is approximately 50-200 nanometres in diameter.

Ans: 50-200 nanometres

Text Corpus



Question

Answer

Text Corpus



Voglibose

From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding sources. Unsourced material may be challenged and removed.

Find sources: "Voglibose" – news · newspapers · books · scholar · JSTOR (September 2008) (Learn how and when to remove this template message)

Voglibose (INN and USAN, trade name **Veglib**, marketed by Maccor Health Series) is an alpha-glucosidase inhibitor used for lowering post-prandial blood glucose levels in people with diabetes mellitus. Voglibose delays the absorption of glucose thereby reducing the risk of macrovascular complications. Voglibose is a research product of Takeda Pharmaceutical Company, Japan's largest pharmaceutical company. Voglibose was first launched in 1994, under the trade name **BASEN**, to improve postprandial hyperglycemia in diabetes mellitus.^[1]

1. Text Retrieval

Question

Answer

Text Corpus



Voglibose
From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding sources. Unsourced material may be challenged and removed.
Find sources: "Voglibose" – news · newspapers · books · scholar · JSTOR (September 2008) Learn how and message me.

Voglibose (INN and USAIN, trade name **Weglib**, marketed by Eisai Health Series) is an alpha-glucosidase inhibitor used for lowering post-prandial blood glucose levels in people with *diabetes mellitus*. Voglibose delays the absorption of glucose thereby reducing the risk of macrovascular complications. Voglibose is a research product of Teikoku Pharmaceutical Company, Japan's largest pharmaceutical company. Voglibose was first launched in 1994, under the trade name BASEN, to improve postprandial hyperglycemia in *diabetes mellitus*.¹

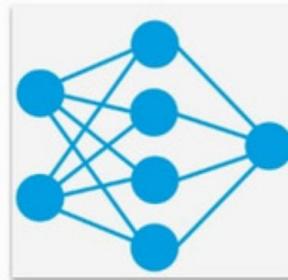
1. Text Retrieval

Question



2. Reading Comprehension

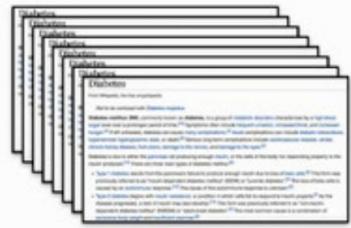
Answer



Who voices the dog in the TV show Family Guy?

Multi-Hop Question

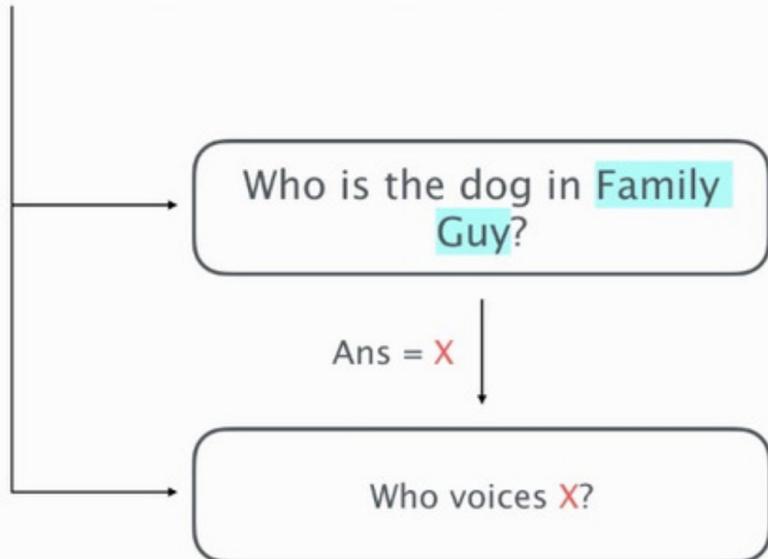
Text Corpus



Who voices the dog in the TV show Family Guy?

Multi-Hop Question

Text Corpus



Who voices the dog in the TV show Family Guy?

Multi-Hop Question

Text Corpus



Who is the dog in Family Guy?

Family Guy includes an anthropomorphic dog Brian.

Ans = Brian Griffin

Who voices Brian Griffin?

Who voices the dog in the TV show Family Guy?

Multi-Hop Question

Text Corpus



Who is the dog in Family Guy?

Family Guy includes an anthropomorphic dog Brian.

Ans = Brian Griffin

Who voices Brian Griffin?

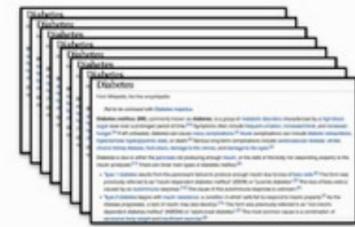
The voice of Brian is provided by Seth Macfarlane.

Ans = Seth Macfarlane

Who voices the dog in the TV show Family Guy?

Multi-Hop Question

Text Corpus



Who is the dog in Family Guy?

Family Guy includes an anthropomorphic dog Brian.

Ans = Brian Griffin

Unknown!

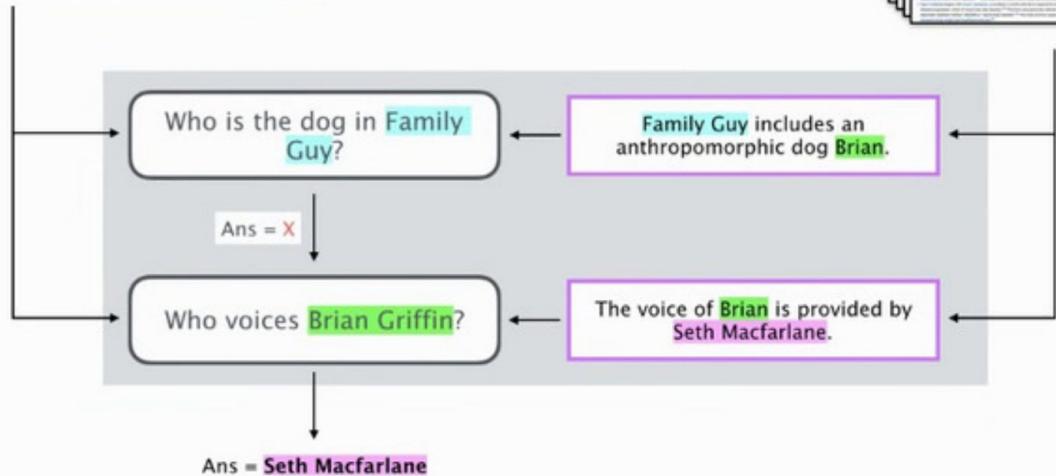
Who voices Brian Griffin?

The voice of Brian is provided by Seth Macfarlane.

Ans = Seth Macfarlane



Who voices the dog in the TV show Family Guy?



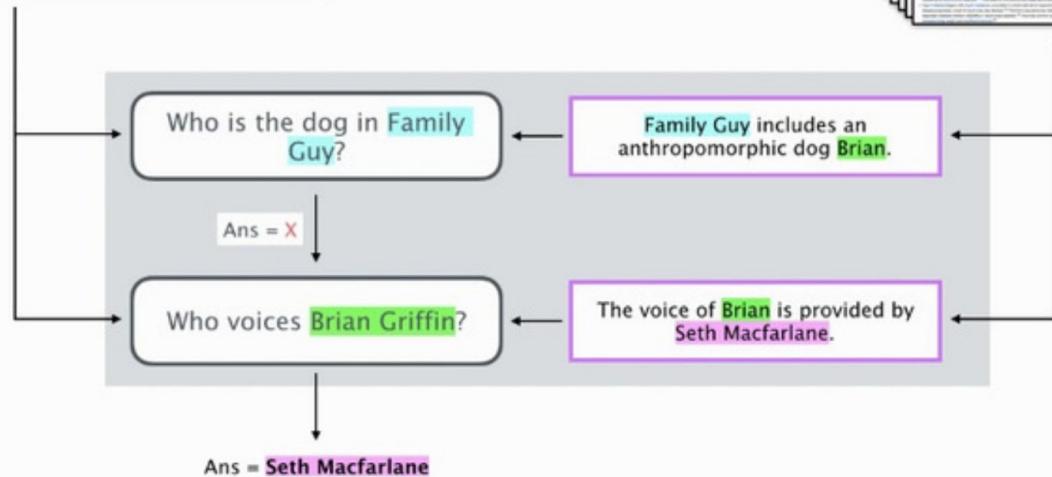
Prior Work:

[Talmor & Berant, 2018; Sun et al, 2019]

- Heuristically identify intermediate steps
Does not generalize
- Iteratively retrieve and read
Slow

Our goals —

Who voices the dog in the TV show Family Guy?



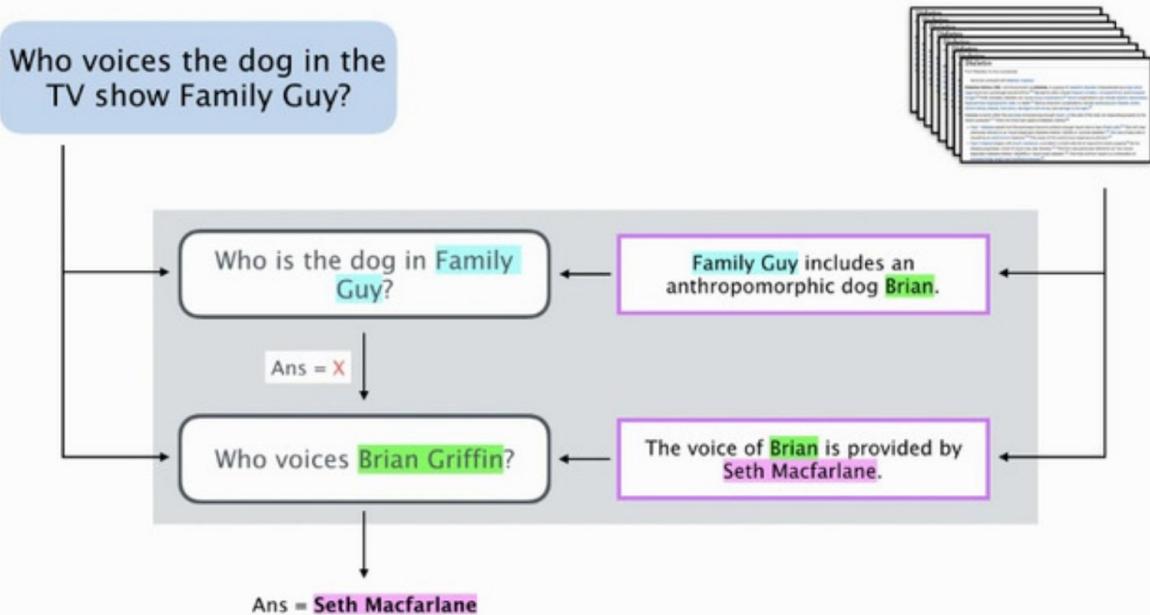
Prior Work:

[Talmor & Berant, 2018; Sun et al, 2019]

- Heuristically identify intermediate steps
Does not generalize
- Iteratively retrieve and read
Slow

Our goals —

End-to-End



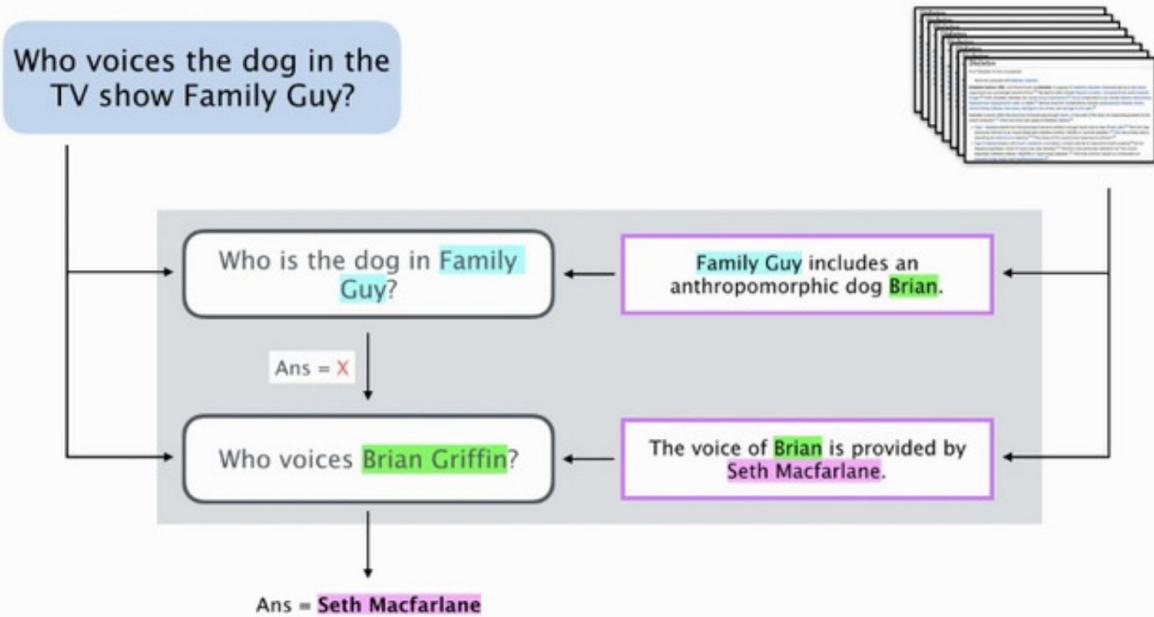
Prior Work:

[Talmor & Berant, 2018; Sun et al, 2019]

- Heuristically identify intermediate steps
Does not generalize
- Iteratively retrieve and read
Slow

Our goals —

- End-to-End
- Efficiently



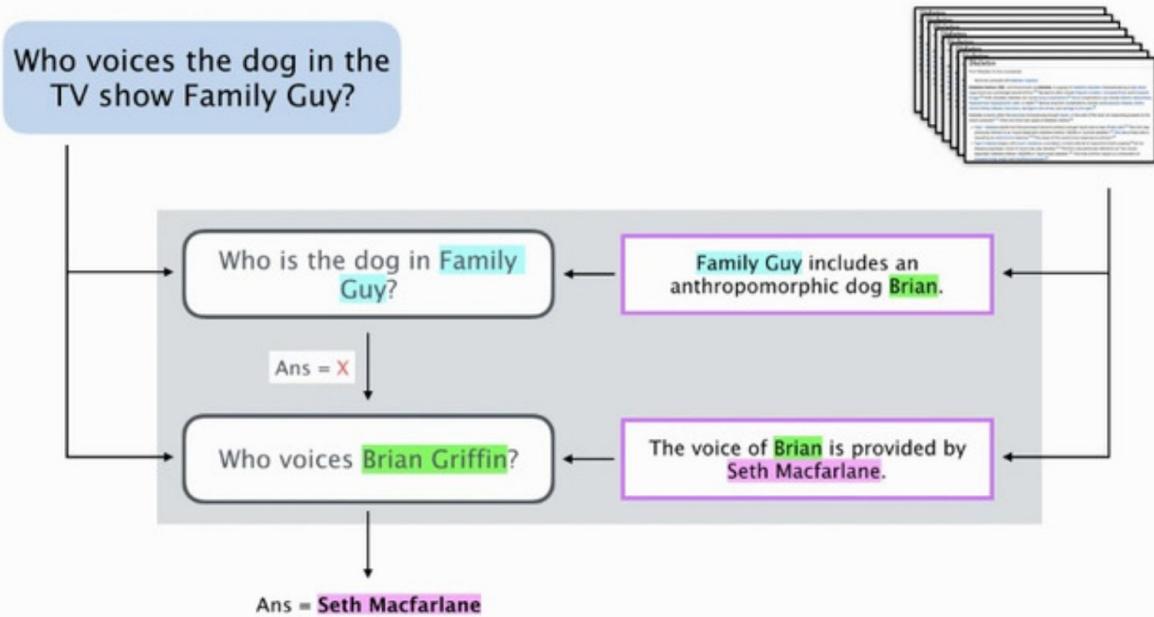
Prior Work:

[Talmor & Berant, 2018; Sun et al, 2019]

- Heuristically identify intermediate steps
Does not generalize
- Iteratively retrieve and read
Slow

Our goals —

- End-to-End
- Efficiently
- Compositionally



Prior Work:

[Talmor & Berant, 2018; Sun et al, 2019]

- Heuristically identify intermediate steps
Does not generalize
- Iteratively retrieve and read
Slow

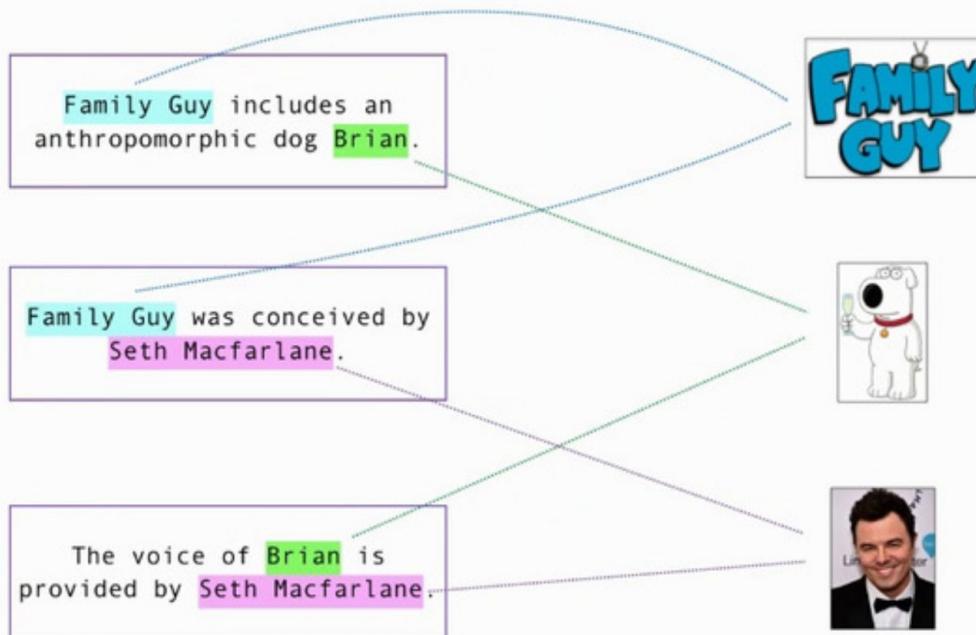
Text Corpus

Family Guy includes an anthropomorphic dog Brian.

Family Guy was conceived by Seth Macfarlane.

The voice of Brian is provided by Seth Macfarlane.

Text Corpus



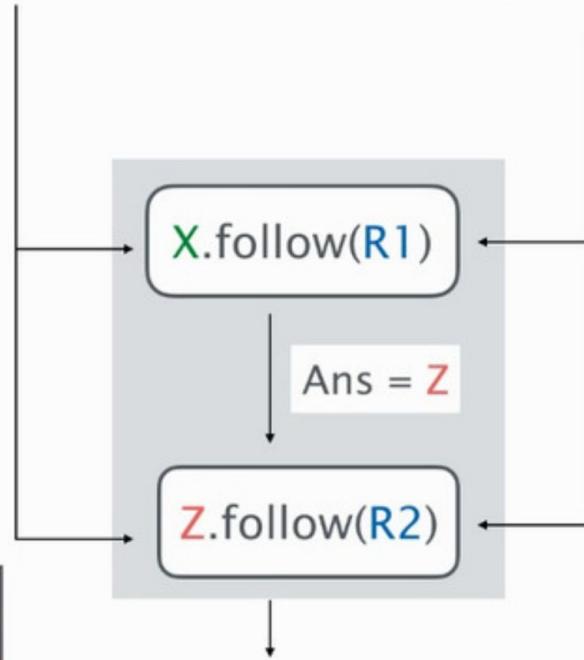
Mentions

Entities

Who voices the dog in
the TV show Family Guy?



$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$

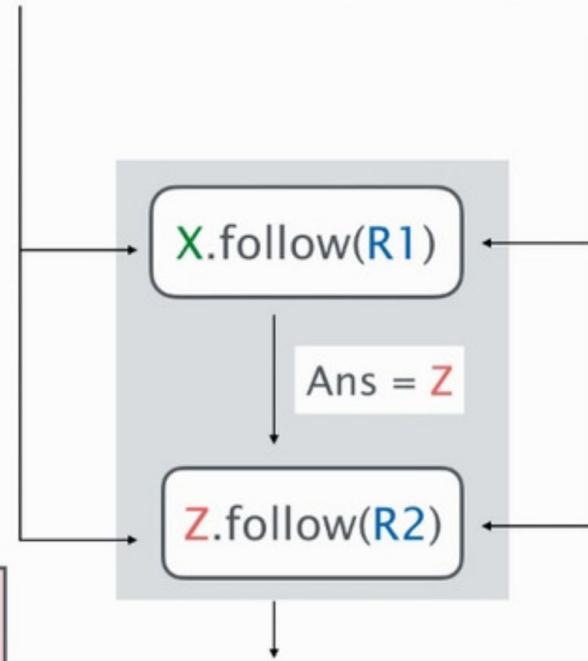


Who voices the dog in
the TV show Family Guy?



- Given a set of entities X

$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$

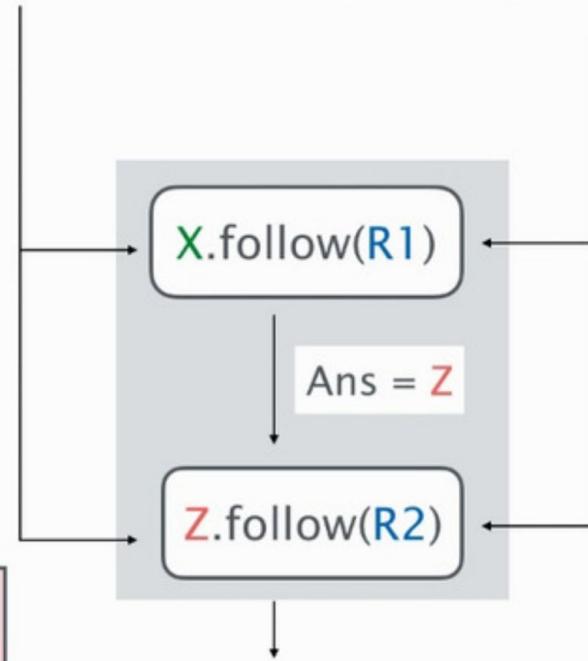


Who voices the dog in
the TV show Family Guy?



- Given a set of entities X
- Follow the relation R

$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$

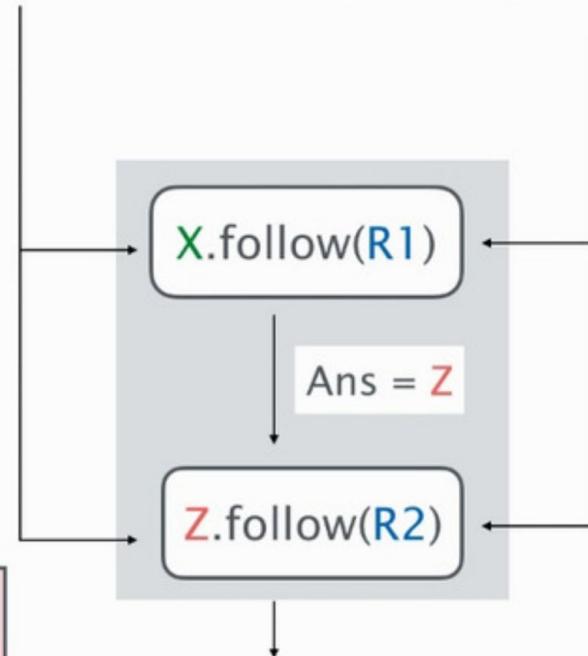


Who voices the dog in
the TV show Family Guy?



- Given a set of entities X
- Follow the relation R
- To arrive at a set of entities Y

$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$



$X \sim \{ \text{Family Guy} \}$

$R1 \sim \text{"dog in show"}$

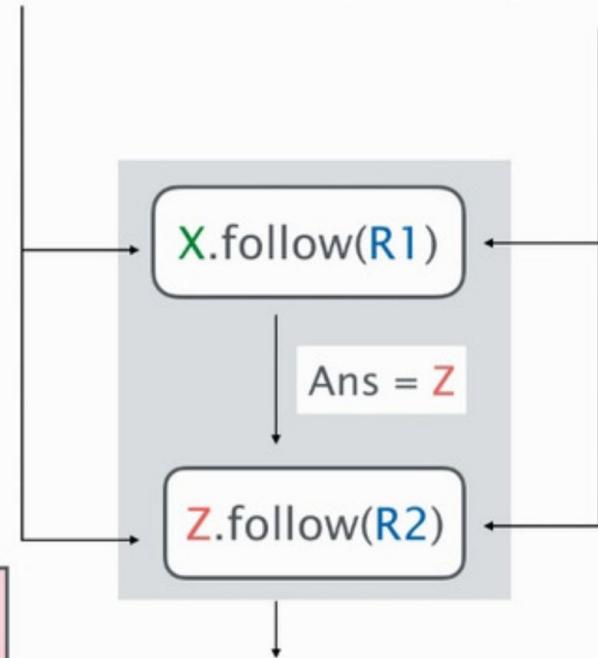
$R2 \sim \text{"voiced by"}$

Who voices the dog in
the TV show Family Guy?



- Given a set of entities X
- Follow the relation R
- To arrive at a set of entities Y

$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$



$X \sim \{ \text{Family Guy} \}$

$R1 \sim \text{"dog in show"}$

$R2 \sim \text{"voiced by"}$

Who voices the dog in
the TV show Family Guy?

- Given a set of entities X
- Follow the relation R
- To arrive at a set of entities Y
$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$

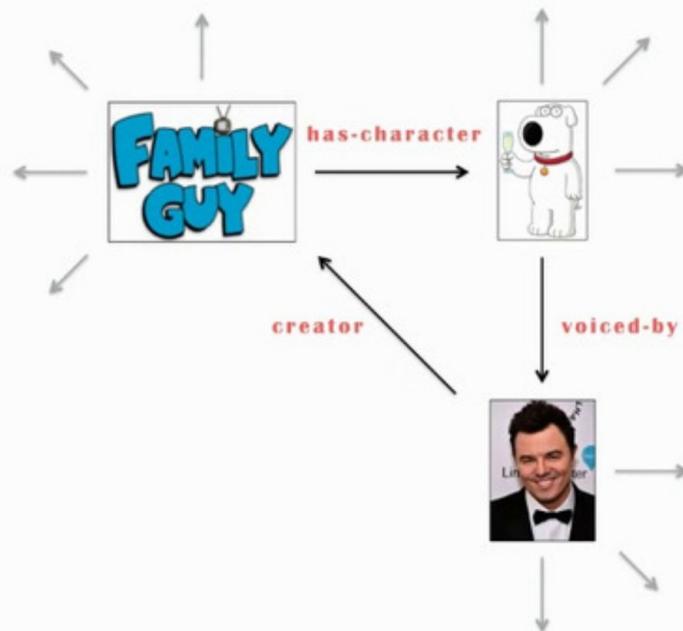
$X.\text{follow}(R1)$

Ans = Z

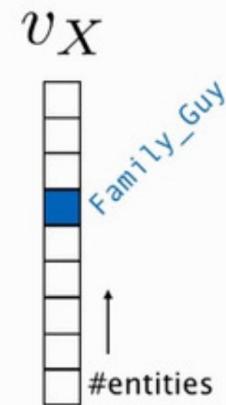
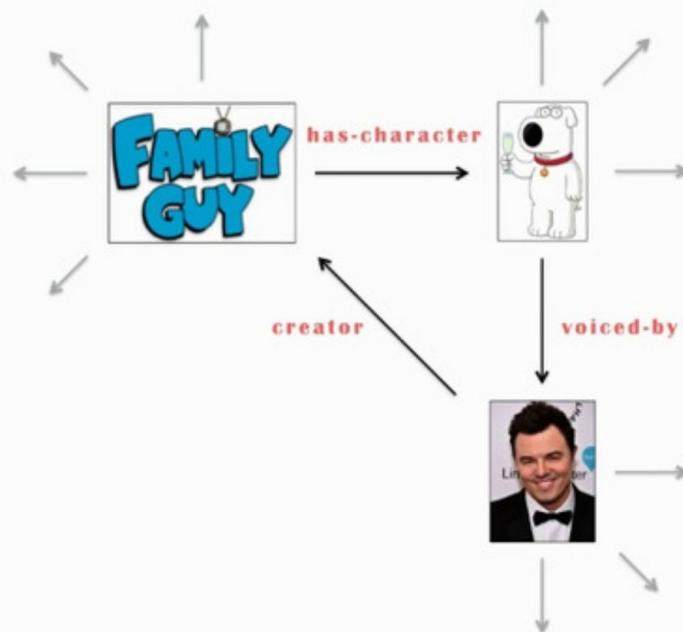
$Z.\text{follow}(R2)$



$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$

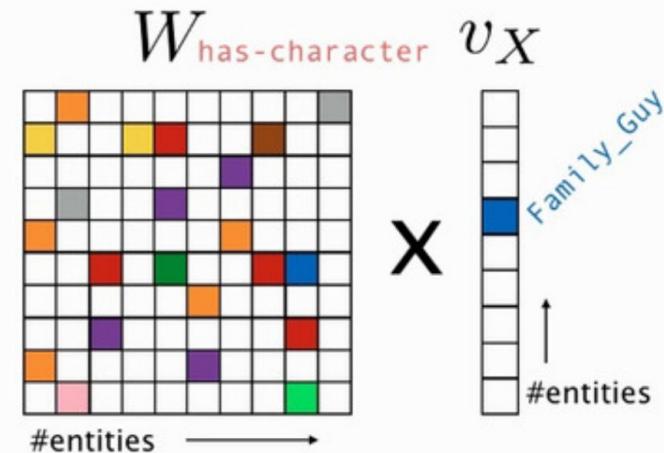
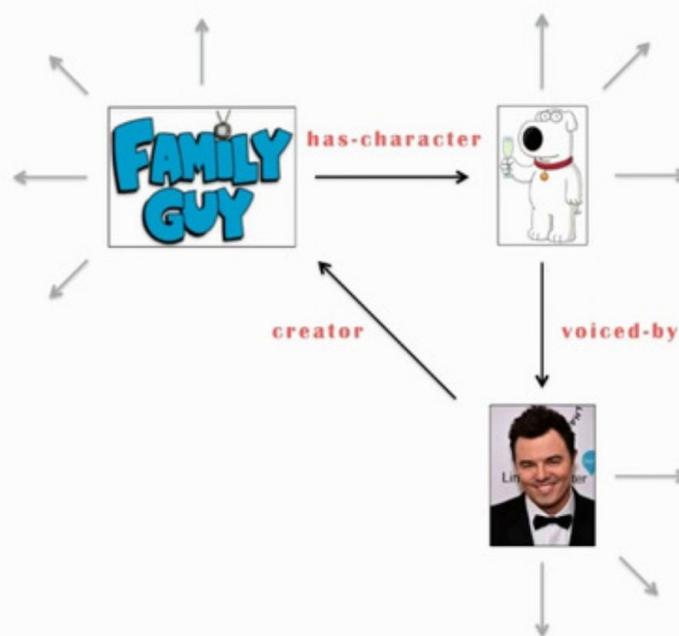


$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$



$$v_X[i] = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{o.w.} \end{cases}$$

$$Y = X.\text{follow}(R) = \{x' \text{ s.t. } R(x, x') \text{ holds}\}$$



$$W_R[i, j] = \begin{cases} 1 & \text{if } R(j, i) \\ 0 & \text{o.w.} \end{cases} \quad v_X[i] = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{o.w.} \end{cases}$$



R = “dog in show”

$$Y = X.\text{follow}(R)$$



X =

R = "dog in show"

$$Y = X.\text{follow}(R)$$

1. Expand X to co-occurring mentions



Brian Griffin is
a character in
the TV series
Family Guy

Family Guy
includes an
anthropomorphic
dog Brian.

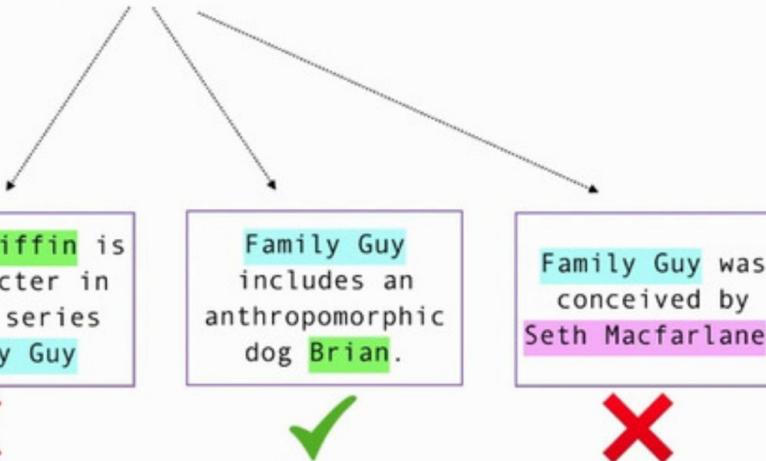
Family Guy was
conceived by
Seth Macfarlane.



X =

R = “dog in show”

$$Y = X.\text{follow}(R)$$



1. Expand X to co-occurring mentions

2. Filter mentions based on R

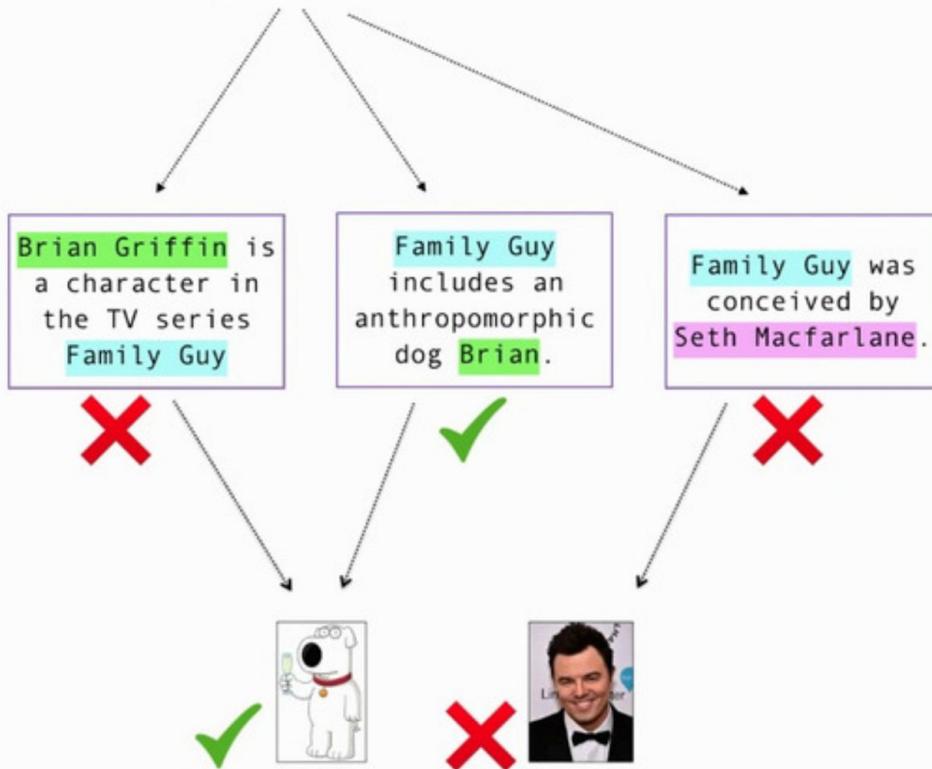
$$\text{score} = f(\text{mention context}, R)$$



X =

R = "dog in show"

$$Y = X.\text{follow}(R)$$



1. Expand X to co-occurring mentions
2. Filter mentions based on R
3. Combine scores of the same entity

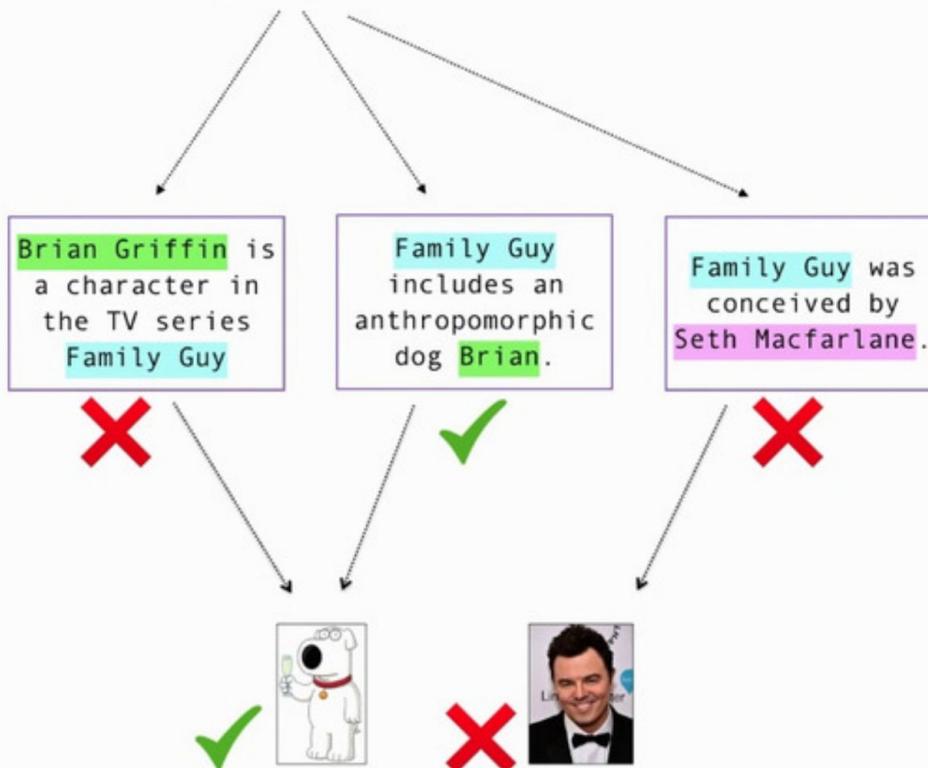
$$\text{score} = f(\text{mention context}, R)$$



X =

R = "dog in show"

$$Y = X.\text{follow}(R)$$

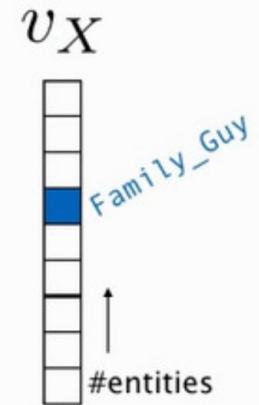


1. Expand X to co-occurring mentions
2. Filter mentions based on R
3. Combine scores of the same entity

$$\text{score} = f(\text{mention context}, R)$$

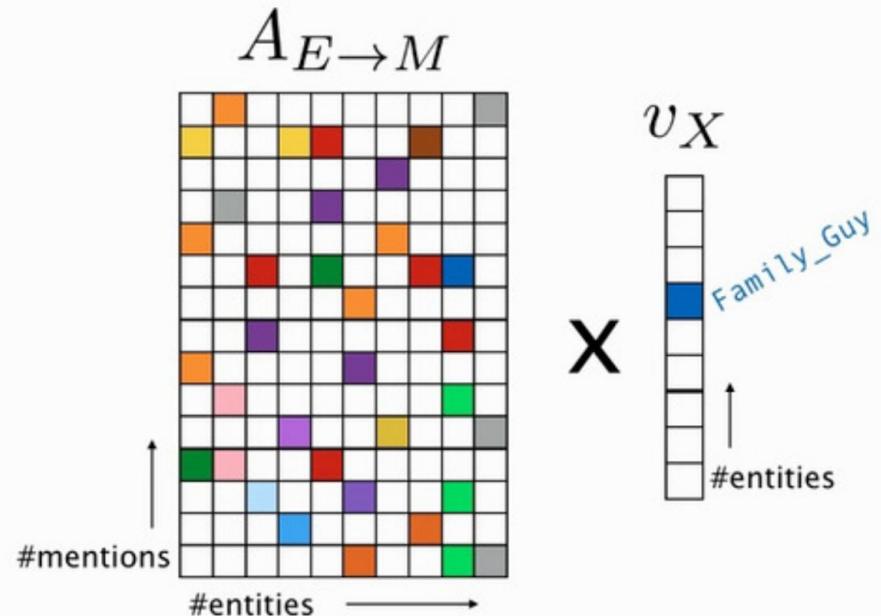
Key Idea: We can do this
efficiently with inner products

1. Expand X to co-occurring mentions



$$v_X[i] = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{o.w.} \end{cases}$$

1. Expand X to co-occurring mentions

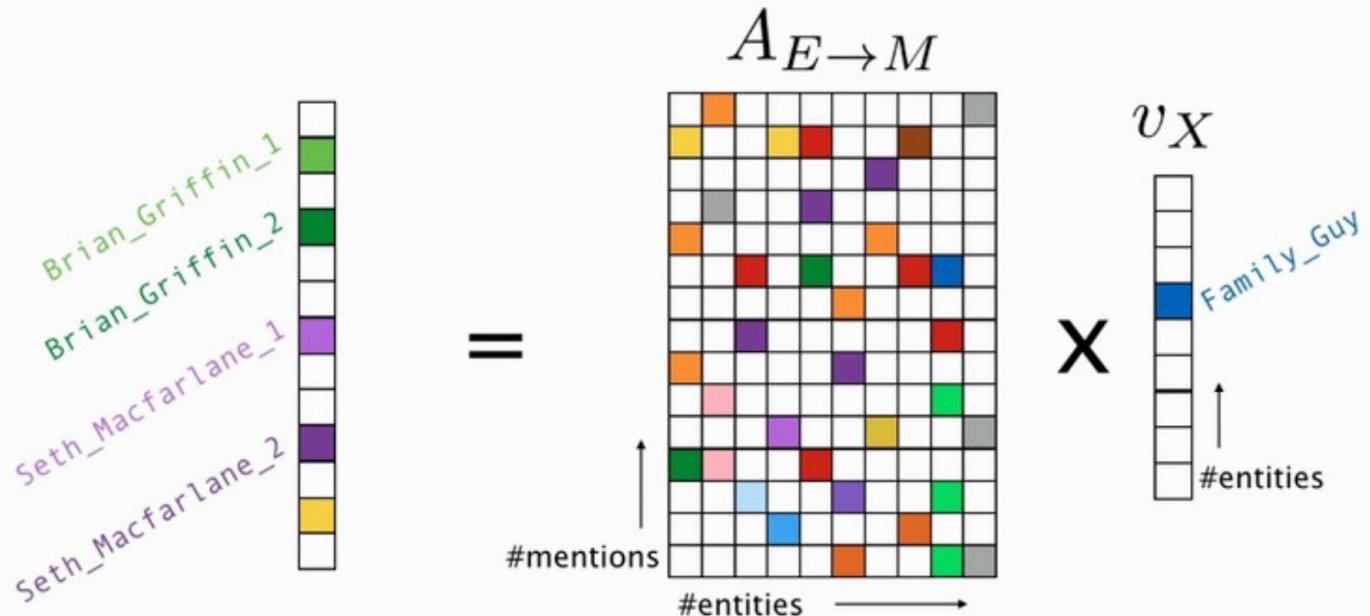


$$A_{E \rightarrow M}[i, j] = \begin{cases} 1 & \text{entity } j \text{ co-occurs with mention } i \\ 0 & \text{o.w.} \end{cases}$$

Pre-computed

$$v_X[i] = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{o.w.} \end{cases}$$

1. Expand X to co-occurring mentions



$$A_{E \rightarrow M}[i, j] = \begin{cases} 1 & \text{entity } j \text{ co-occurs with mention } i \\ 0 & \text{o.w.} \end{cases}$$

Pre-computed

$$v_X[i] = \begin{cases} 1 & \text{if } i \in X \\ 0 & \text{o.w.} \end{cases}$$

2. Filter mentions based on R

R = “dog in show”

Family Guy includes an anthropomorphic dog Brian.

m

2. Filter mentions based on R

R = “dog in show”



q_R

Family Guy includes an anthropomorphic dog Brian.

\underline{m}

2. Filter mentions based on R

$R = \text{"dog in show"}$ —→



q_R

$f(m)$



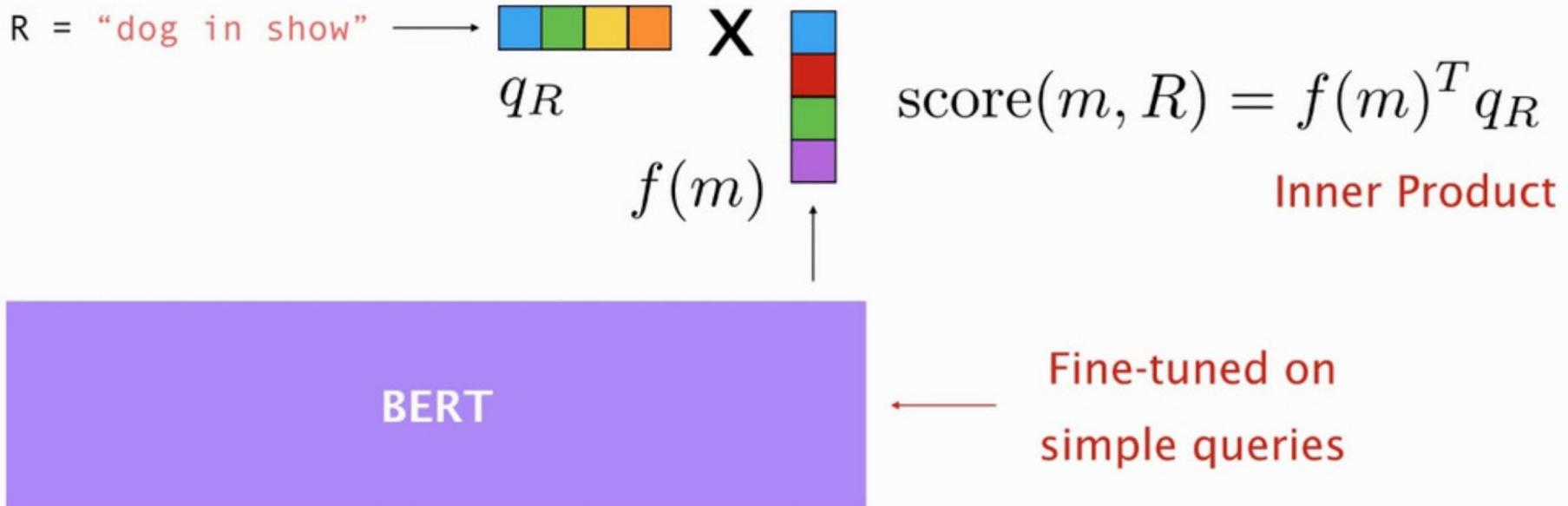
Fine-tuned on
simple queries

Family Guy includes an anthropomorphic dog Brian.

m

Levy et al,
CoNLL 2017

2. Filter mentions based on R



Family Guy includes an anthropomorphic dog Brian.

m

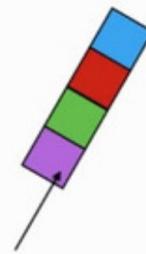
Levy et al,
CoNLL 2017

2. Filter mentions based on R

R = “dog in show”



q_R



Family Guy in

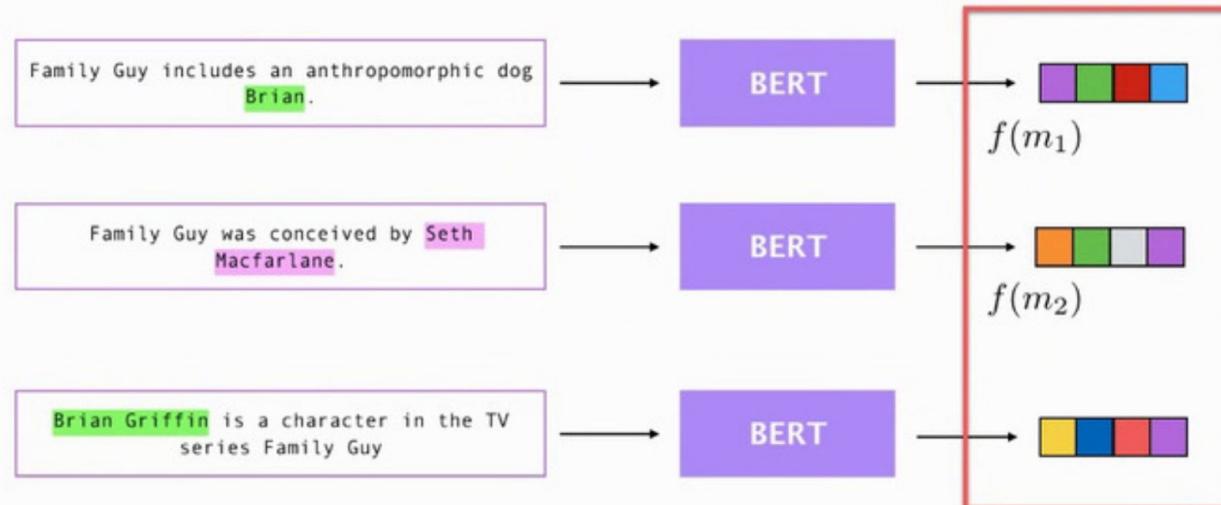
BERT



2. Filter mentions based on R

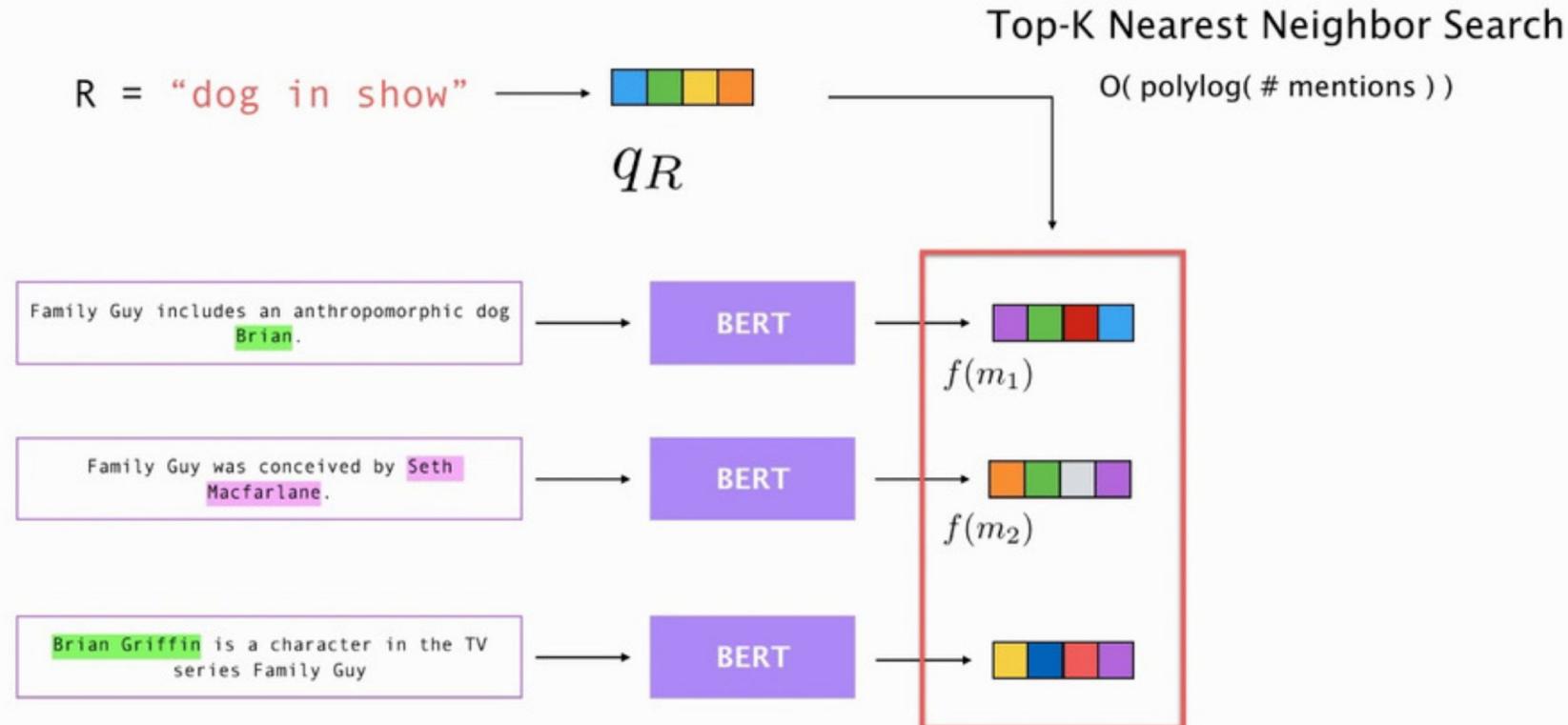
$R = \text{"dog in show"}$ \longrightarrow 

q_R



2. Filter mentions based on R

Andoni et al,
NIPS 2015



Seo et al,
EMNLP 2018; ACL 2019

Offline index

2. Filter mentions based on R



2. Filter mentions based on R

$$\mathcal{T}_K(q_R)$$



Mentions whose type
matches R

$$A_{E \rightarrow M} v_X$$



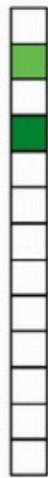
Mentions co-occurring
with X

2. Filter mentions based on R

$$\mathcal{T}_K(q_R)$$

$$A_{E \rightarrow M} v_X$$

Mentions which satisfy
 $X.\text{follow}(R)$



=



•
Element-wise
Product



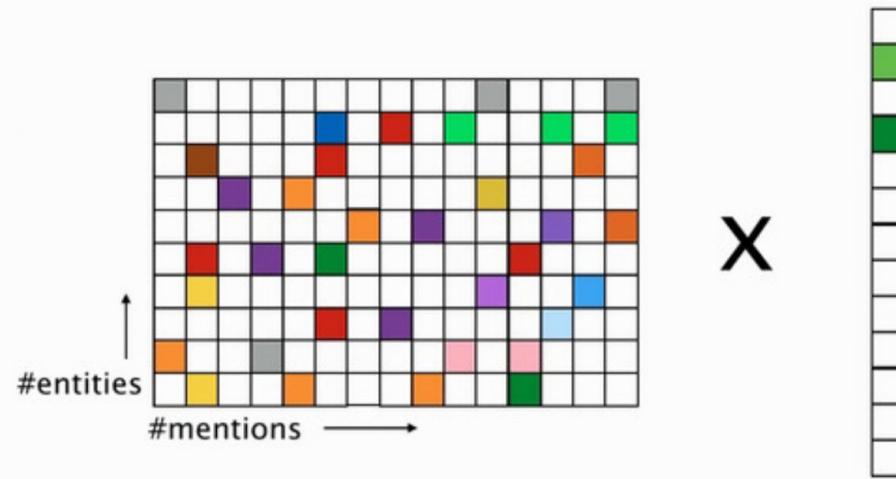
Mentions whose type
matches R

Mentions co-occurring
with X

3. Combine scores of the same entity

$$B_{M \rightarrow E}$$

$$\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X$$



Pre-computed

$$B_{M \rightarrow E}[i, j] = \begin{cases} 1 & \text{if mention } j \text{ refers entity } i \\ 0 & \text{o.w.} \end{cases}$$

3. Combine scores of the same entity

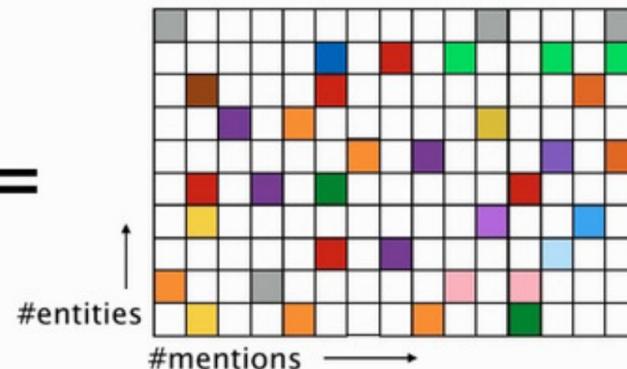
$v_{X.\text{follow}(R)}$

$B_{M \rightarrow E}$

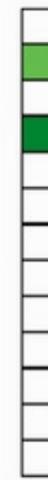
$\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X$



=



X



$B_{M \rightarrow E}[i, j] = \begin{cases} 1 & \text{if mention } j \text{ refers entity } i \\ 0 & \text{o.w.} \end{cases}$

Pre-computed

$$v_{X.\text{follow}(R)} = B_{M\rightarrow E}\left[\mathcal{T}_K(q_R)\odot A_{E\rightarrow M}v_X\right]$$

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

1. Expand X to co-occurring mentions

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

1. Expand **X** to co-occurring mentions
2. Filter mentions based on **R**

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

1. Expand **X** to co-occurring mentions
2. Filter mentions based on **R**
3. Combine scores of the same entity

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

 Efficient

$O(\mu + \text{polylog}|\mathcal{M}|)$

out-degree # mentions

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

Efficient

$O(\mu + \text{polylog}|\mathcal{M}|)$

out-degree # mentions

Closed under composition

$X.\text{follow}(R_1).\text{follow}(R_2)$

$$v_{X.\text{follow}(R)} = B_{M \rightarrow E} [\mathcal{T}_K(q_R) \odot A_{E \rightarrow M} v_X]$$

Efficient

$O(\mu + \text{polylog}|\mathcal{M}|)$

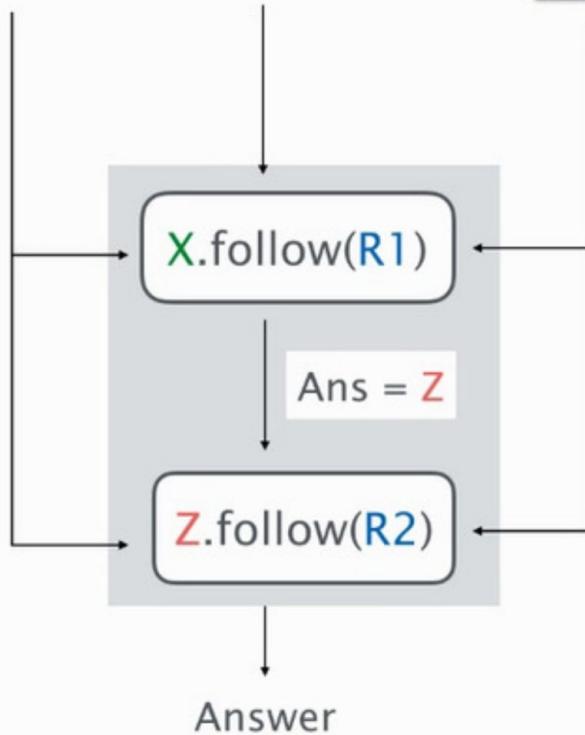
out-degree # mentions

Closed under composition

$X.\text{follow}(R_1).\text{follow}(R_2)$

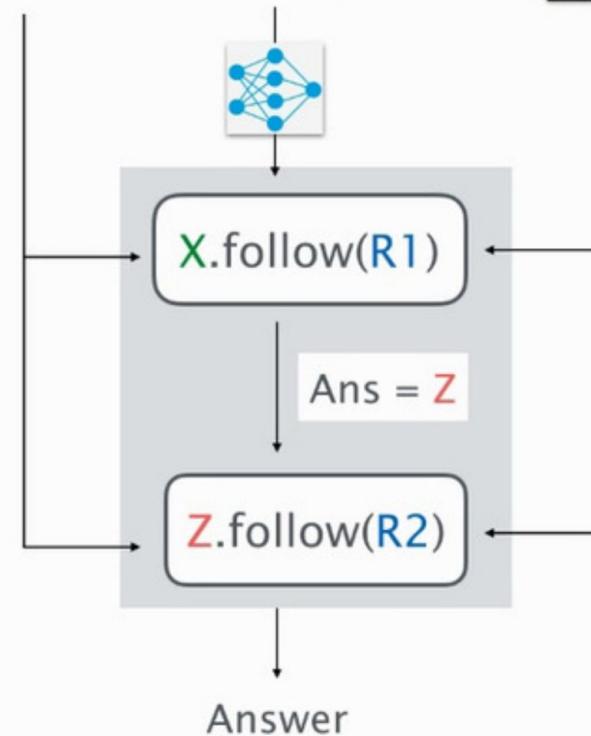
Differentiable

Who voices the dog in
the TV show Family Guy?



X = Soft-Entity Linking(Q)

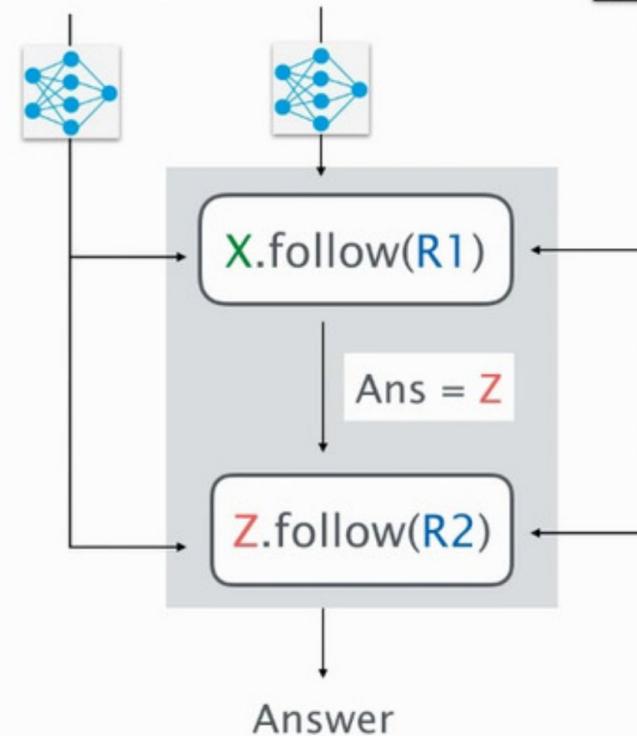
Who voices the dog in
the TV show Family Guy?



X = Soft-Entity Linking(Q)

$R1$ = Transformer-1(Q)

Who voices the dog in
the TV show Family Guy?

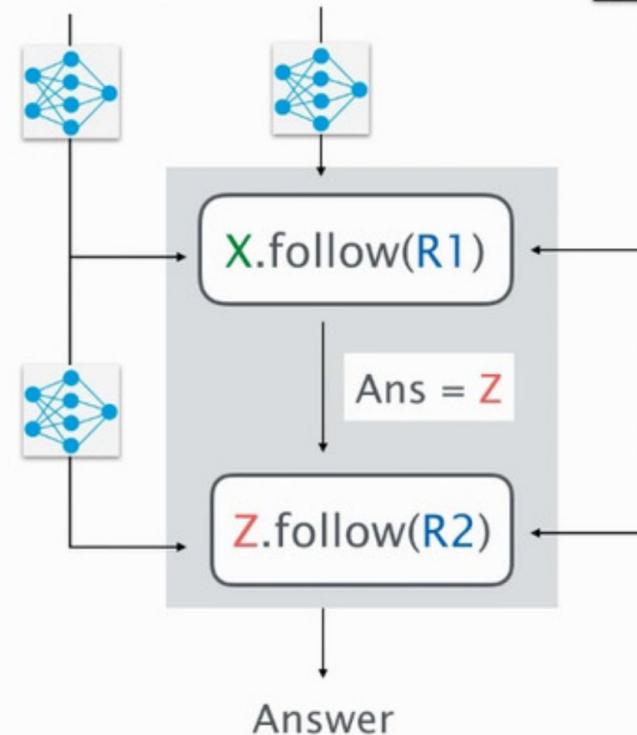


X = Soft-Entity Linking(Q)

$R1$ = Transformer-1(Q)

$R2$ = Transformer-2(Q)

Who voices the dog in
the TV show Family Guy?



$X = \text{Soft-Entity Linking}(Q)$

$R1 = \text{Transformer-1}(Q)$

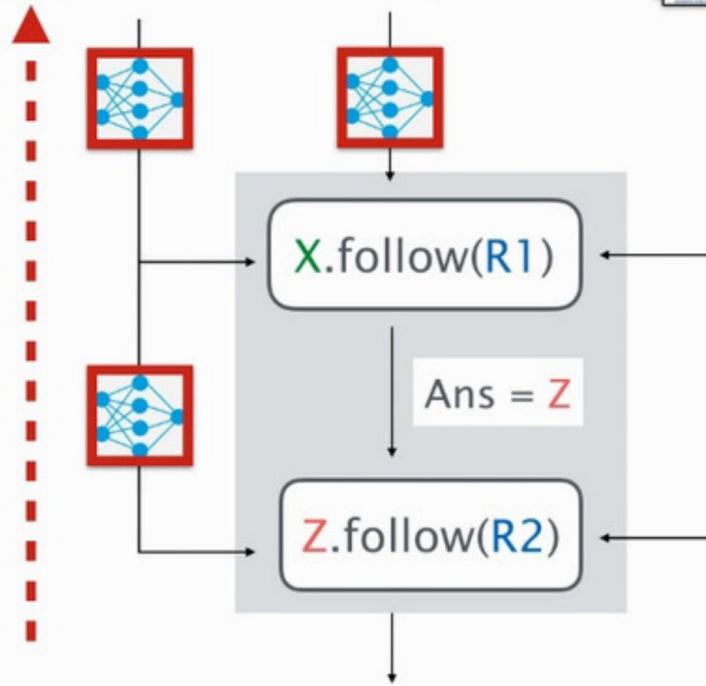
$R2 = \text{Transformer-2}(Q)$

Who voices the dog in
the TV show Family Guy?

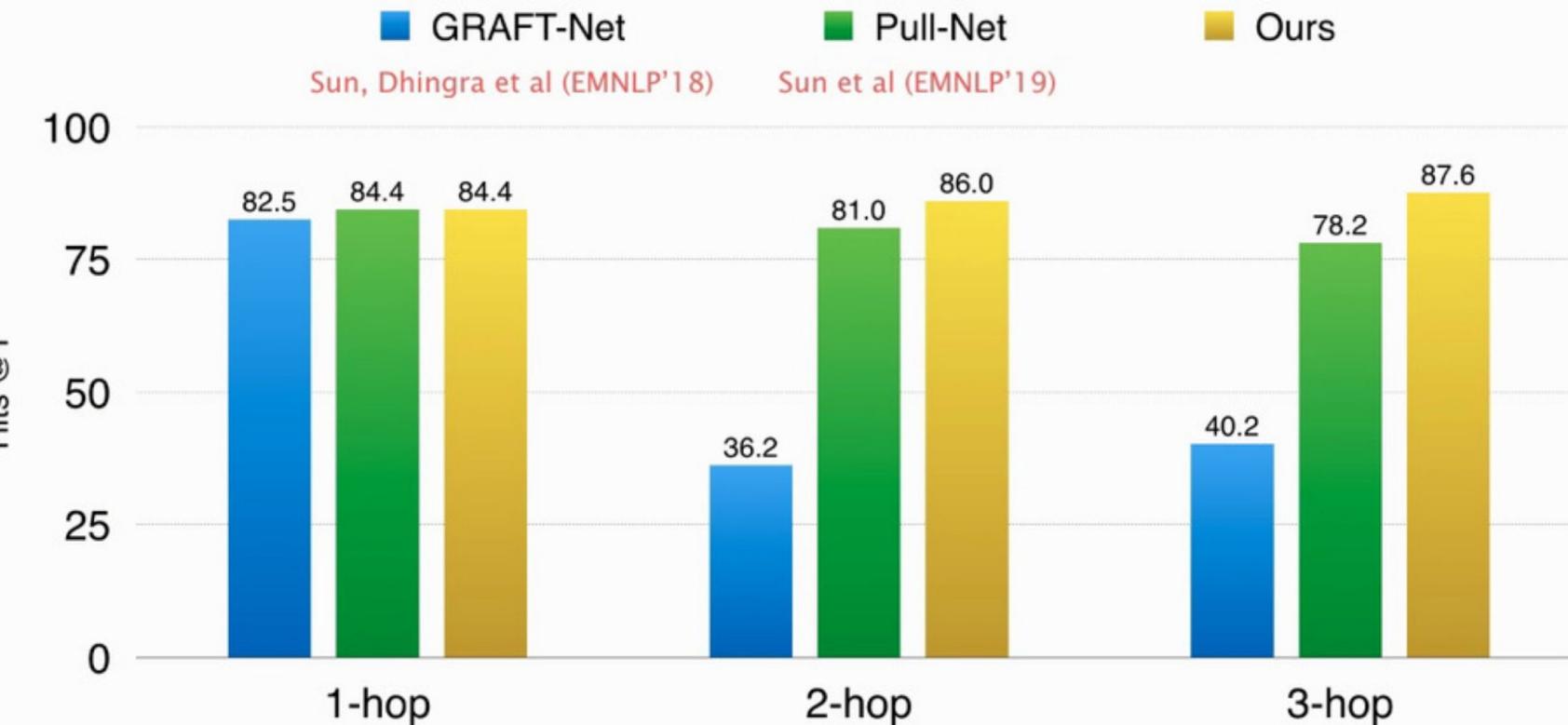


End-to-End!

Answer

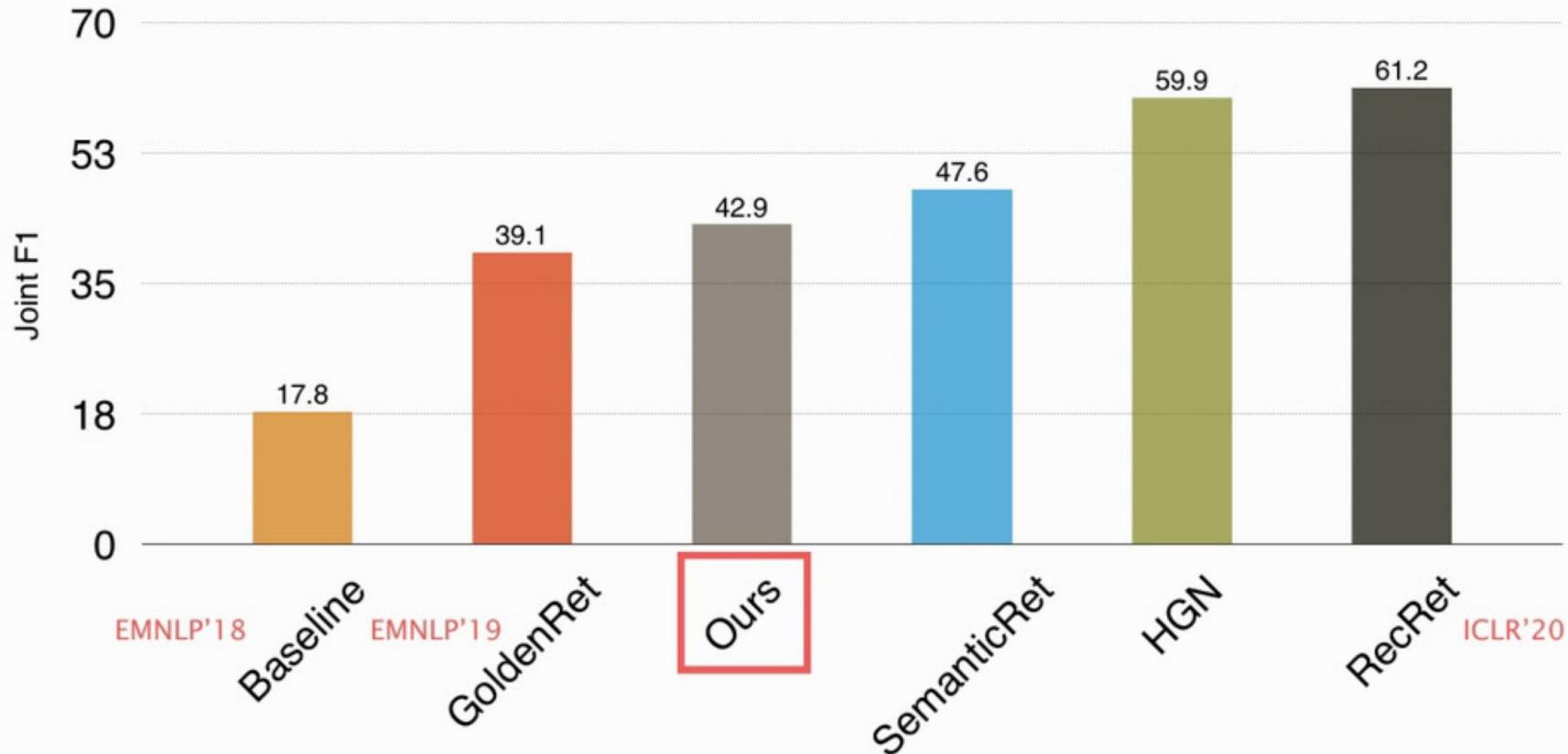


MetaQA* Results

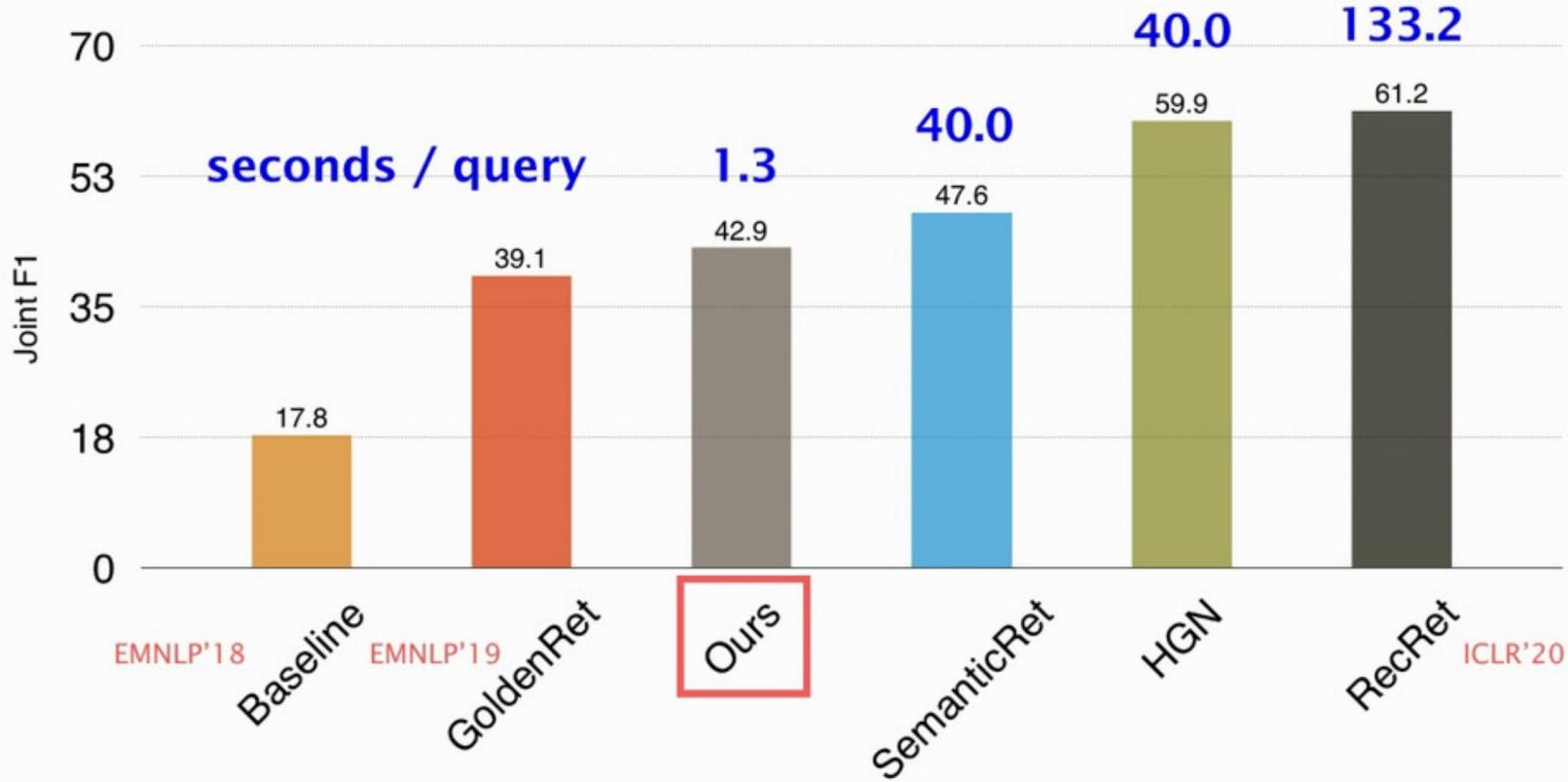


* Zhang, Dai et al, *Variational Reasoning for Question Answering with Knowledge Graph*, in AAAI 2018.

HotpotQA* Results

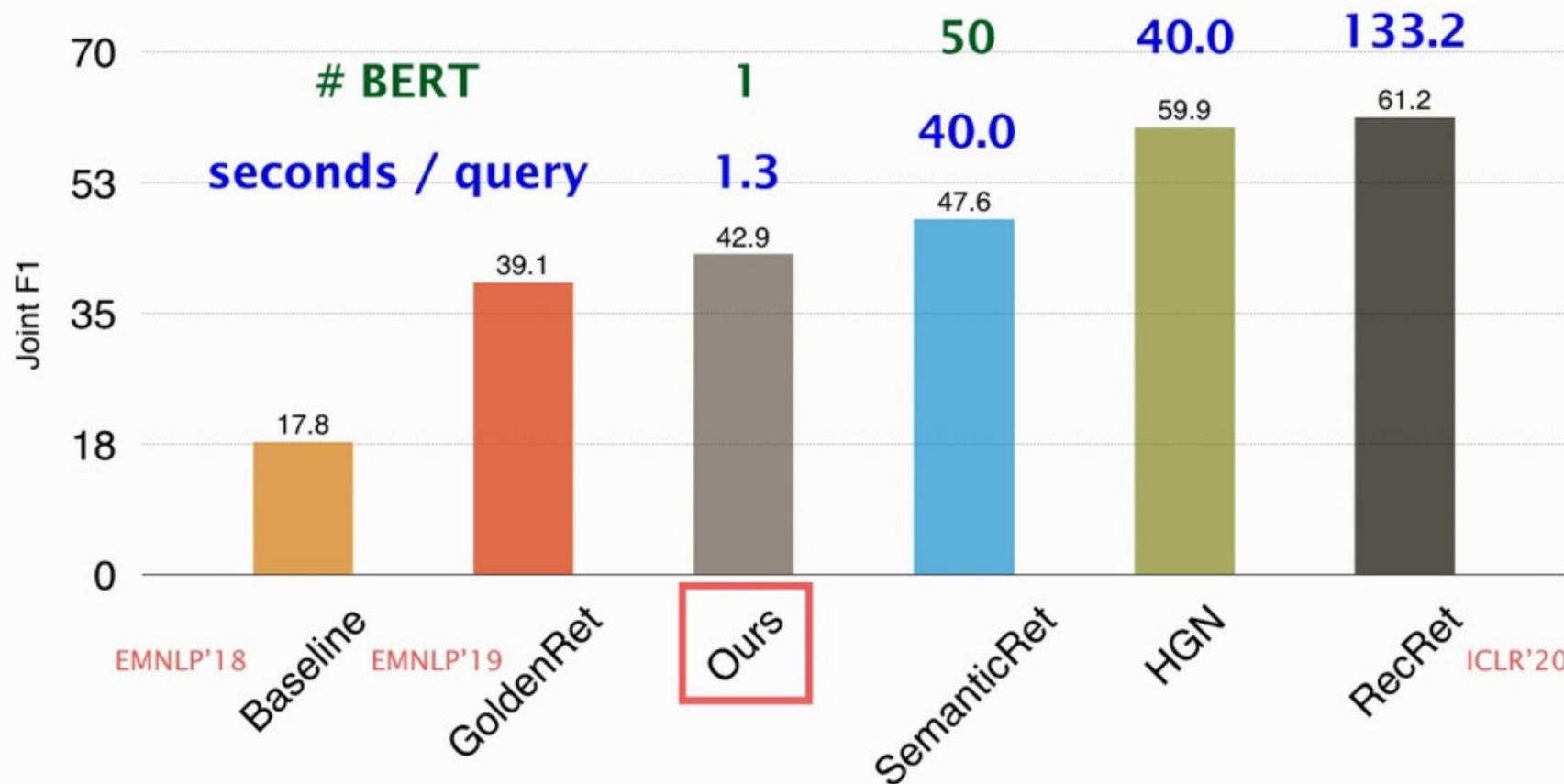


HotpotQA* Results



* Yang, Qi, Zhang et al, *HotpotQA: A dataset for diverse, explainable multi-hop question answering*, in EMNLP 2018.

HotpotQA* Results



Question

What is the shape of the family of viruses containing Coronavirus?

SUBMIT

What is the shape of the family of viruses containing Coronavirus?

SUBMIT

spherical

Coronavirus **Coronaviruses** are species of **virus** belonging to the subfamily " **Coronavirinae** " in the family " **Coronaviridae** " , in the order **Nidovirales** . Coronaviruses are **enveloped virus** es with a **positive-sense** single-stranded **RNA genome** and with a nucleocapsid of helical symmetry . The **genomic size** of coronaviruses ranges from approximately 26 to 32 **kilobases** , the largest for an RNA virus .

Coronaviridae **Coronaviridae** is a family of enveloped , positive-stranded **RNA** viruses . The viral genome is 26 – 32 kb in length . **Virions** are spherical , 120 – 160 nm across

Coronavirinae **Coronavirinae** is one of two subfamilies in the family **Coronaviridae** . It is subdivided into the genera alphacoronavirus , **betacoronavirus** , gammacoronavirus , and deltacoronavirus . These include phylogenetically compact genogroups of enveloped , positive-sense , single-stranded RNA and with a nucleocapsid of helical symmetry . The genomic size of coronaviruses ranges from approximately 26 to 32 kilobases , extraordinarily large for an RNA virus . Their numbers are growing rapidly with several novel coronaviruses being recently discovered including Middle East respiratory syndrome coronavirus MERS-CoV discovered in 2012 .

Nidovirales **Nidovirales** is an order of **virus** es with **animal** and human hosts (MERS-CoV and SARS-CoV) . It includes the families " **Coronaviridae** " , " **Arteriviridae** " , " **Roniviridae** " and " **Mesoniviridae** " .

Sense (molecular biology) In **molecular biology** and **genetics** , the **sense** of **nucleic acid** molecules (often **DNA** or **RNA**) is the nature of their roles and their complementary molecules ' nucleic acid units ' roles in specifying amino acids . Depending on the context within molecular biology , sense may have slightly different meanings .

Question

The Church of St George is in an Ethiopian city famous for what?

SUBMIT

Analyzing 5,233,329 Wikipedia articles ...

Question

The Church of St George is in an Ethiopian city famous for what?

SUBMIT

monolithic rock-cut churches

Church of
Saint
George,
Lalibela

The Church of St. George (Amharic :) is one of eleven **rock-hewn monolithic church es** in **Lalibela** , a city in the **Amhara Region of Ethiopia** . Originally named " Röha " (" Warwar ") , the historical and religious site was named Lalibela after the King **Gebre Mesqel Lalibela** of the **Zagwe dynasty** , who commissioned its construction . He is regarded as a saint by the **Ethiopian Orthodox Tewahedo Church** .

Lalibela

Lalibela (Amharic :) is a town in **Amhara Region** , northern **Ethiopia** famous for **monolithic rock-cut churches** . The whole of Lalibela offers an exceptional testimony to the medieval and post-medieval civilization of **Ethiopia** . Lalibela is one of Ethiopia 's holiest cities , second only to **Aksum** , and a center of **pilgrimage** . Unlike Aksum , the population of Lalibela is almost completely **Ethiopian Orthodox Christian** . Ethiopia is one of the earliest nations to adopt Christianity in the first half of the fourth century , and its historical roots date to the time of the Apostles . The churches themselves date from the seventh to thirteenth centuries , and are traditionally dated to the reign of the Zagwe dynasty king **Gebre Mesqel Lalibela** (r. ca. 1181 – 1221 AD) .

Ethiopia

Ethiopia (; Amharic : , " ኢትዮጵያ " ,) , officially the Federal Democratic Republic of Ethiopia (ዲሞክራሲያዊ ሪፐብሊክ , " ye ኢትዮጵያ Fédéralawi Démokrasiyawi Ripebilik ") , is a **country** located in the **Horn of Africa** . It shares borders with **Eritrea** to the north and northeast , **Djibouti** and **Somalia** to the east , **Sudan** and **South Sudan** to the west , and **Kenya** to the south . With over 100 million inhabitants , Ethiopia is the most populous **landlocked country** in the world , as well as the second-most populous nation on the African continent . It occupies a total area of 1,100,000 km² , and its capital and largest city is **Addis Ababa** .

Pilgrimage

A **pilgrimage** is a **journey** or search of **moral** or **spiritual** significance . Typically , it is a journey to a **shrine** or other location of importance to a person 's **belief**s and **faith** , although sometimes it can be a metaphorical journey into someone 's own beliefs . Many religions attach spiritual importance to particular places : the place of birth or death of founders or saints , or to the place of their " calling " or spiritual awakening , or of their connection (visual or verbal) with the divine , to locations where miracles were performed or witnessed , or locations where a deity is said to live or be " housed " , or any site that is seen to have special spiritual powers . Such sites may be commemorated with shrines or temples that devotees are encouraged to visit for their own spiritual benefit : to be healed or have questions answered or to achieve some other spiritual benefit . A person who makes such a journey is called a **pilgrim** . As a common human experience , pilgrimage has been proposed as a Jungian archetype by **Wallace Clift** and **Jean Dalby Clift** .

Amhara
Region

Amhara (Amharic : አማራ) is one of the nine **ethnic divisions** (" killoch ") of **Ethiopia** , containing the homeland of the **Amhara people** . Previously known as Region 3 , its capital is **Bahir Dar** . Ethiopia 's largest inland body of water , **Lake Tana** , which is the source of the **Blue Nile** river , is located within Amhara . The region also contains the **Semien Mountains National Park** , which includes **Ras Dashan** , the highest point in Ethiopia . Amhara is bordered by the state of **Sudan** to the west and northwest , and in other directions by other regions of Ethiopia : **Tigray** to the north , **Afar** to the east , **Benishangul-Gumuz** to the west and southwest , and **Oromia** to the south .

- The sparse matrices and dense index form a *Virtual KB*

- The sparse matrices and dense index form a *Virtual KB*
- Our method learns a (soft) semantic parser from denotations

- The sparse matrices and dense index form a *Virtual KB*
- Our method learns a (soft) semantic parser from denotations
- Future work should explore more expressive logical forms

- The sparse matrices and dense index form a *Virtual KB*
- Our method learns a (soft) semantic parser from denotations
- Future work should explore more expressive logical forms

Thank You.

Code: <http://www.cs.cmu.edu/~bdhingra/pages/drkit.html>

Carnegie
Mellon
University



Language
Technologies
Institute

ML
MACHINE LEARNING
DEPARTMENT

The logo features a cluster of colored dots in blue, green, yellow, and red, followed by the text "Google AI".