

CS 563: Question Answering

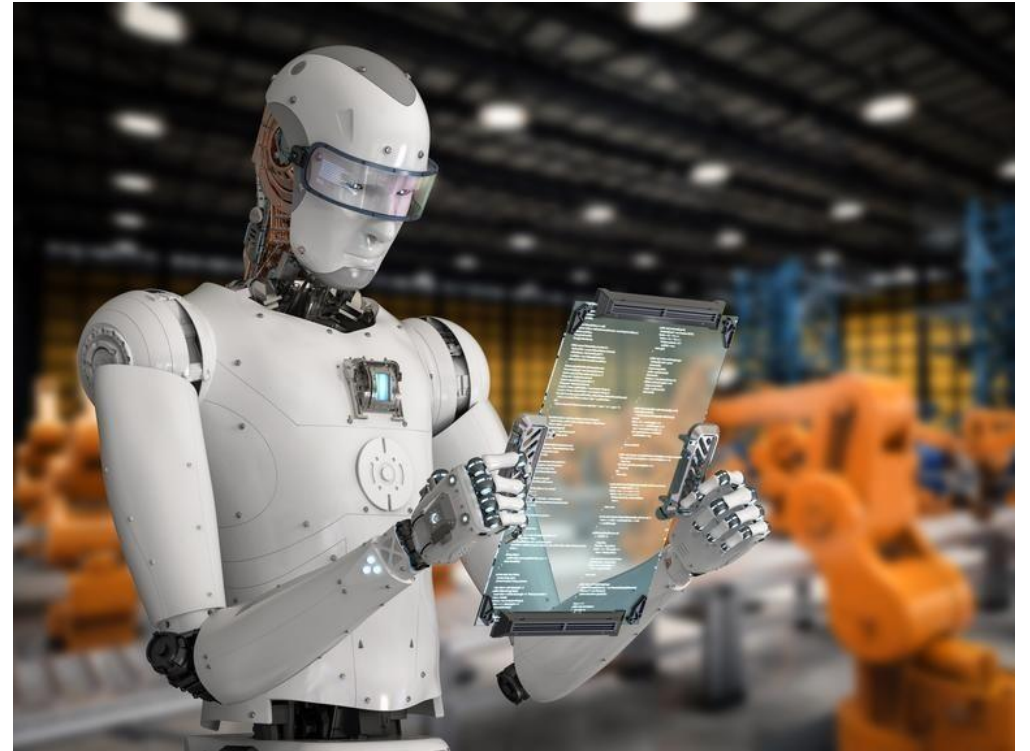
Asif Ekbal

Department of Computer Science
and Engineering

IIT Patna, Patna

Email:

asif.ekbal@gmail.com, asif@iitp.ac.in



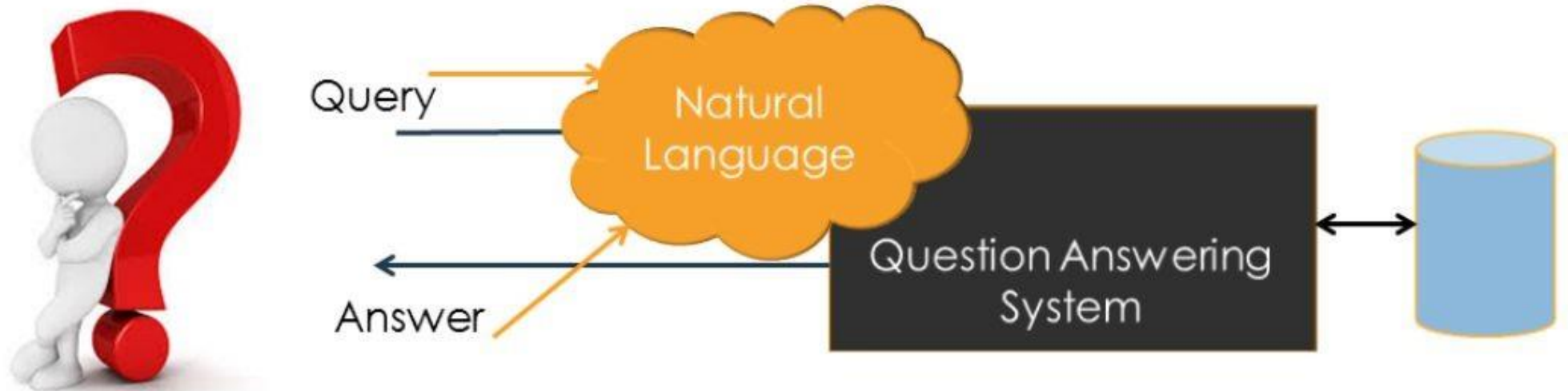
Outline

- Definition and Background
- IR based Approach for QA
 - Motivation and History
 - AskMSR: A shallow approach
 - Common Evaluation Metrics
- Machine Comprehension
 - Motivation and History
 - MC Datasets
 - Machine Learning Approach
 - Sliding Window
 - Logistic Regression
 - Deep Learning Approach
 - Stanford Attentive Reader
 - Stanford Attentive Reader++
 - BiDAF
- References

Question Answering

What is question answering?

Systems that automatically answer questions posed by humans in natural language query.



Search Engine: Stepping towards QA!



[All](#) [Maps](#) [News](#) [Images](#) [Books](#) [More](#) [Settings](#) [Tools](#)

About 8,79,000 results (0.83 seconds)

Telangana / Capital

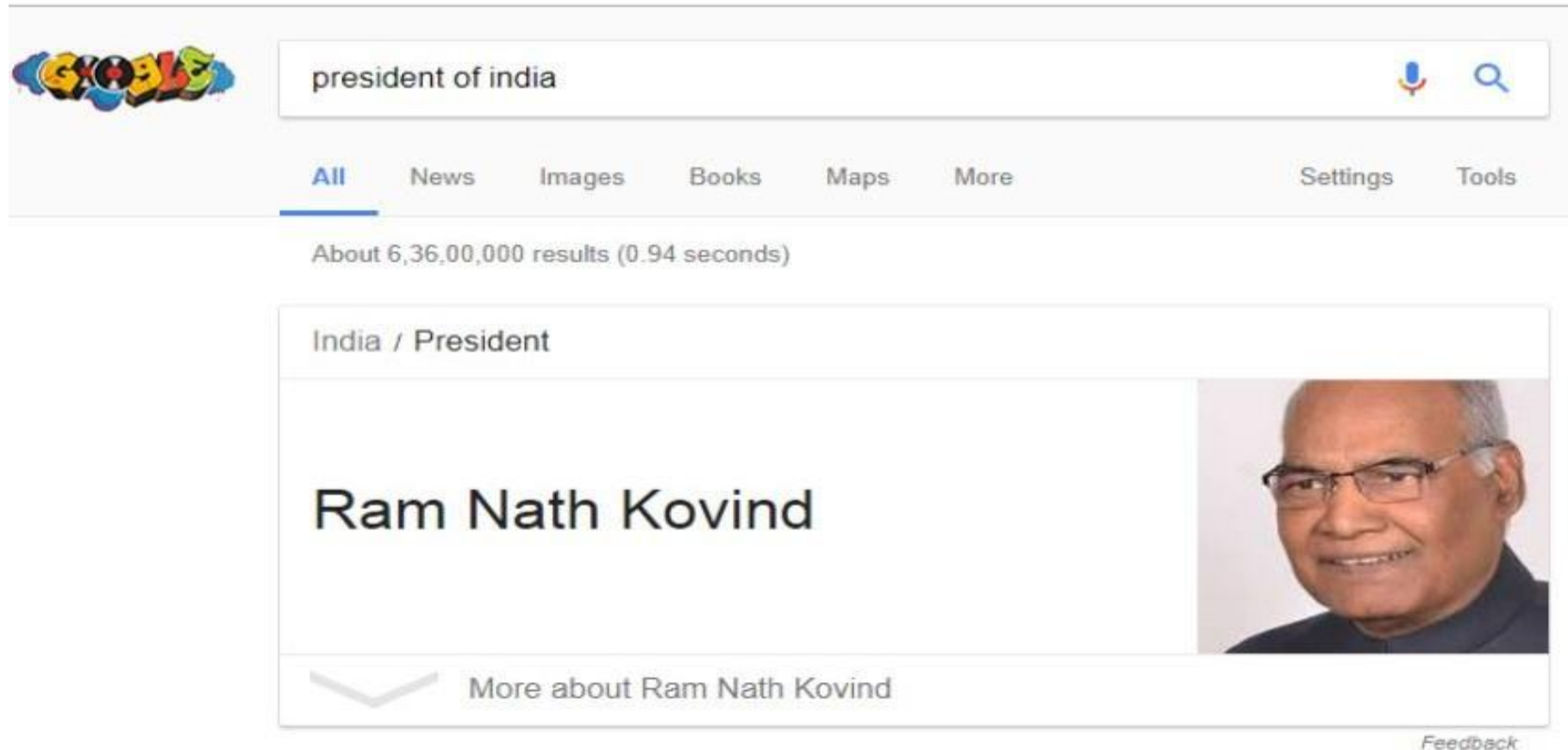


Hyderabad

Plan a trip and points of interest

[Feedback](#)

Search Engine: Stepping towards QA!



Search Engine: Stepping towards QA!

United Nations ka headquarters kha hai?

All News Maps Images Videos More Settings Tools

About 8,12,000 results (0.93 seconds)

United Nations Office at Geneva - Wikipedia

https://en.wikipedia.org/wiki/United_Nations_Office_at_Geneva

The United Nations Office at Geneva (UNOG) is the second-largest of the four major office sites ... UN specialized agencies and other UN entities with offices in Geneva hold bi-weekly briefings at the Palais des Nations, organized by the United ...

Country: Switzerland Construction started: 1929

Town or city: Geneva

Constituent agencies · Directors-General · Administrative history

List of United Nations organizations by location - Wikipedia

https://en.wikipedia.org/wiki/List_of_United_Nations_organizations_by_location

... the second most important UN centre, after the United Nations Headquarters. While the Secretariat of the United Nations is headquartered in New York City, its many bodies, ...

Missing: kha | Must include: kha

Locations · Europe · North America

United Nations - Wikipedia

https://en.wikipedia.org/wiki/United_Nations

The United Nations (UN) is an intergovernmental organization tasked to promote international ... The headquarters of the UN is in Manhattan, New York City, and is subject to extraterritoriality. Further main offices are situated in Geneva, Nairobi ...

Missing: kha | Must include: kha

where is the headquarter of united nations?

All News Maps Images Videos More Settings Tools

About 5,96,00,000 results (1.33 seconds)

United Nations / Headquarters



New York City, New York, United States

People also search for

View 15+ more



New York



United States of America



Manhattan



Brooklyn



Los Angeles



London



Washington, D.C.

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!
- IBM's Watson is a Question Answering System

Jeopardy!

- Jeopardy! is an American television quiz competition in which contestants are presented with general knowledge clues in the form of *answers*, and must phrase their responses in the form of *questions*.
- The original daytime version debuted on NBC on March 30, 1964,

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!
- IBM's Watson is a Question Answering System.

Watson's performance

- With the answer: "You just need a nap. You don't have this sleep disorder that can make sufferers nod off while standing up," Watson replied, "What is narcolepsy?"

Narcolepsy: a condition characterized by an extreme tendency to fall asleep whenever in relaxing surroundings

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!
- IBM's Watson is a Question Answering system

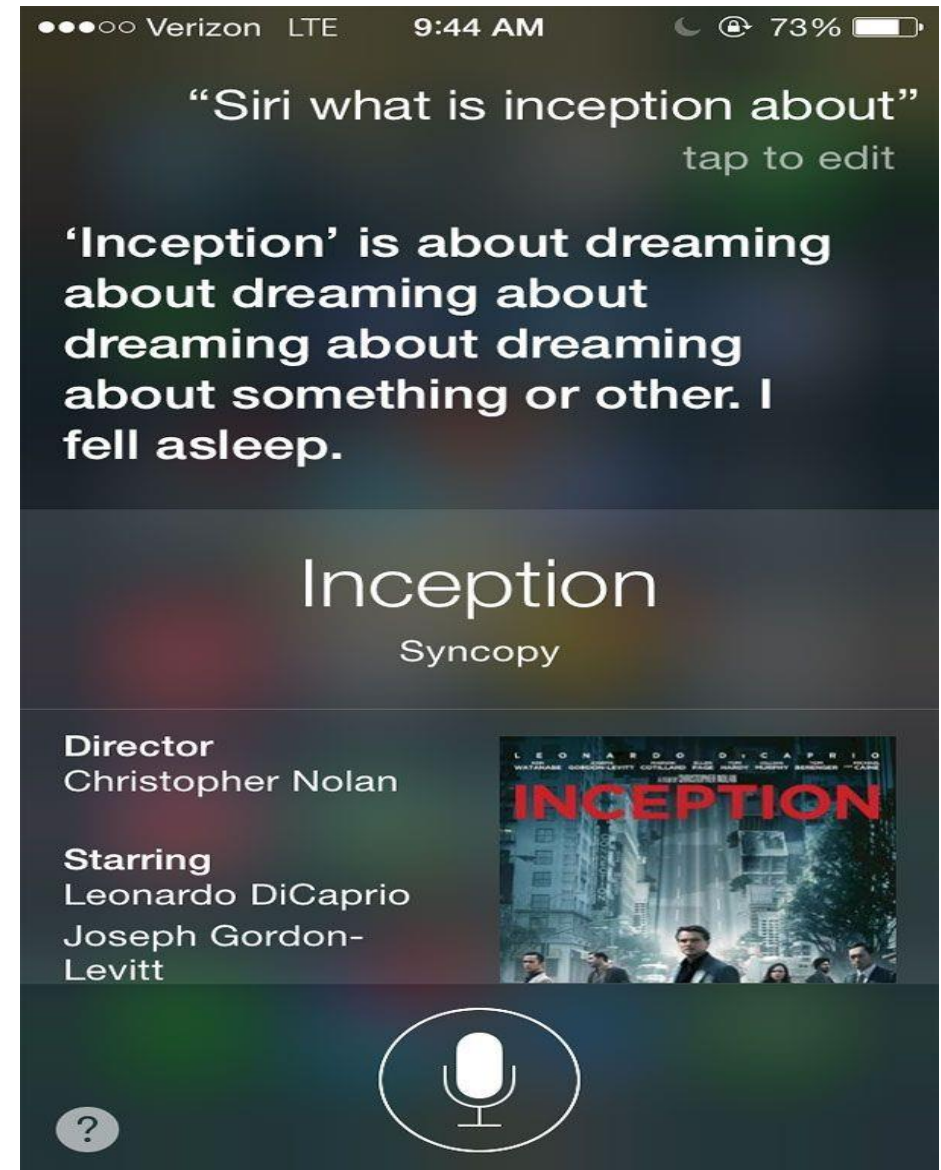
- The winning reply!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL



Bram Stoker

Apple Siri



WolframAlpha: *from Wikipedia*

WolframAlpha (/ˈwʊlf.rəm-/ *WUULf-rəm-*) is a computational knowledge engine and [answer engine](#) developed by [Wolfram Research](#).^[3] It answers factual queries directly by computing the answer from externally sourced data.^{[4][5]}

WolframAlpha was released on May 18, 2009, and is based on Wolfram's earlier product [Wolfram Mathematica](#), a computational platform for calculation, visualization, and statistics capabilities.^[1] Additional data is gathered from both academic and commercial websites such as the CIA's [The World Factbook](#), the [United States Geological Survey](#), a Cornell University Library publication called *All About Birds*, [Chambers Biographical Dictionary](#), [Dow Jones](#), the [Catalogue of Life](#),^[3] [CrunchBase](#),^[6] [Best Buy](#),^[7] and the [FAA](#).^[8]



Japanese: 自然言語処理 (common noun)

Other notations: [Enlarge](#) [Data](#) [Customize](#) [Plain Text](#)

Internet domains

naturallanguageprocessing.com | naturallanguageprocessing.net | naturallanguageprocessing.org

Sources

POWERED BY THE WOLFRAM LANGUAGE

[Company fundamentals provided by [Finnhub Stock API](#).]

Related Queries:

= repulsively

= how many words can ...

= language

= gray

= translator

typical translation leng...

Assuming any type of pie, banana cream | Use pie, banana cream, prepared from recipe or pie, banana cream, no-bake type, prepared from this instead



Have a question about using Wolfram|Alpha?

[Contact Pro Premium](#)[Expert Support >>](#)[Give us your feedback >>](#)[Pro](#) [Web Apps](#) [Mobile Apps](#) [Products](#) [Business](#) [API & Developer Solutions](#)

Input interpretation:

[Resources & Tools](#) [About](#) [Contact](#) [Connect](#) [Facebook](#) [Twitter](#) [LinkedIn](#)

English

©2022 Wolfram Alpha LLC

[Terms](#)[Privacy](#)

pie

amount



WOLFRAM

[wolfram.com](#)[Wolfram Language](#)[Mathematica](#)[Wolfram Demonstrations](#)[Wolfram for Education](#)[MathWorld](#)

banana cream

total calories

12/2

Average result:

[Show details](#)

702 Cal (dietary Calories)

how many calories are in two slices of banana cream pie?



[Examples](#) [Random](#)

Assuming any type of pie, banana cream | Use [pie, banana cream, prepared from recipe](#) or [pie, banana cream, no-bake type, prepared from mix](#) instead

Input interpretation:

pie	amount	2 slices	total calories
	type	banana cream	

Average result:

[Show details](#)

702 Cal (dietary Calories)

Motivation

- **Conversational Agents:** Facebook (M), Apple (Siri), Google etc.
- **Google Assistant:** Ask it questions. Tell it to do things
- **Jeopardy!:** In 2011, the IBM Watson computer system competed on Jeopardy! against former winners and won the first prize
- **Biomedical and Clinical QA:** Urgent need of system that accepts the queries from medical practitioners in natural language and returns the answers quickly and efficiently from biomedical literature, EMR etc.
- **Online knowledge service:** The online service provide the answer of various question from science, mathematics etc.

Motivation and History

- Open domain QA systems received larger attention in the 90s
 - Combination of NLP and IR/IE techniques
 - One of the most famous: MIT START system
 - Wolfram Alpha
- Advanced systems use a combination of “shallow” methods together with knowledge bases and more complex NLP methods
- In the last 20 years, TREC, SemEval and ACL provided workshops and tracks for various flavor of QA tasks (closed and open-domain)

Motivation and History (cont'd...)

- Lately, a large number of new datasets and tasks have become available which have improved the performance of (open-domain) QA systems
 - VisualQA
 - Given an image and a question in natural language, provide the correct answer
 - 600,000+ questions on more than 200,000 images
 - SQuAD - Stanford QA Dataset
 - Open-domain question answering
 - 100,000+ Q-A pairs on 500+ articles
 - NewsQA dataset
 - Crowd-sourced machine reading comprehension dataset
 - 120,000 answered questions Over 12,000 news articles

Types: QA

- Single vs Multiple
- Simple vs Complex
- Text vs Visual
- Open-domain vs Closed-domain
- IR-based vs Knowledge-based

Single vs Multiple

- A **single document Q/A task** involves questions associated with one particular document
- In most cases, the assumption is that the answer appears somewhere in the document and probably once
- Applications involve searching an individual resource, such as a book, encyclopedia, or manual
- Reading comprehension tests are also a form of single document QA
ex. SQuAD

- A **multiple document Q/A** task involves questions posed against a **collection** of documents
- The answer may appear in the collection multiple times or **may not appear at all!**
- Applications include WWW search engines, and searching text repositories such as news archives, medical literature, or scientific articles

ex. MS MARCO

Simple vs Complex

- **Simple (factoid) questions (most commercial systems)**
 - Who wrote the Declaration of Independence?
 - What is the average age of the onset of autism?
 - Where is Apple Computer based?
 - ex. SQuAD
- **Complex (narrative) questions**
 - What do scholars think about Jefferson's position on dealing with pirates?
 - What is a Hajj?
 - In children with an acute febrile illness, what is the efficacy of single medication therapy with acetaminophen or ibuprofen in reducing fever?
 - ex. Narrative QA
- **Complex (opinion) questions**
 - Was the Trump/Hilary election fair?

Text vs Visual

Input : Document and Question

Output : Answer

Example

- University of Washington
- [Allennlp](#)

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

Input: Picture or Video and Question.

Output :Answer.

- What is in the image?
- Are there any humans?
- What sport is being played?
- Who has the ball?
- How many players are in the image?
- Who are the teams?
- Is it raining?

Example

- <http://vqa.cloudcv.org>



IR-based vs Knowledge-based

- Information Retrieval:
QA can be viewed as short passage retrieval
- Information Extraction:
QA can be viewed as open-domain information extraction

- Build a semantic representation of the query
 - Times, dates, locations, entities, numeric quantities
- Map from this semantics to query structured data or resources
 - Geospatial databases
 - Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
 - Restaurant review sources and reservation services
 - Scientific databases
- Examples: Siri

Web Question Answering: Is More Always Better?

Dumais, Banko, Brill, Lin, Ng (Microsoft, MIT, Berkeley)

Q: “Where is the Louvre located?”
Want “Paris”
or “France”
or “75058
Paris Cedex 01”
or a map
Don’t just
want URLs

The screenshot shows a Google search results page for the query "Where is the Louvre museum located?". The search bar at the top contains the query and a "Google Search" button. Below the search bar, a message states: "The following words are very common and were not included in your search: **Where is the** [details]". The search results are displayed in a list format, with tabs for "Web", "Images", "Groups", "Directory", and "News-Now". The first result is a PDF document titled "An Analysis of the AskMSR Question-Answering System" by Dumais et al., with a file format of PDF/Adobe Acrobat and a link to "View as HTML". The second result is a webpage titled "hotel montpensier - located near louvre museum, opera house, ..." with a link to "View as HTML". The third result is a webpage titled "hotel montpensier - located near louvre museum, opera house, ..." with a link to "View as HTML". The fourth result is a webpage titled "Louvre Museum Official Website: Publications" with a link to "View as HTML". On the right side of the page, there is a sidebar with a "Sponsored" section featuring a "Paris Metro & ... explore Paris and ... www.gotopar ... Interest: ... See your ...".

AskMSR: Shallow approach

In what year did Abraham Lincoln die?
Ignore hard documents and find easy ones

Abraham Lincoln, 1809-1865

***LINCOLN, ABRAHAM** was born near Hodgenville, Kentucky, on February 12, 1809. In 1816, the Lincoln family moved to Pigeon Creek in Perry (now Spencer) County. Two years later, Abraham Lincoln's mother died and his father married a woman his "angel" mother. Lincoln attended a formal school for only a few months but acquired knowledge through the reading of books. In 1830, he moved to Illinois, in 1830 where he obtained a job as a store clerk and the local postmaster. He served without distinction in the Black Hawk War, lost his attempt at the state legislature, but two years later he tried again, was successful, and Lincoln was admitted to the bar and became noteworthy as a witty, honest, competent circuit lawyer. He served a one-year term in the U.S. House in 1846, at which time he opposed the war with Mexico. By 1858, Lincoln had gained national attention for his series of debates with Stephen A. Douglas. After losing the election, he became a significant figure in his party. On March 4, 1861, at his inauguration on March 4, seven southern states had seceded, for a total of 11. Lincoln immediately took action to suppress the rebellion. He called for 75,000 volunteers (approximately 40,000 were accepted), and he issued the Emancipation Proclamation which expanded the purpose of the war to the abolition of slavery. The dedication of a national cemetery in Gettysburg, Lincoln's explanation of the war, and his final address to the nation are among his most famous speeches. Lincoln's death on April 15, 1865, at Ford's Theatre in Washington, D.C., was a national tragedy.



Sixteenth President
1861-1865
Married to Mary Todd Lincoln



ABRAHAM LINCOLN

**Sixteenth President
of the United States**

Born in 1809 - Died in 1865

Abraham Lincoln

16th President of the United States (March 4, 1861 to April 15, 1865)

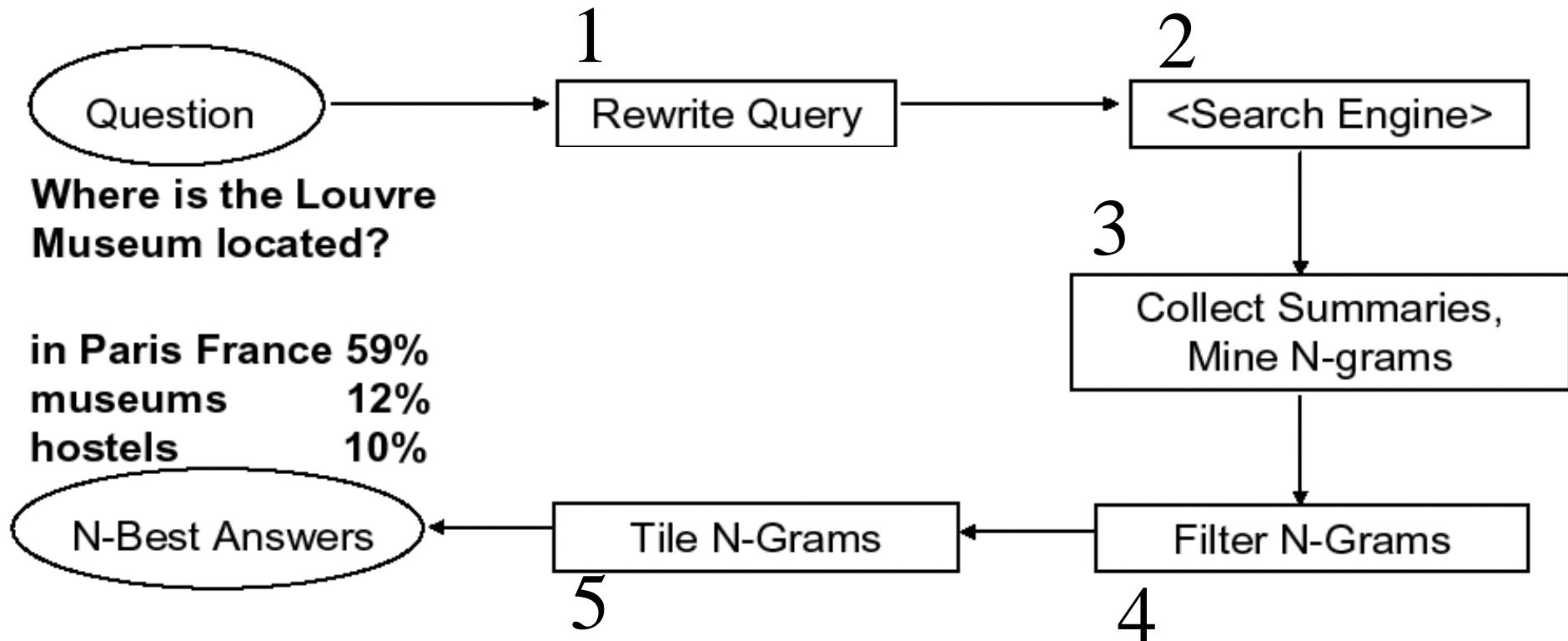
Born: February 12, 1809, in Hardin County, Kentucky

Died: April 15, 1865, at Petersen's Boarding House in Washington, D.C.

"I was born February 12, 1809, in Hardin County, Kentucky. My parents were both born in Virginia, of undistinguished families, perhaps I should say. My mother, who died in my tenth year, was of a family of the name of Hanks."



AskMSR: Details



Step 1: Rewrite queries

Intuition: The user's question is often syntactically quite close to sentences that contain the answer

Where is the Louvre Museum located?

The Louvre Museum is located in ***Paris***

Who created the character of Scrooge?

Charles Dickens created the character of Scrooge.

Query rewriting

- Classify question into seven categories
 - Who is/was/are/were...?
 - When is/did/will/are/were ...?
 - Where is/are/were ...?

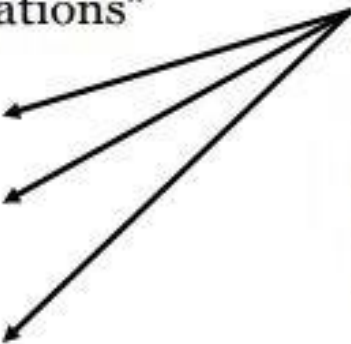
a. Category-specific transformation rules

eg “For Where questions, move ‘is’ to all possible locations”

“Where is the Louvre Museum located”

- “is the Louvre Museum located”
- “the is Louvre Museum located”
- “the Louvre is Museum located”
- “the Louvre Museum is located”
- “the Louvre Museum located is”

Nonsense,
but who
cares? It’s
only a few
more queries
to Google.



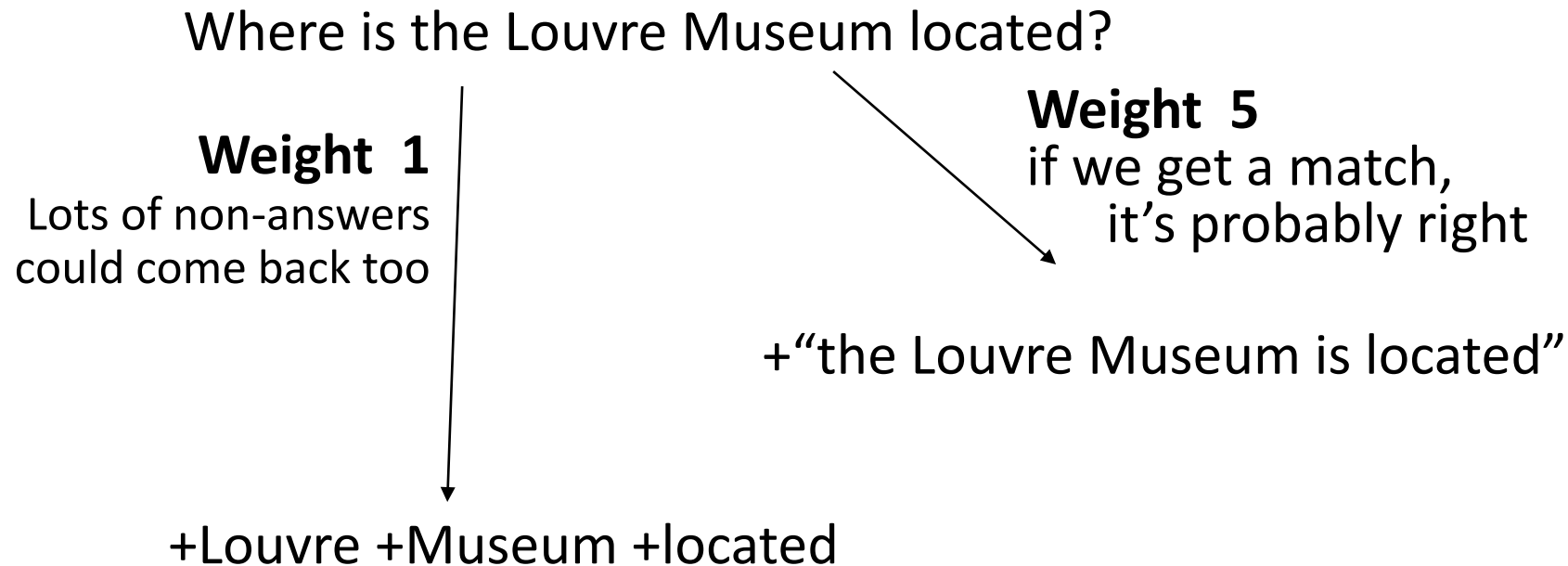
b. Expected answer “Datatype” (eg, Date, Person, Location, ...)

When was the French Revolution? → DATE



Query Rewriting: Weights

One wrinkle: Some query rewrites are more reliable than others



Step 2: Query search engine

Send all rewrites to a search engine

Retrieve top N answers (100?)

For speed, rely just on search engine's
“snippets”, not the full text of the actual
document

Step 3: Mining N-Grams

Simple: Enumerate all N-grams (N=1,2,3 say) in all retrieved snippets

Weight of an n-gram: occurrence count, each weighted by “reliability” (weight) of rewrite that fetched the document

Example: “Who created the character of Scrooge?”

- Dickens - 117
- Christmas Carol - 78
- Charles Dickens - 75
- Disney - 72
- Carl Banks - 54
- A Christmas - 41
- Christmas Carol - 45
- Uncle - 31

The returned summaries contain the query terms, usually with a few words of surrounding context. The summary text is processed in accordance with the patterns specified by the rewrites. Unigrams, bigrams and trigrams are extracted and subsequently scored according to the weight of the query rewrite that retrieved it. These scores are summed across all summaries containing the n-grams

Step 4: Filtering N-Grams

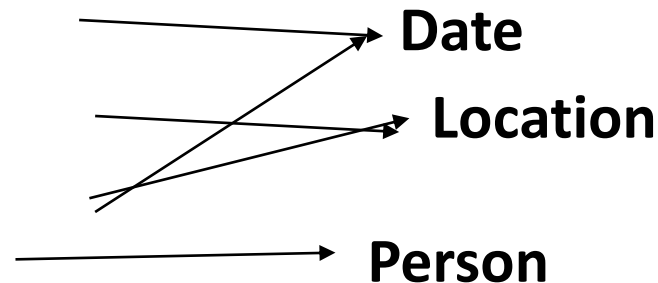
Each question type is associated with one or more
“**data-type filters**” = regular expression

When...

Where...

What ...

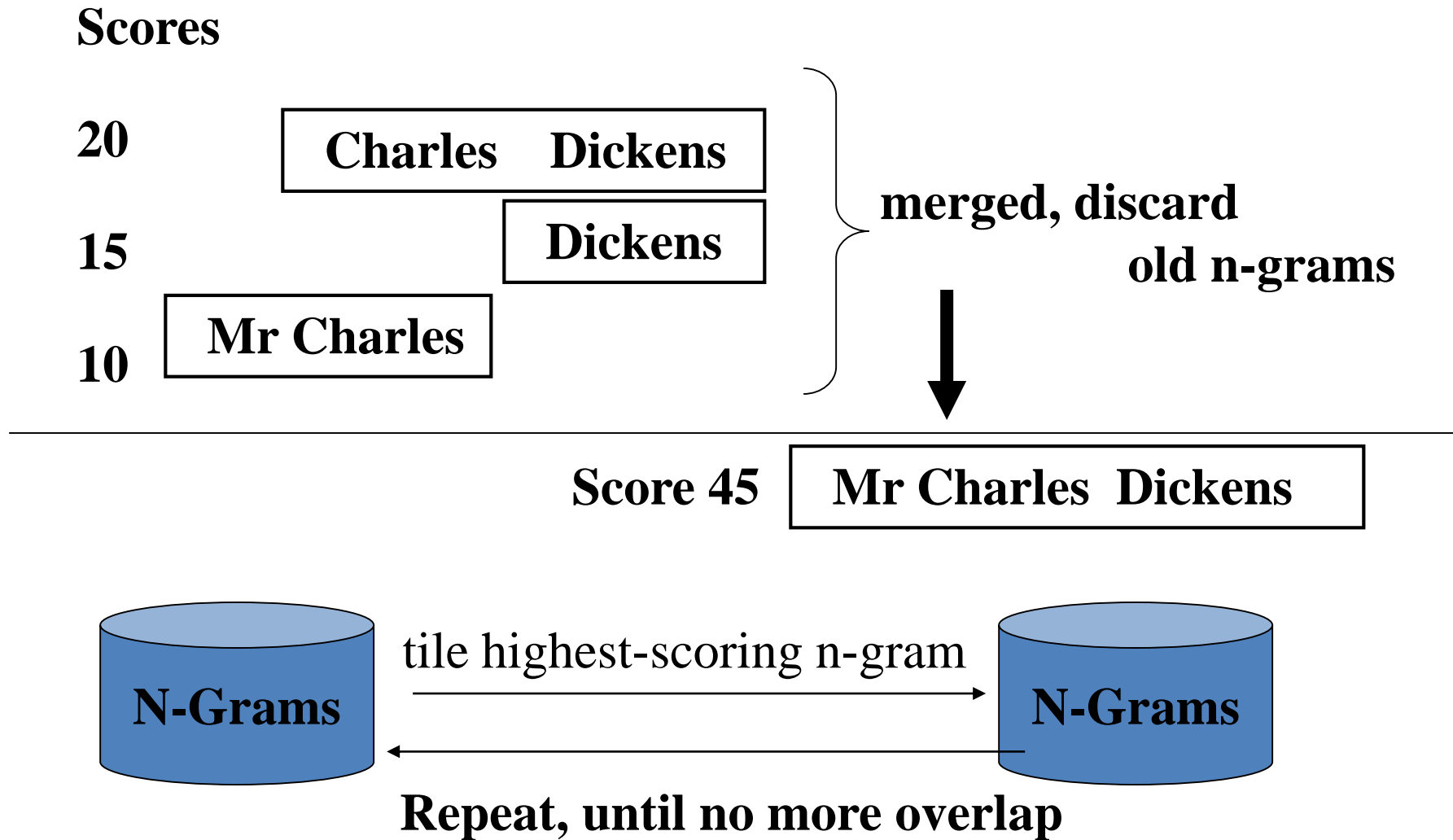
Who ...



Boost score of n-grams that do match regexp

Lower score of n-grams that don't match regexp

Step 5: Tiling the Answers



Results

Standard TREC contest test-bed:

~1M documents; 900 questions

Technique doesn't do too well (though would have placed in top 9 of ~30 participants!)

MRR = 0.262 (ie, right answered ranked about #4-#5 on average)

Why? Because it relies on the redundancy of the Web

Using the Web as a whole, not just TREC's 1M documents... MRR = 0.42 (ie, on average, right answer is ranked about #2-#3)

Issues

In many scenarios (e.g., monitoring an individual's email...) we only have a small set of documents

Works best/only for “Trivial Pursuit”-style fact-based questions

Limited/brittle repertoire of
question categories
answer data types/filters
query rewriting rules

Basic Q/A Architecture

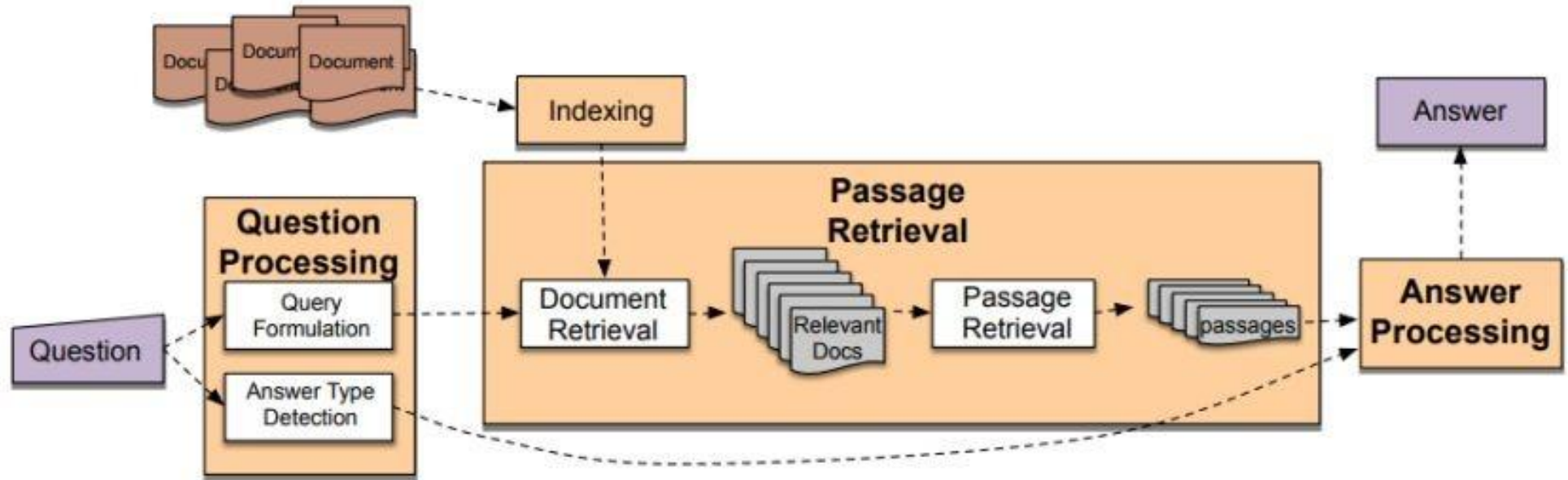


Figure: IR based question answering system. Image courtesy Jurafsky and Martin [2000]

Common Evaluation Metrics

- **Accuracy** (does answer match gold-labeled answer?)
- **Mean Reciprocal Rank:**
 - The reciprocal rank of a query response is the inverse of the rank of the first correct answer.
 - The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries Q .

(ex adapted from Wikipedia)

- 3 ranked answers for a query, with the first one being the one it thinks is most likely correct
- Given those 3 samples, we could calculate the mean reciprocal rank as $(1/3 + 1/2 + 1)/3 = 11/18$ or about 0.61.

Query	Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, cats	cats	3	1/3
torus	torii, tori , toruses	tori	2	1/2
virus	viruses , virii, viri	viruses	1	1

Machine Comprehension

- **Machine Comprehension or Machine Reading Comprehension (MRC)** is all about answering a query about a given context paragraph
- “A machine **comprehends** a passage of **text** if, for any **question** regarding that text that can be **answered** correctly by a majority of native speakers, that machine can provide a string which those speakers would agree both answers that question, and does not contain information irrelevant to that question.”
(Burges 2013)

Machine Comprehension: History

- Much early NLP work attempted reading comprehension
 - Schank, Abelson, Lehnert et al. c. 1977 – “Yale A.I. Project”
- Revived by Lynette Hirschman in 1999
 - Could NLP systems answer human reading comprehension questions for 3rd to 6th graders? **Simple methods attempted.**
- Revived again by Chris Burges in 2013 with MCTest
 - Again answering questions over simple story texts
- Floodgates opened in 2015/16 with the production of large datasets which permit supervised neural systems to be built
 - Hermann et al. (NIPS 2015) DeepMind CNN/DM dataset
 - Rajpurkar et al. (EMNLP 2016) SQuAD
 - MS MARCO, TriviaQA, RACE, NewsQA, NarrativeQA, HotpotQA

Machine Comprehension

Passage (P) + Question (Q) → Answer (A)

P

Alyssa got to the beach after a long trip. She's from Charlotte. She traveled from Atlanta. She's now in Miami. She went to Miami to visit some friends. But she wanted some time to herself at the beach, so she went there first. After going swimming and laying out, she went to her friend Ellen's house. Ellen greeted Alyssa and they both had some lemonade to drink. Alyssa called her friends Kristin and Rachel to meet at Ellen's house.....

Q

What city is Alyssa in?

A

Miami

Motivation (1)

- Teaching machines to understand human language is a long-standing challenge in AI
- Requires various aspects of text understanding
 - Part-of-speech Tagging
 - Named Entity Recognition
 - Syntactic Parsing
 - Coreference resolution
- Is there a comprehensive evaluation that can test all these aspects and probe even deeper levels of understanding?
 - **Machine Comprehension**

Motivation (2)

- Reading comprehension: tests to measure how well a human has understood a piece of text
- Machine comprehension: how well computer systems understand human language
- Machine comprehension could be the most suitable task for evaluating language understanding

Datasets

Before 2015

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

After 2015

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

QA vs. Machine Comprehension

- Reading comprehension as an instance of question answering because it is essentially a question answering problem over a short passage of text
- Question answering is to build computer systems which are able to automatically answer questions posed by humans from various sources
- Machine comprehension puts more emphasis on text understanding with answering questions regarded as a way to measure language understanding

Approaches

- **Machine Learning Approaches**

- Sliding Window (Richardson et. al, 2013)
 - Compute the unigram/bigram overlap between the sentence containing the candidate answer and the question
 - Use TF-IDF based similarity to select the best candidate answer
- Logistic Regression (Rajpurkar et. al, 2013)
 - Extract several types of features for each candidate answer
 - Features
 - Matching Word Frequencies
 - Matching Bigram Frequencies
 - Lengths
 - Span POS Tags
 -

Deep Learning Approaches

CNN/Daily Mail Datasets

- Still noisy and artificial (not real questions)
- Not hard enough for reasoning and inference
- Does it work for a real QA problem?

Stanford Question Answering Dataset (SQuAD)

- Passage + Question → Answer
 - **Passage**: selected from Wikipedia
 - **Question**: crowdsourced
 - **Answer**: must be a span in the passage

Extractive Question Answering

Stanford Question Answering Dataset (SQuAD)

- Passage + Question → Answer

Who did **Genghis Khan** **unite** before he began **conquering** the rest of **Eurasia**?

He came to power by **uniting** many of the nomadic tribes of Northeast Asia. **After** founding the Mongol Empire and being proclaimed "**Genghis Khan**", he started the Mongol invasions that resulted in the **conquest** of most of **Eurasia**. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezm Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

Stanford Question Answering Dataset (SQuAD)

- Passage + Question → Answer

Who did **Genghis Khan** unite before he began **conquering** the rest of **Eurasia**?

He came to power by **uniting** many of the nomadic tribes of Northeast Asia. **After** founding the Mongol Empire and being proclaimed "**Genghis Khan**", he started the Mongol invasions that resulted in the **conquest** of most of **Eurasia**. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezmid Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

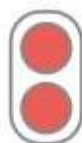
Deep Learning Approaches

Stanford Attentive Reader

Bidirectional LSTMs

Q

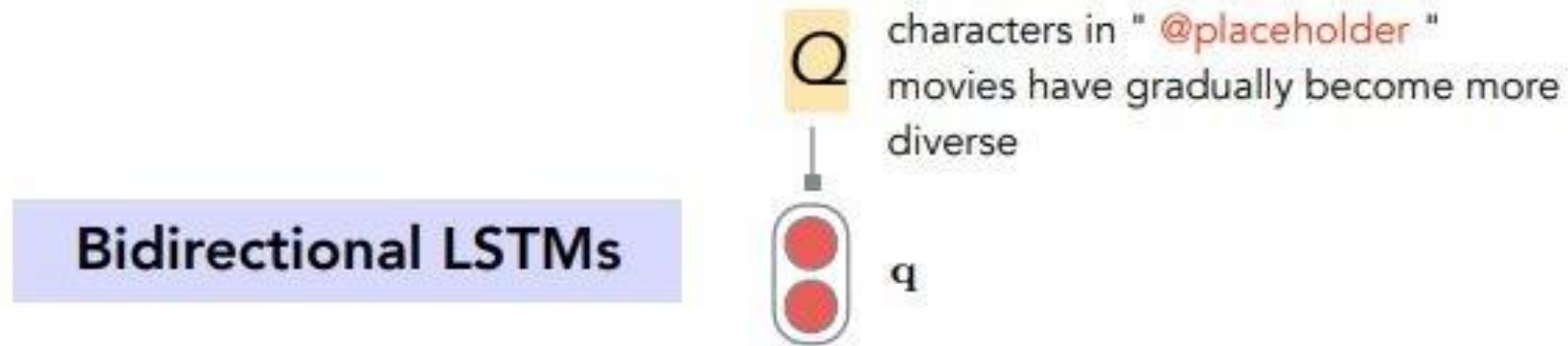
characters in " @placeholder "
movies have gradually become more
diverse



q

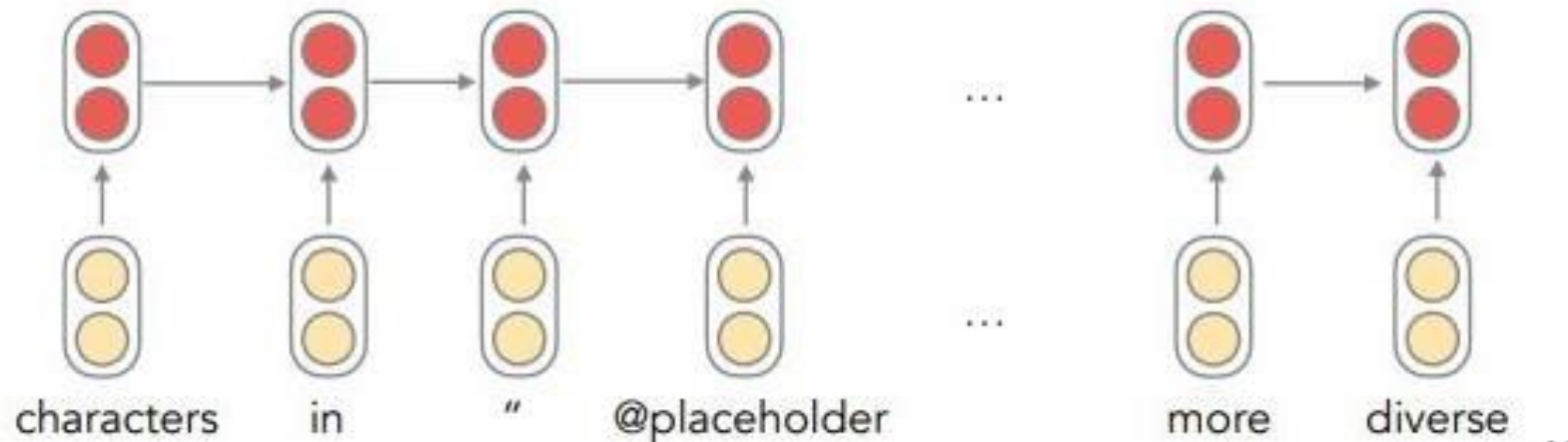
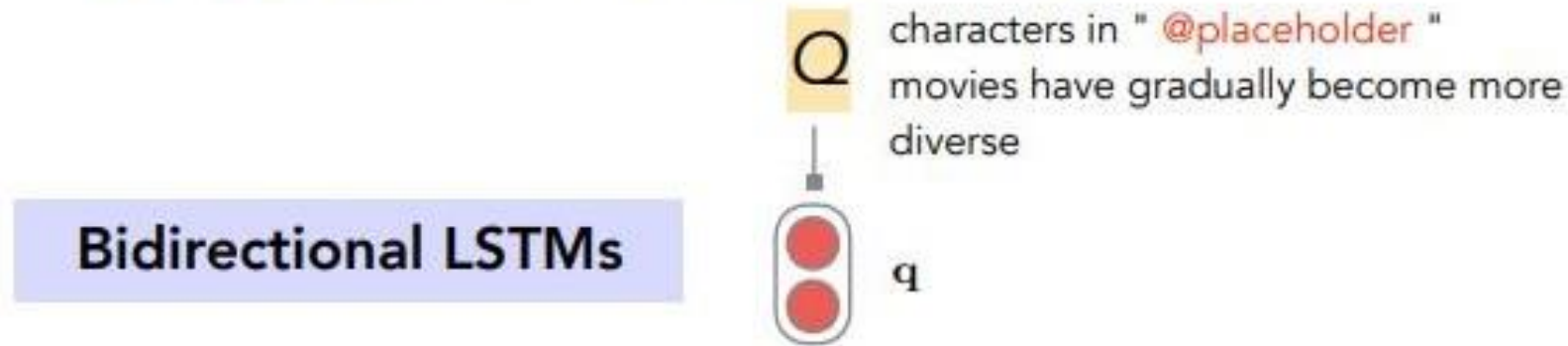
Deep Learning Approaches

Stanford Attentive Reader



Deep Learning Approaches

Stanford Attentive Reader



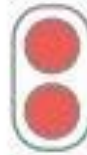
Deep Learning Approaches

Stanford Attentive Reader

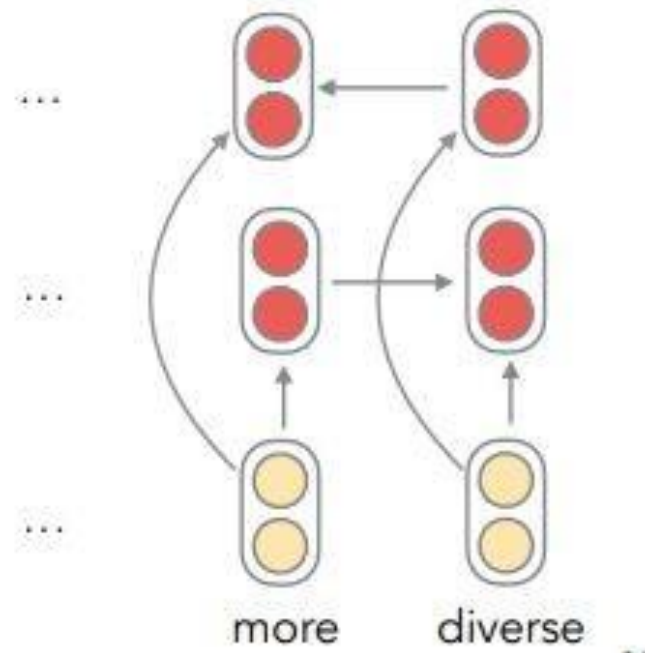
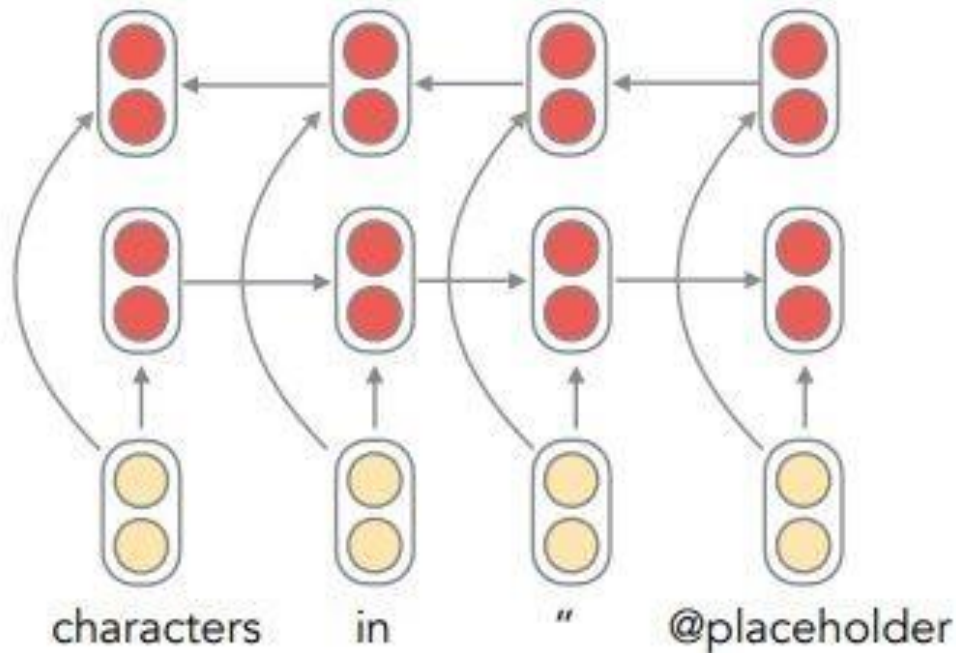
Bidirectional LSTMs

Q

characters in " @placeholder "
movies have gradually become more
diverse

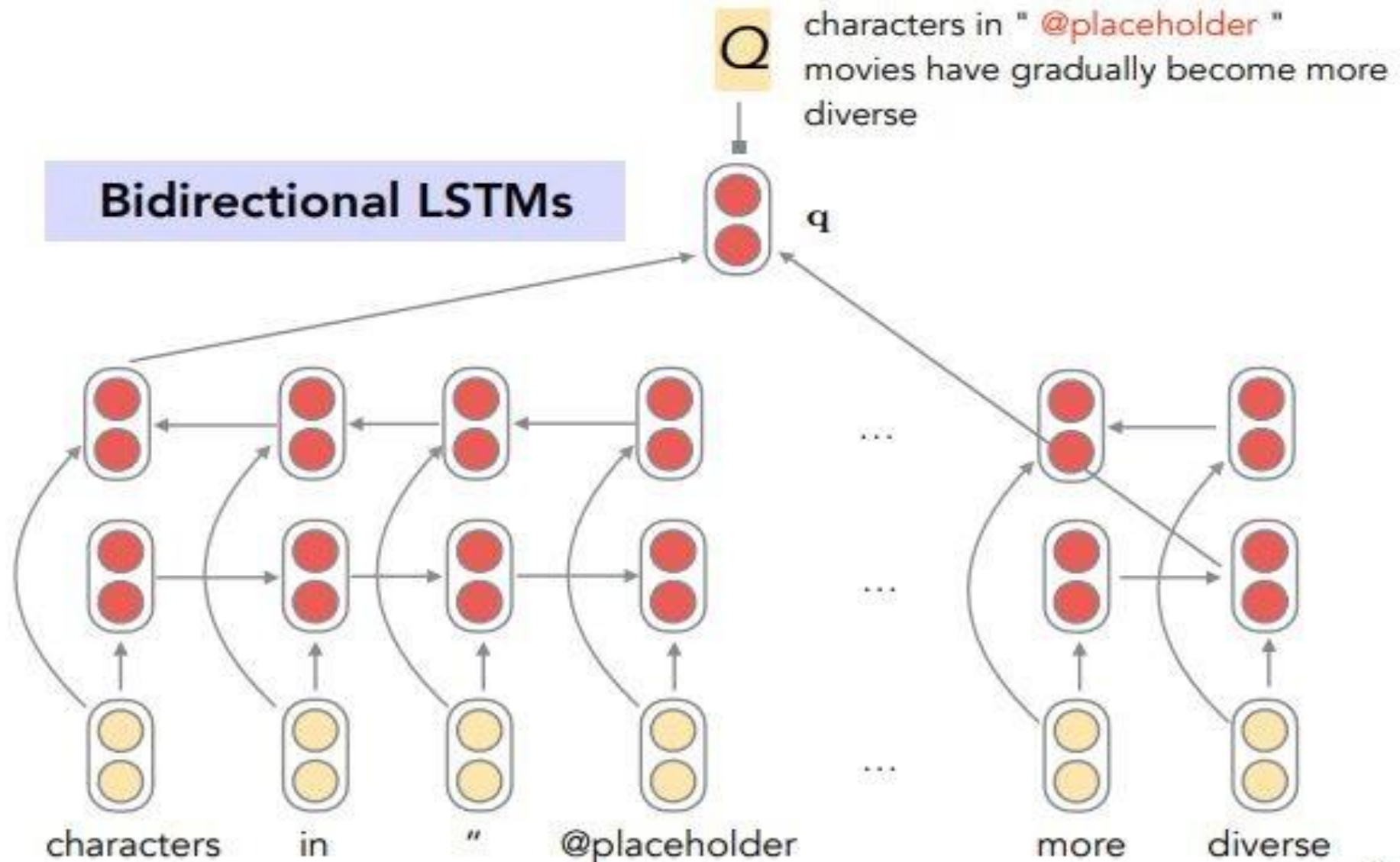


q



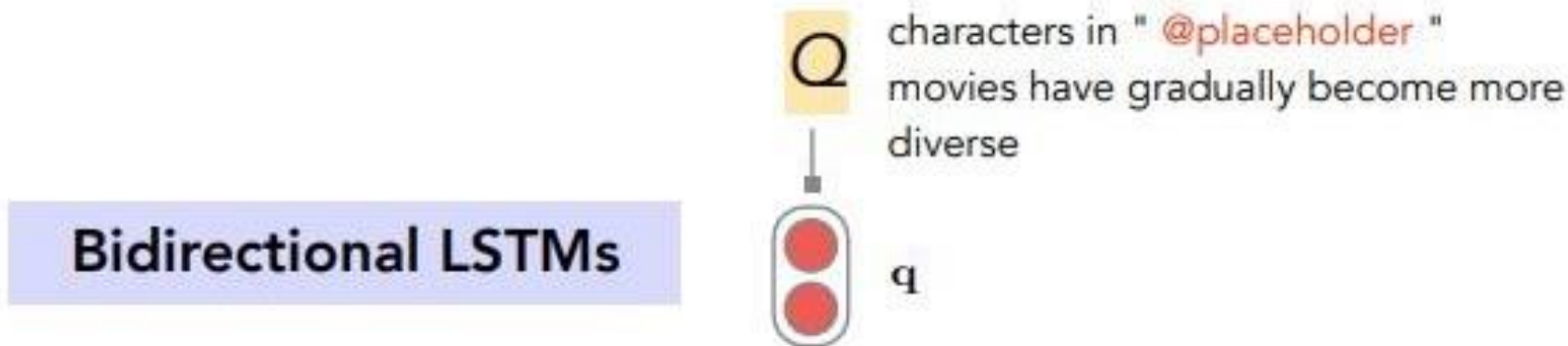
Deep Learning Approaches

Stanford Attentive Reader



Deep Learning Approaches

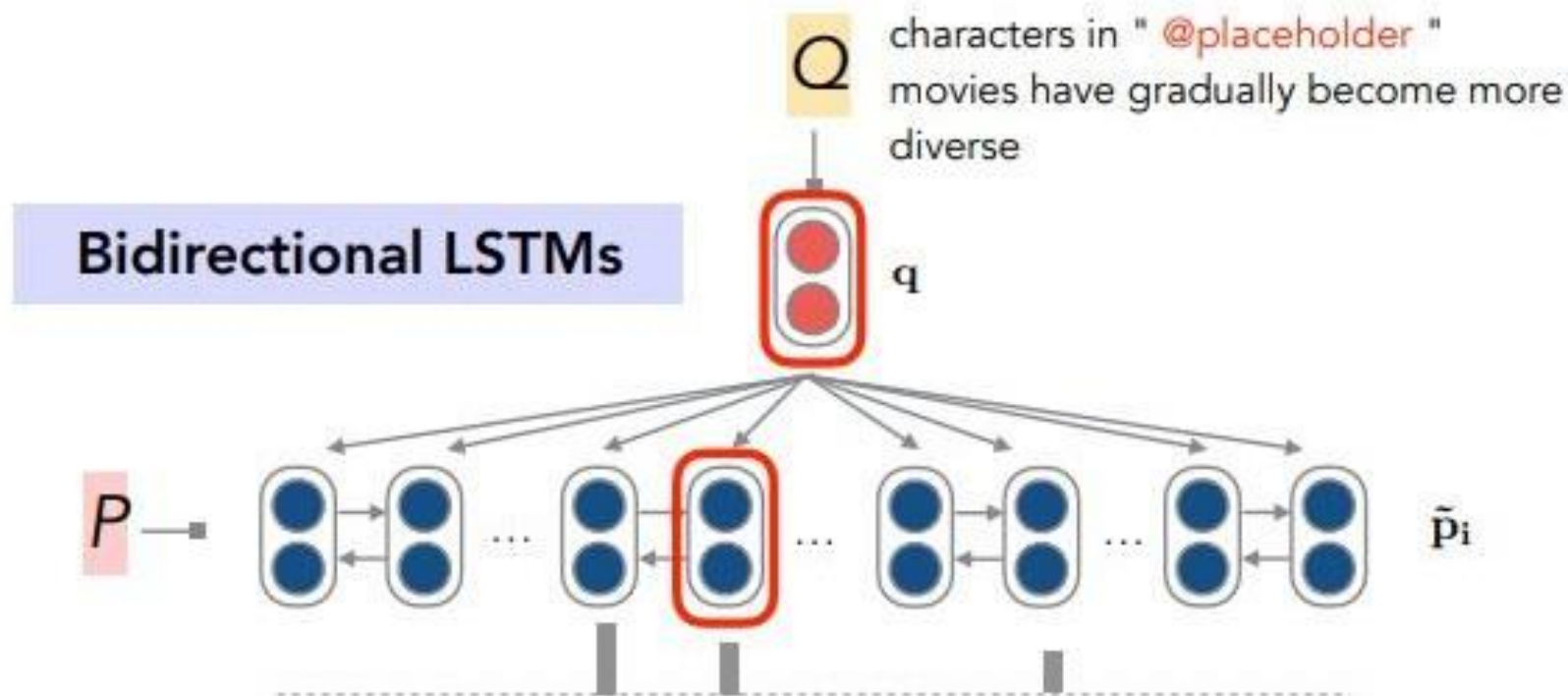
Stanford Attentive Reader



(@entity4) if you feel a ripple in the force today , it may be the news that the official @entity6 is getting its first gay character . according to the sci-fi website @entity9 , the upcoming novel " @entity11 " will feature a capable but flawed @entity13 official named @entity14 who " also happens to be a lesbian . " the character is the first gay figure in the official @entity6 -- the movies , television shows , comics and books approved by @entity6 franchise owner @entity22 -- according to @entity24 , editor of " @entity6 "

Deep Learning Approaches

Stanford Attentive Reader

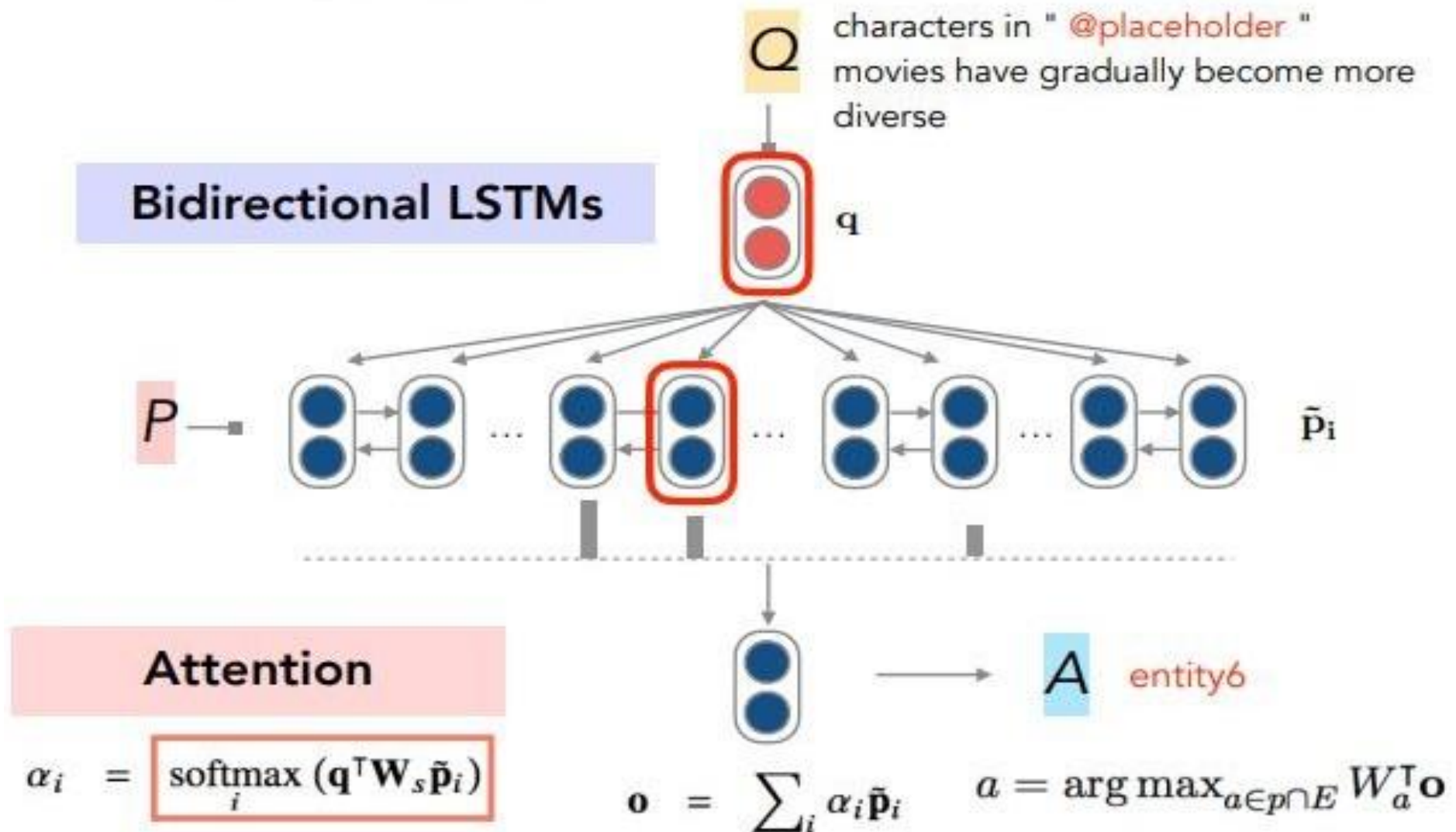


Attention

$$\alpha_i = \text{softmax}_i (q^T W_s \tilde{p}_i)$$

Deep Learning Approaches

Stanford Attentive Reader



Deep Learning Approaches

Stanford Attentive Reader++

Q

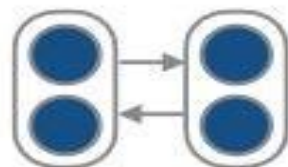
Who did **Genghis Khan** unite before he began **conquering** the rest of **Eurasia**?

Bidirectional RNNs

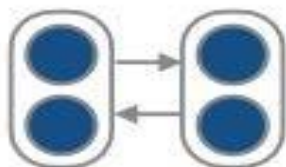


q

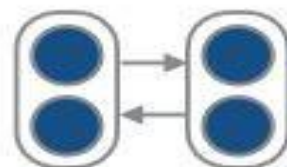
P



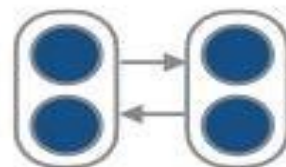
...



...



...

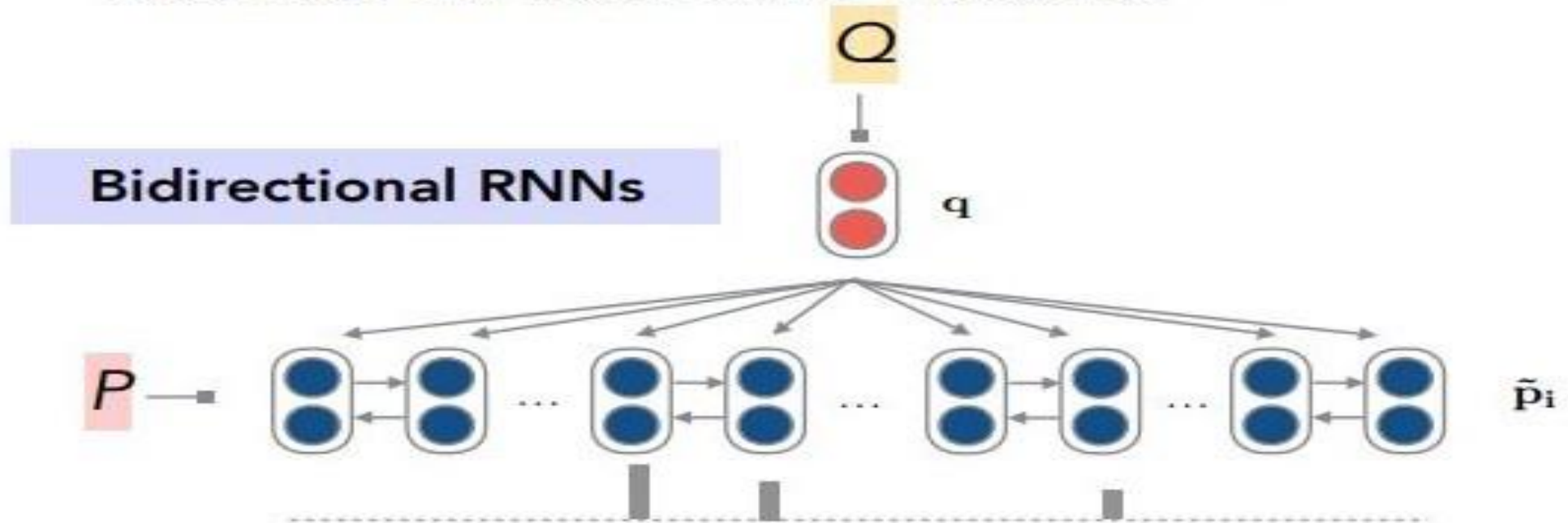


\tilde{p}_i

He came to power by **uniting** many of the nomadic tribes of Northeast Asia. **After** founding the Mongol Empire and being proclaimed "**Genghis Khan**", he started the Mongol invasions that resulted in the **conquest** of most of **Eurasia**. These included raids or invasions of the Qara Khitai, Caucasus, Khwarezmid Empire, Western Xia and Jin dynasties. These campaigns were often accompanied by wholesale massacres of the civilian populations – especially in the Khwarezmian and Xia controlled lands. By the end of his life, the Mongol Empire occupied a substantial portion of Central Asia and China.

Deep Learning Approaches

Stanford Attentive Reader++



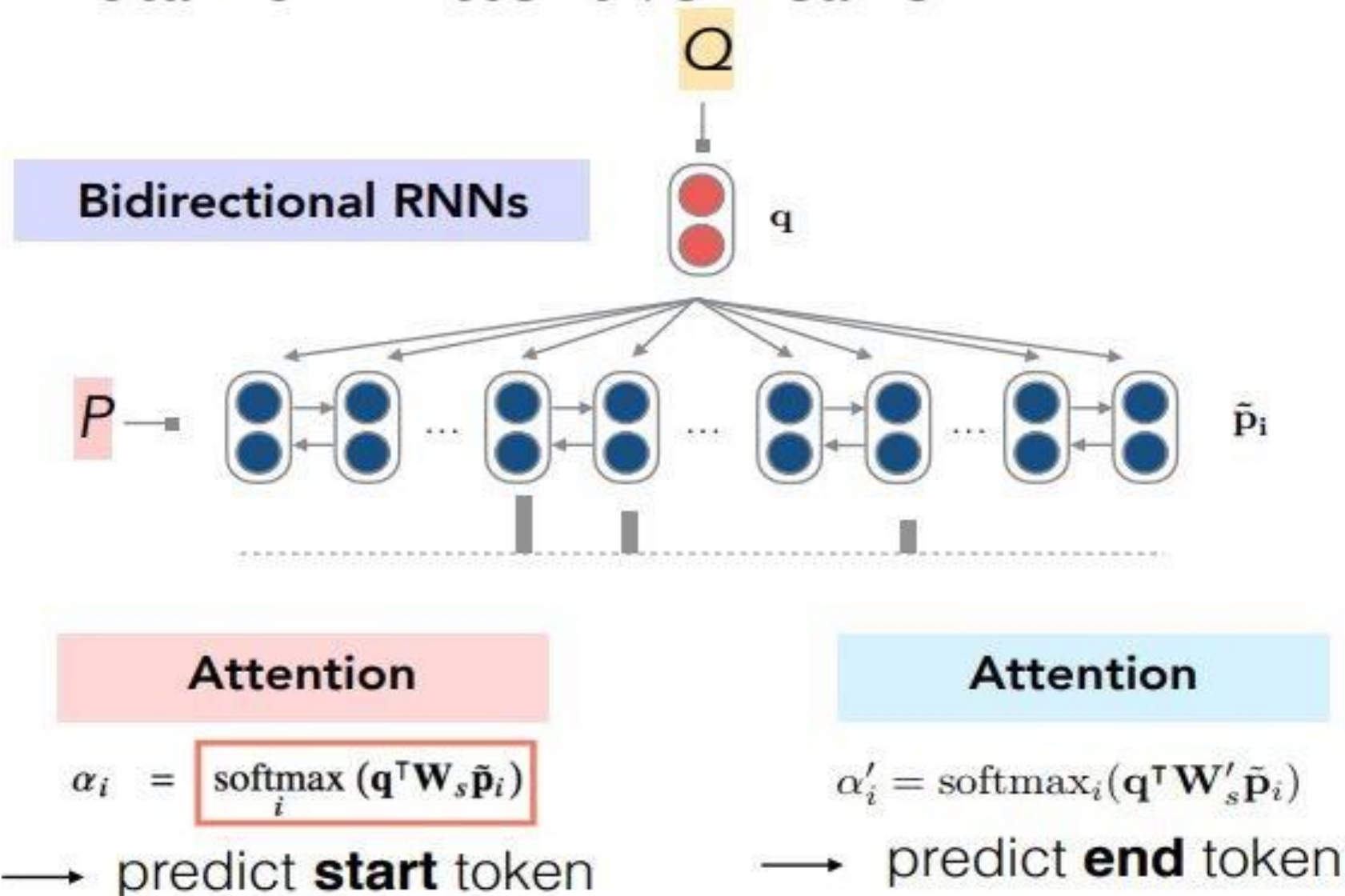
Attention

$$\alpha_i = \text{softmax}_i (\mathbf{q}^\top \mathbf{W}_s \tilde{\mathbf{p}}_i)$$

→ predict **start** token

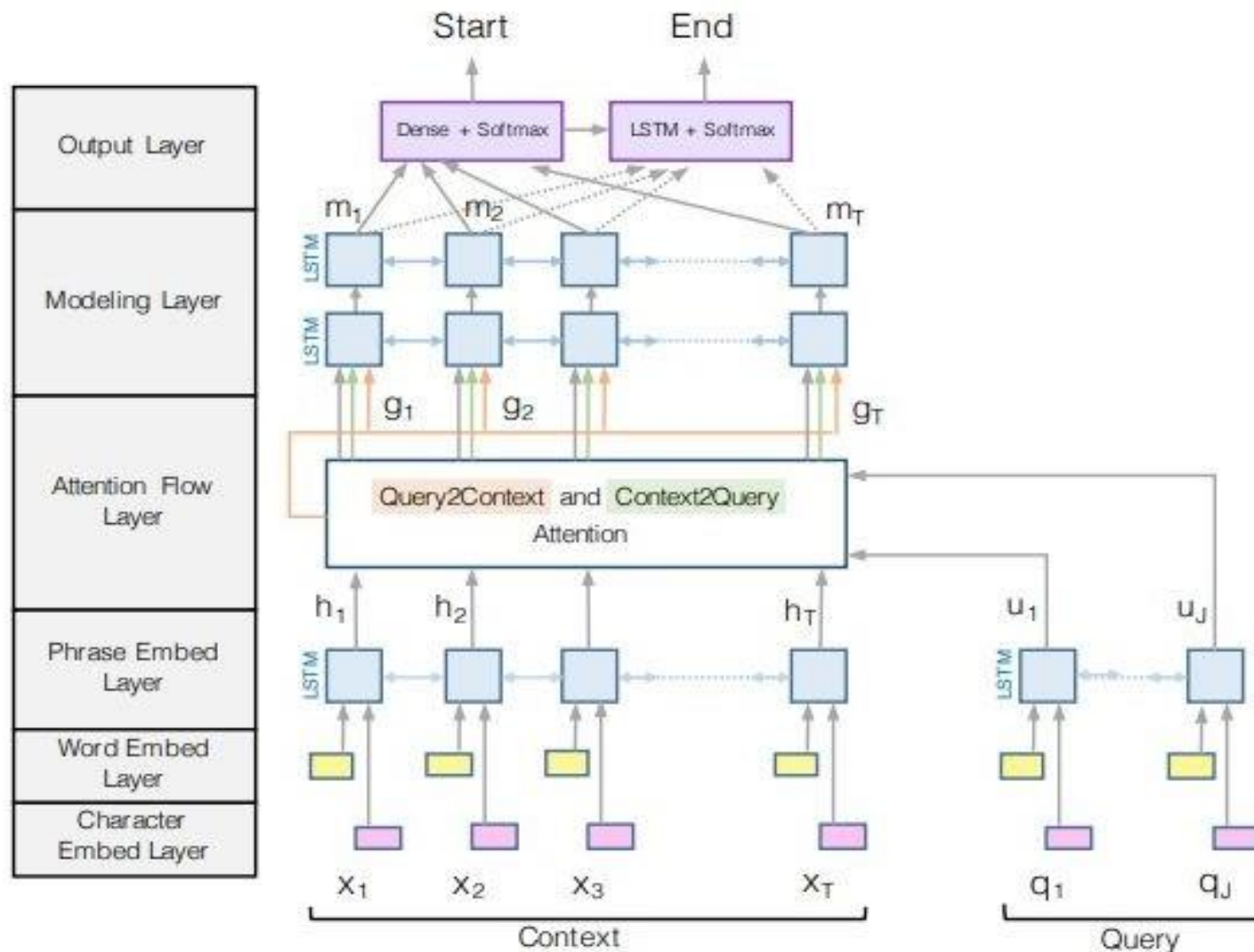
Deep Learning Approaches

Stanford Attentive Reader++



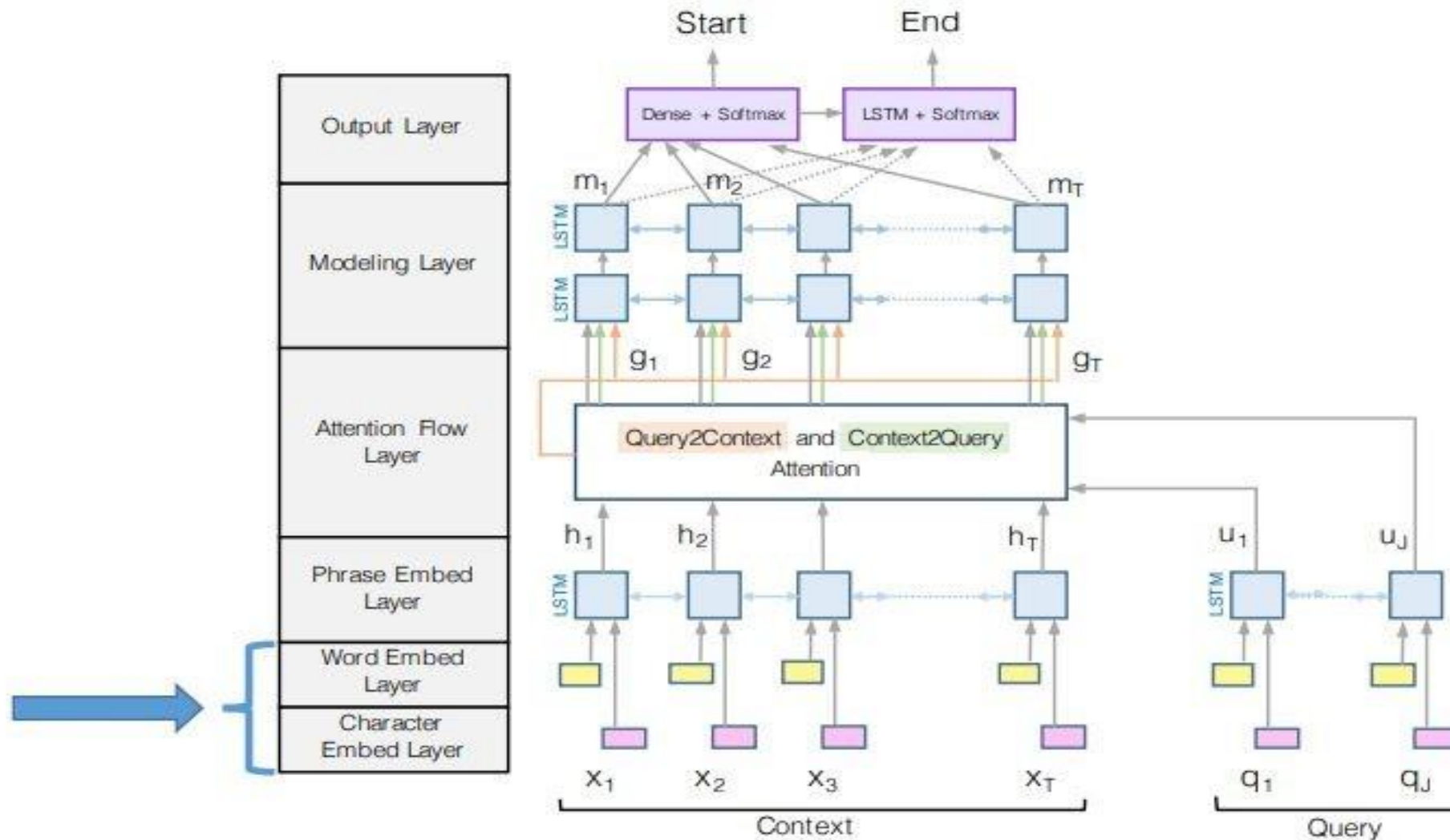
Deep Learning Approaches

(Bidirectional) Attention Flow (Minjoon et. al, 2018)



(Bidirectional) Attention Flow

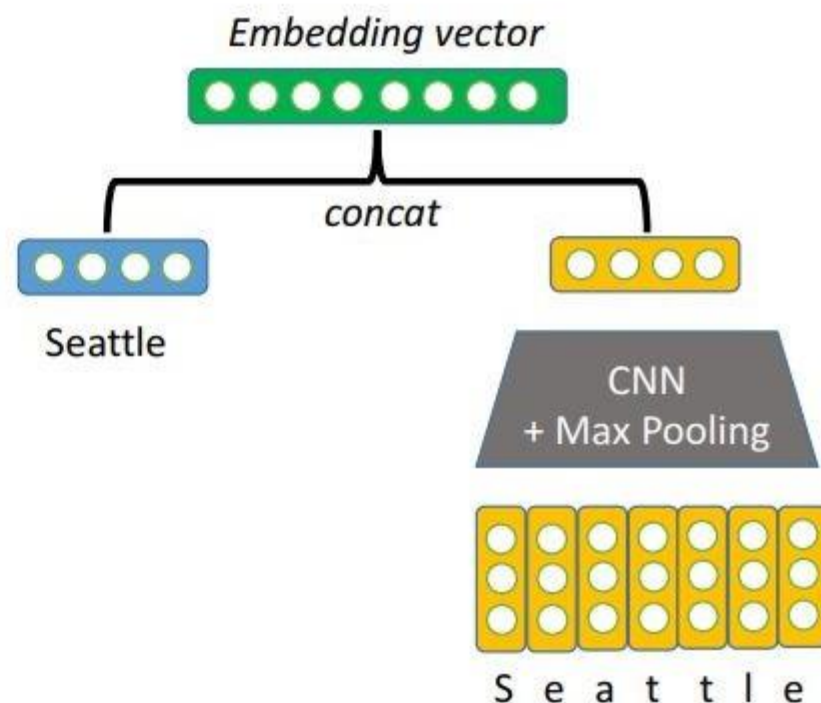
Char/Word Embedding Layers



(Bidirectional) Attention Flow

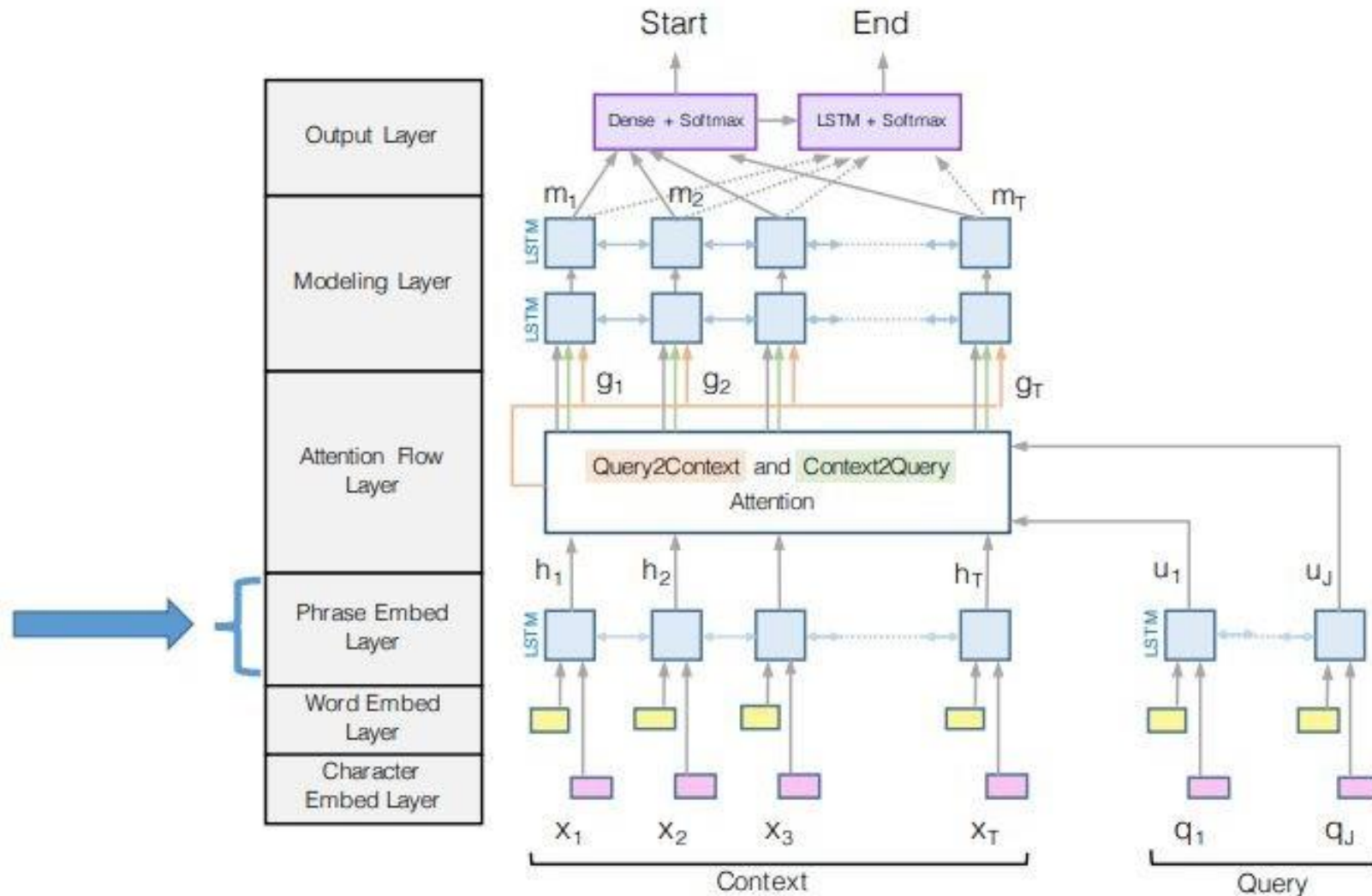
Character and Word Embedding

- Word embedding is fragile against unseen words
 - Char embedding can't easily learn semantics of words
 - Use both!
-
- Char embedding as proposed by Kim (2015)



(Bidirectional) Attention Flow

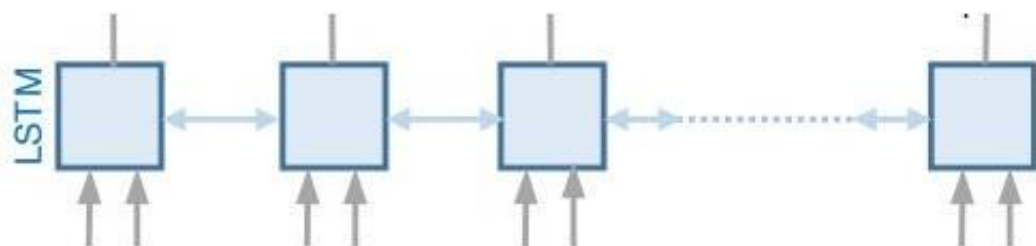
Phrase Embedding Layer



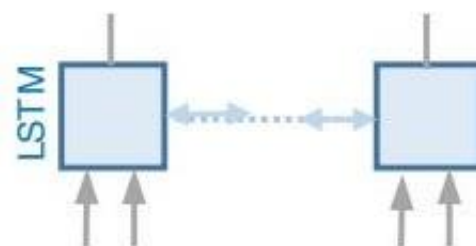
(Bidirectional) Attention Flow

Phrase Embedding Layer

- **Inputs:** the char/word embedding of query and context words
- **Outputs:** word representations aware of their neighbors (phrase-aware words)
- Apply bidirectional RNN (LSTM) for both query and context



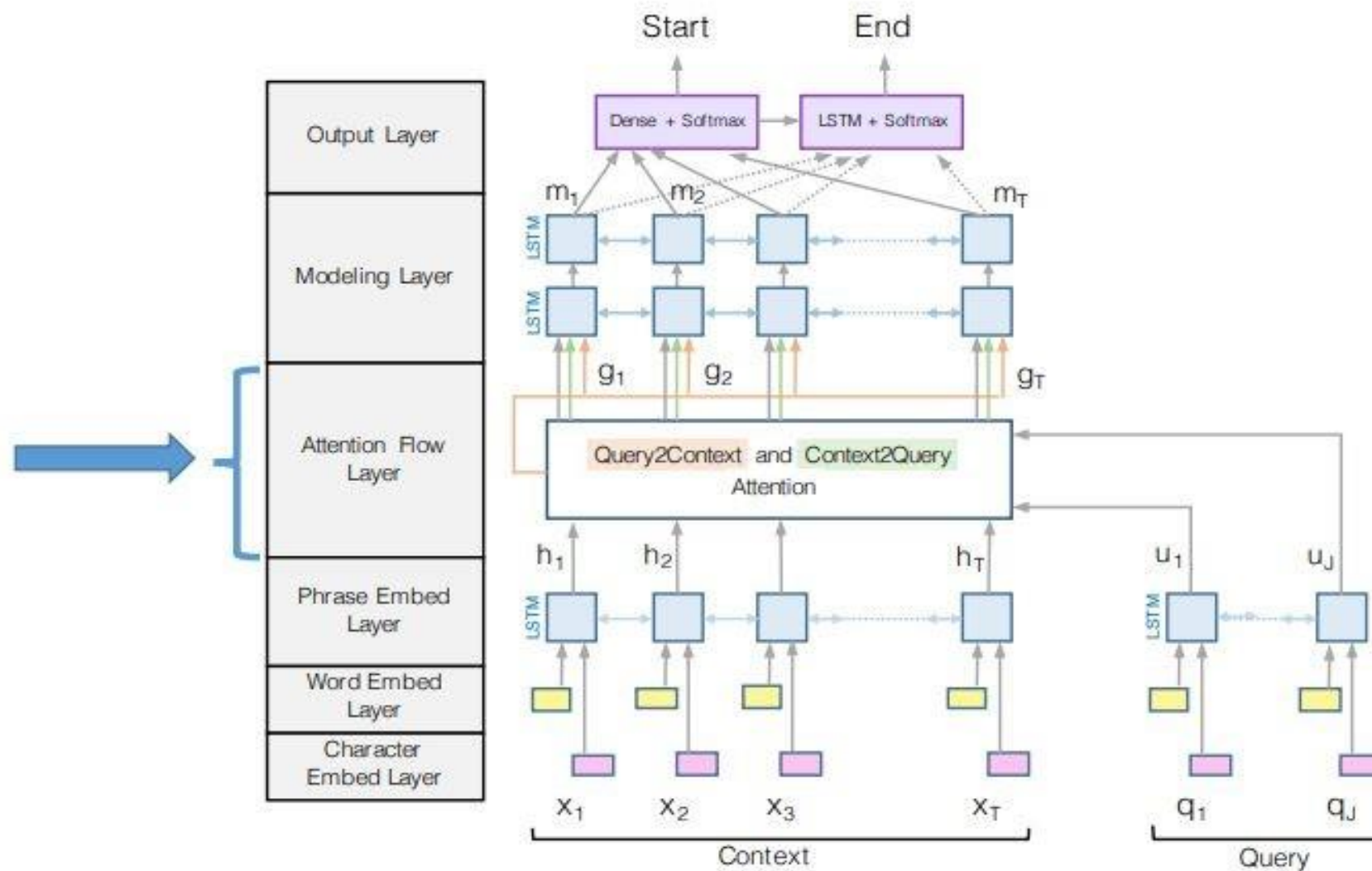
Context



Query

(Bidirectional) Attention Flow

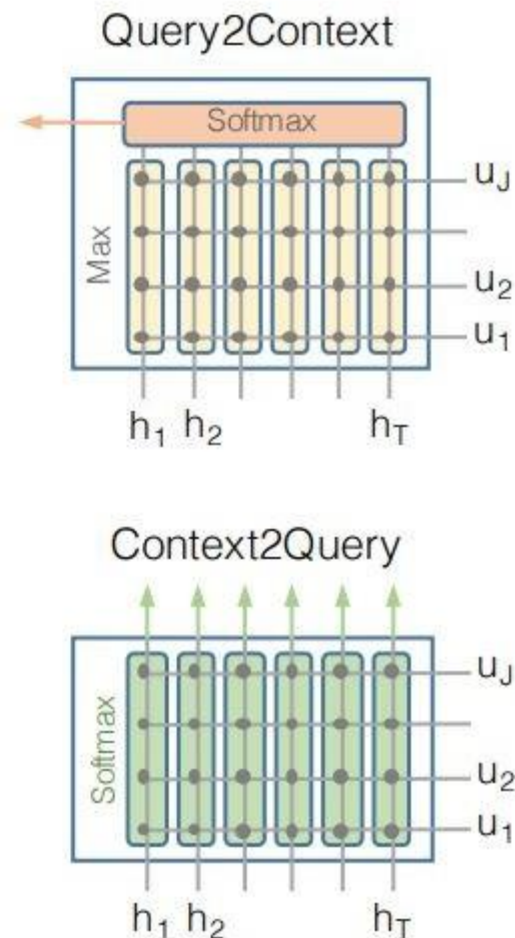
Attention Layer



(Bidirectional) Attention Flow

Attention Layer

- **Inputs:** phrase-aware context and query words
- **Outputs:** query-aware representations of context words
- **Context-to-query attention:** For each (phrase-aware) context word, choose the most relevant word from the (phrase-aware) query words
- **Query-to-context attention:** Choose the context word that is most relevant to any of query words.

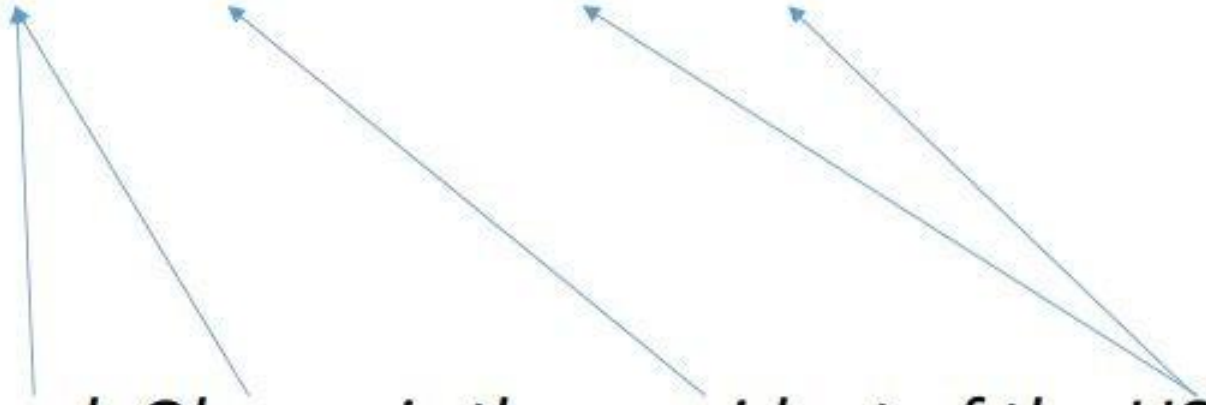


(Bidirectional) Attention Flow

Context-to-Query Attention (C2Q)

Q: *Who leads the United States?*

C: *Barak Obama is the president of the USA.*



For each context word, find the most relevant query word.

(Bidirectional) Attention Flow

Query-to-Context Attention (Q2C)

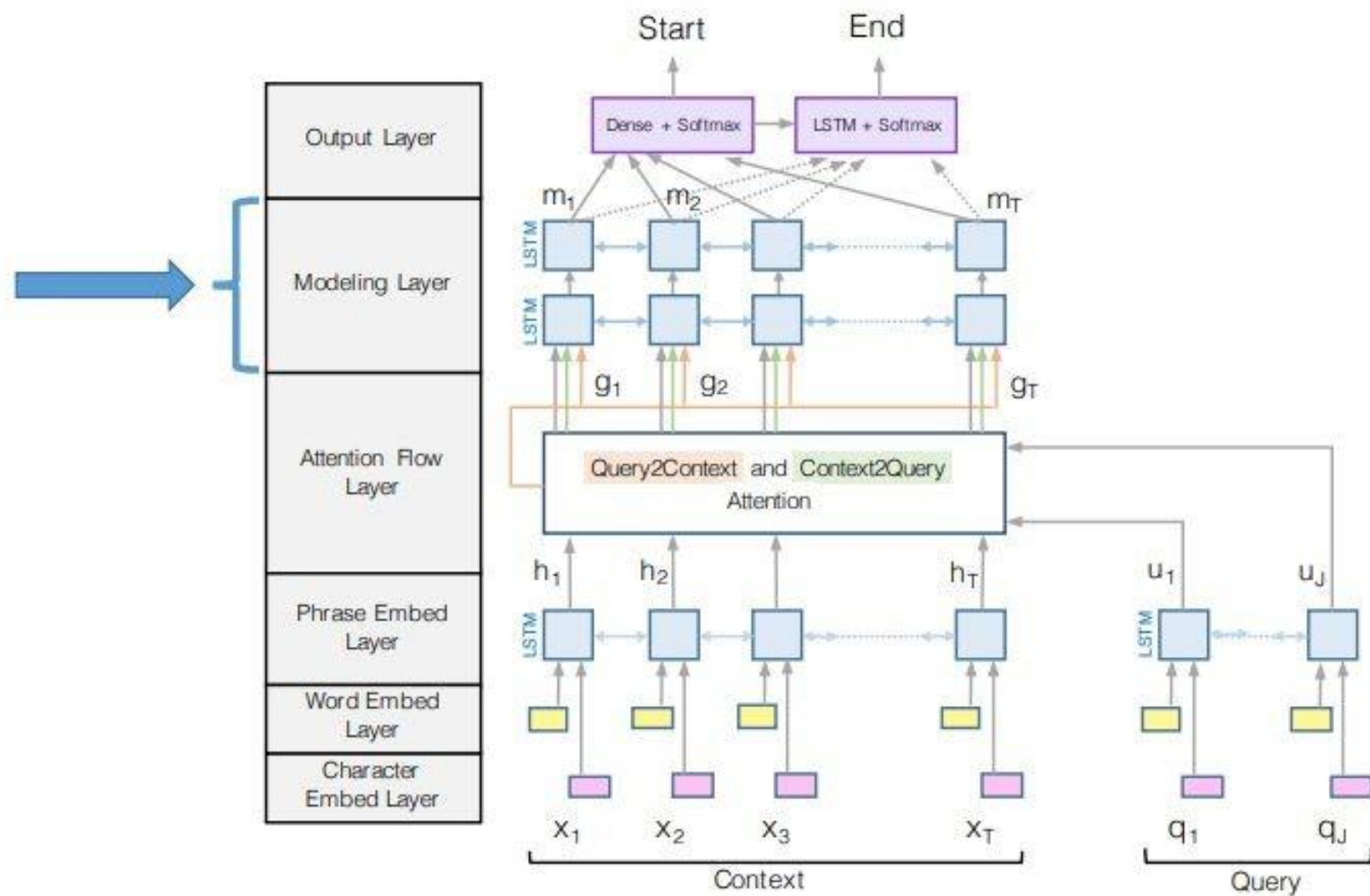
*While **Seattle**'s weather is very nice in summer, its weather is very rainy in winter, making it one of the most gloomy cities in the U.S. LA is ...*

Q: Which city is gloomy in winter?



(Bidirectional) Attention Flow

Modeling Layer



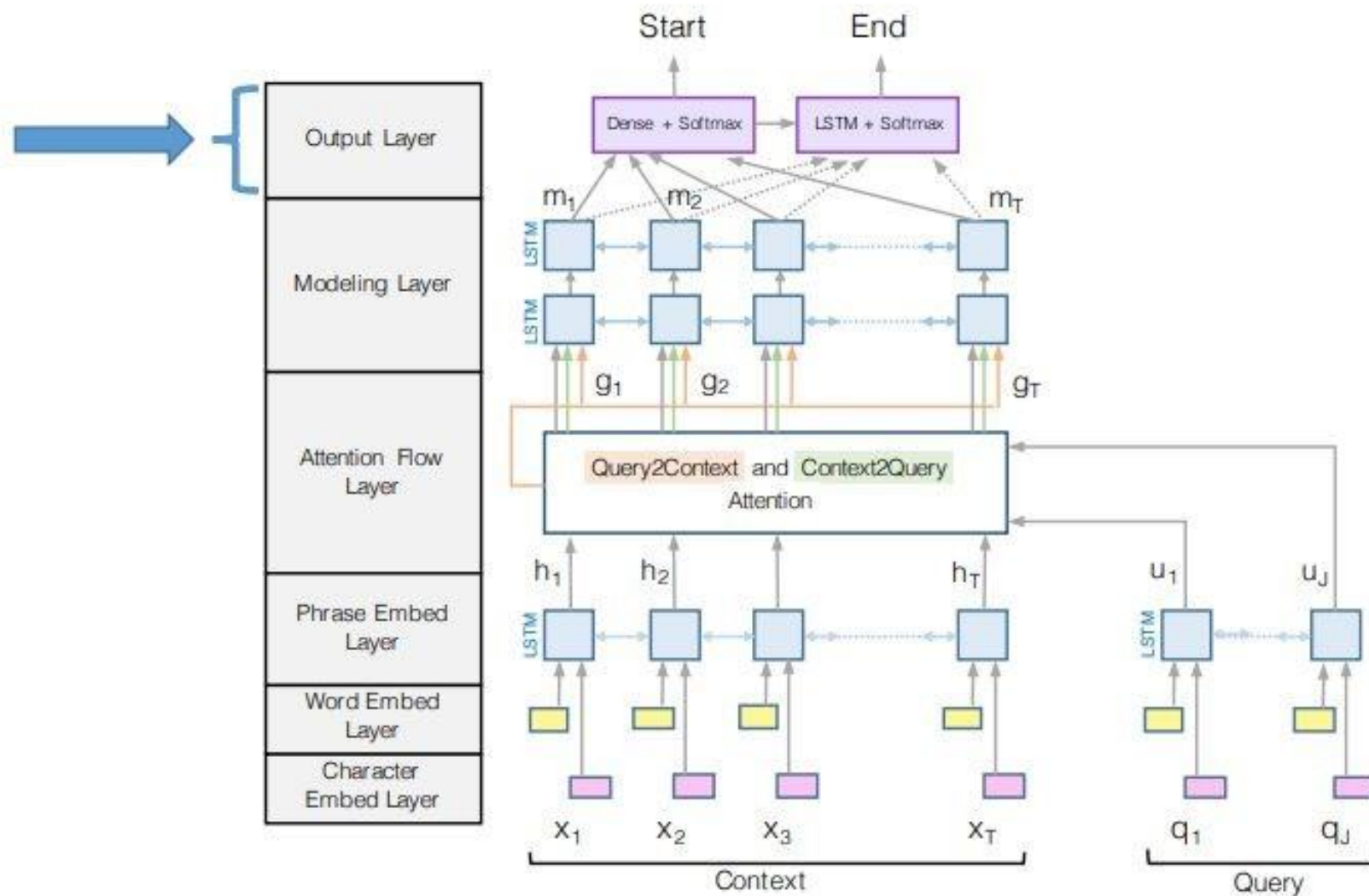
(Bidirectional) Attention Flow

Modeling Layer

- **Attention layer:** modeling interactions between query and context
- **Modeling layer:** modeling interactions within (query-aware) context words via RNN (LSTM)

(Bidirectional) Attention Flow

Output Layer



References

- Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *arXiv preprint arXiv:1611.01603* (2016).
- Chen, Danqi, Jason Bolton, and Christopher D. Manning. "A thorough examination of the cnn/daily mail reading comprehension task." *arXiv preprint arXiv:1606.02858* (2016).
- Manning, Christopher. "Natural Language Processing with Deep Learning CS224N/Ling284. Lecture 11." (2017).
- Brill, Eric, Susan Dumais, and Michele Banko. "An analysis of the AskMSR question-answering system." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

Credit: Some of the slides are taken from the following lectures:

- <https://www.slideshare.net/marinasantini1/lecture-question-answering>
- <https://web.stanford.edu/class/cs224n/slides/cs224n-2019-lecture10-QA.pdf>

Thank you for your attention!