Natural Language Processing
Lecture 10 Conversational AI; Question Answering; Dialog Systems

Qun Liu, Valentin Malykh
Huawei Noah's Ark Lab

Spring 2022
A course delivered at KFU, Kazan

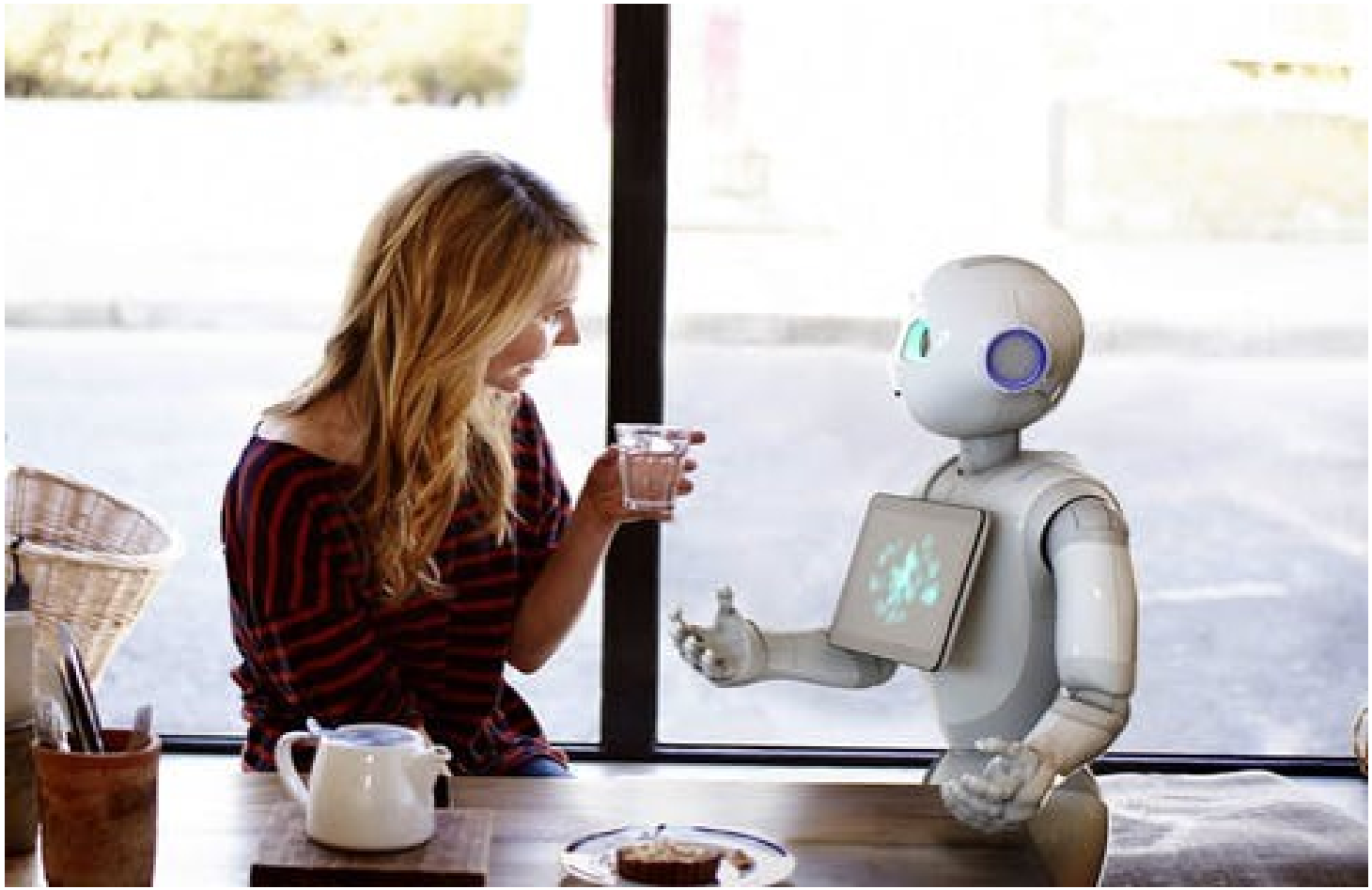
Content

Content

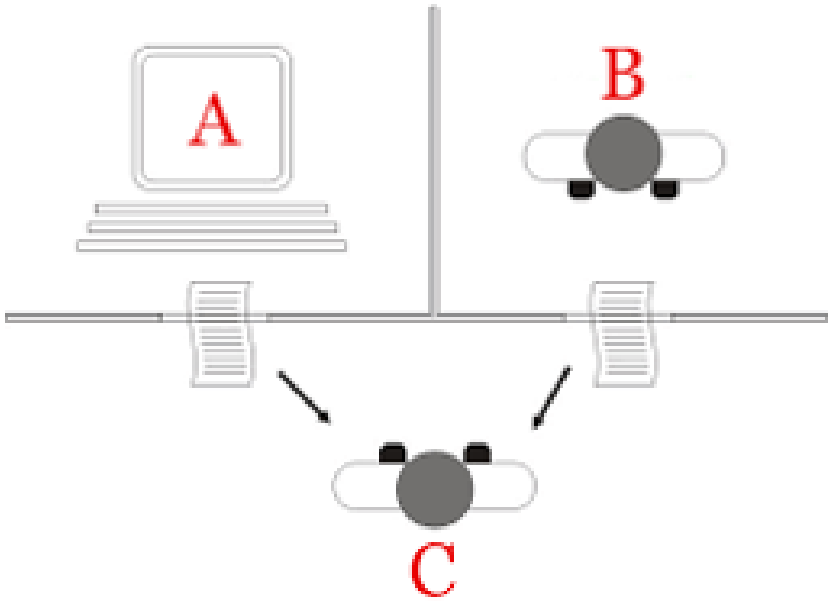
Holy grails of NLP (Recap)

- Accurate machine translation between human languages
- Free conversation between humans and computers

Free human machine conversation (Recap)



Turing test (Recap)



By Juan Alberto Sánchez Margallo, CC BY 2.5, from Wikipedia

Classifications of conversional systems

- Question Answering (QA) Systems
  - Single turn conversation: no dialog context is involved
  - The objective is to answer user's questions
- Dialog Systems
  - Multi-turn conversation: dialog context is involved
  - Diverse objectives: task completion, chitchat, QA
- Multimodal QA/Dialog Systems
  - Additional modality is involved
  - Single turn (Visual QA) or multi-turn (Visual Dialog)
  - Ojbective: conversation around the information given in the additional modalities

Content

Content

Early QA systems

- BASEBALL(1961) and LUNAR(1971-1973)
- SHRDLU(1968-1970)
- Eliza (1964)
- Expert Systems (1970s-1980s)

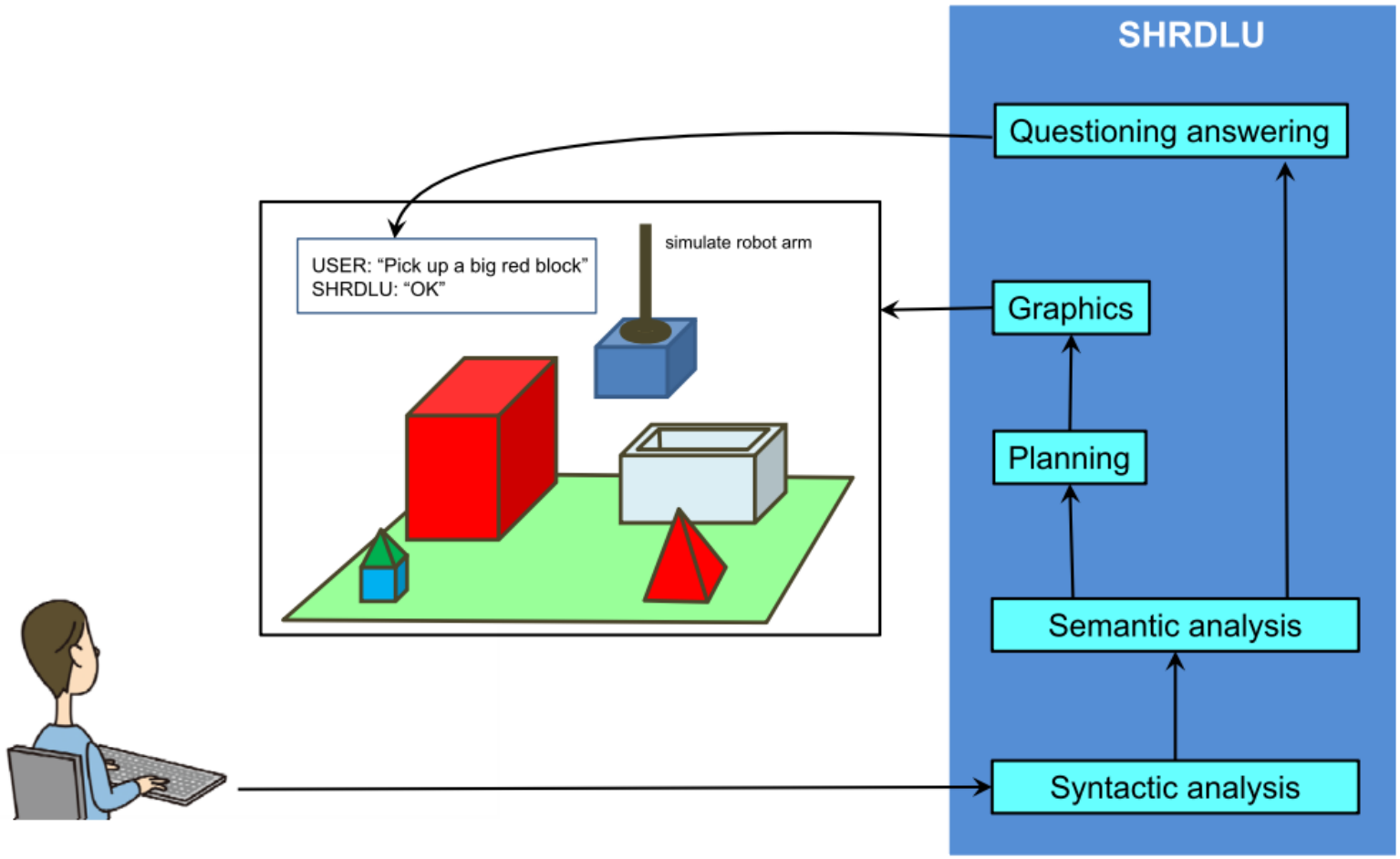
BASEBALL(1961) and LUNAR(1971-1973)

- Two early question answering systems were BASEBALL and LUNAR.
- BASEBALL answered questions about the US baseball league over a period of one year. LUNAR, in turn, answered questions about the geological analysis of rocks returned by the Apollo moon missions.
- Both question answering systems were very effective in their chosen domains. In fact, LUNAR was demonstrated at a lunar science convention in 1971 and it was able to answer 90% of the questions in its domain posed by people untrained on the system.

SHRDLU (1968-1970)

- SHRDLU was an early natural language understanding computer program, developed by Terry Winograd at MIT in 1968–1970.
- In it, the user carries on a conversation with the computer, moving objects, naming collections and querying the state of a simplified "blocks world", essentially a virtual box filled with different blocks.
- SHRDLU was a famous system in the AI history, which firstly demonstrated the ability of an AI system to interact with humans with natural languages to undertake certain tasks.

SHRDLU (1968-1970)



Winograd, Terry, *Understanding Natural Language*, New York: Academic Press, 1972. http://hci.stanford.edu/~winograd/shrdlu/

Eliza (1964)

- ELIZA is an early natural language processing computer program created from 1964 to 1966 at the MIT Artificial Intelligence Laboratory by Joseph Weizenbaum.
- Eliza simulated conversation by using a "pattern matching" and substitution methodology that gave users an illusion of understanding on the part of the program, but had no built in framework for contextualizing events.
- Many early users were convinced of ELIZA's intelligence and understanding, despite Weizenbaum's insistence to the contrary.
- ELIZA was one of the first chatterbots and one of the first programs capable of attempting the Turing test.

Eliza (1964)

```
Welcome to
                 EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                 EE      LL       II       ZZ   AA   AA
                 EEEEE   LL       II     ZZZ    AAAAAAA
                 EE      LL       II    ZZ      AA   AA
                 EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Expert Systems (1970s-1980s)

- An expert system is a computer system that emulates the decision-making ability of a human expert.
- Expert systems solve complex problems by reasoning through bodies of knowledge, represented mainly as if–then rules.
- Expert systems rely heavily on expert-constructed and organized knowledge bases.
- Expert systems were firstly created in the 1970s and then proliferated in the 1980s.

Content

Big data era: open domain QA

- Open Domain QA and Web-based QA
- Community QA (2000s)
- Wolfram Alpha (2009)
- IBM Watson (2011)

Open domain QA

- Open-domain question answering is a category of QA which deals with questions about nearly anything, and can only rely on unstructured data (raw text). On the other hand, these systems usually have much more data available from which to extract the answer.
- The returned answer is in the form of short texts rather than a list of relevant documents (unlike information retrieval systems).
- The system uses a combination of techniques from computational linguistics, information retrieval and knowledge representation for finding answers.

Open domain QA: a brief history

- Simmons et al. (1964) did first exploration of answering questions from an expository text based on matching dependency parses of a question and answer
- Murax (Kupiec 1993) aimed to answer questions over an online encyclopedia using IR and shallow linguistic processing
- The NIST TREC QA track begun in 1999 first rigorously investigated answering fact questions over a large collection of documents
- IBM's Jeopardy! System (DeepQA, 2011) brought attention to a version of the problem; it used an ensemble of many methods
- DrQA (Chen et al. 2016) uses IR followed by neural reading comprehension to bring deep learning to Open-domain QA

Thomas Lukasiewicz, Advanced Machine Learning: Deep Learning for NLP: Lecture 11: Question Answering, 2019

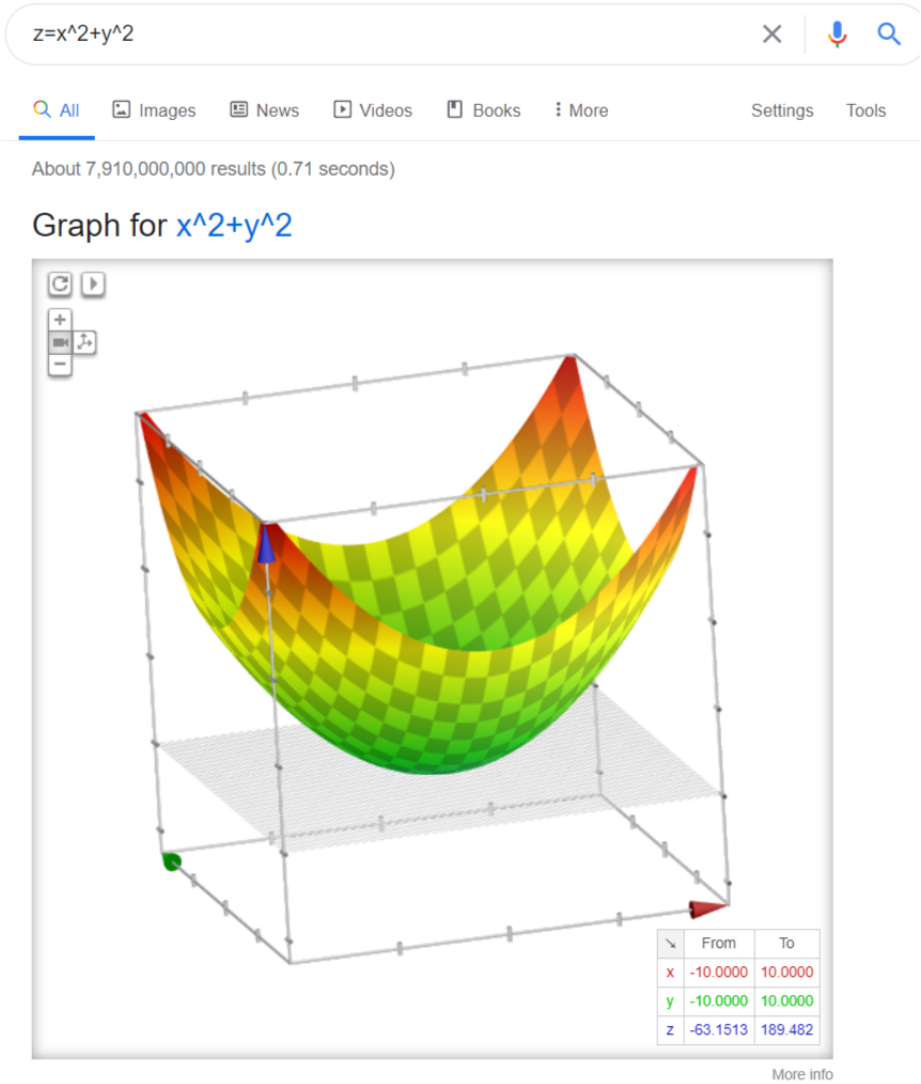
Web-based QA

- As a special case of open-domain QA, big search engines like Google provide direct answers to user queries rather than a list of web pages, when it feels confident.

Web-based QA


how many countries are adjacent to china in land?
All
Images
Maps
News
Videos
More
Settings
Tools
About 638,000,000 results (0.72 seconds)
14
As the most populous country in the world and third largest in area, China also has the largest number of neighbours (14) sharing its 22,000km land borders namely: North Korea, Russia, Mongolia, Kazakhstan, Kyrgyzstan, Tajikistan, Afghanistan, Pakistan, India, Nepal, Bhutan, Myanmar, Laos and Vietnam. Mar 1, 2012
www.eu-asiacentre.eu › pub_details
China and its Neighbours: troubled relations - eu asia centre
About Featured Snippets
Feedback

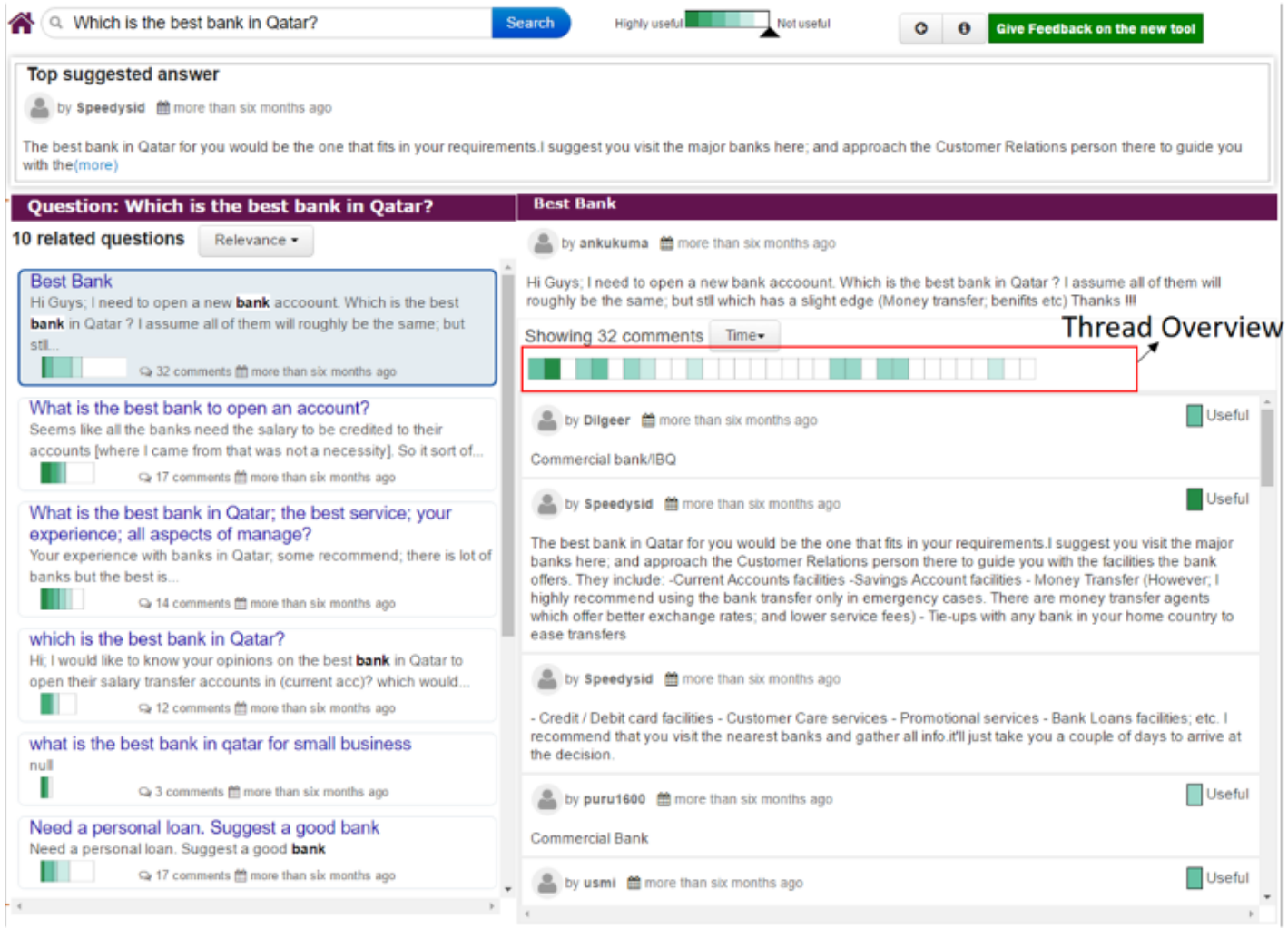
Web-based QA


z=x^2+y^2
All
Images
News
Videos
Books
More
Settings
Tools
About 7,910,000,000 results (0.71 seconds)
Graph for x^2+y^2
From
To
x
-10.0000
10.0000
y
-10.0000
10.0000
z
-63.1513
189.482
More info

Community QA

- Community QA is a category of QA which is based on QA forums.
- A community QA system searchs the forum to find an existing quesiton which is equivalent to the user's input and return a best answer to that question from the forum.
- The questions which can be answered by a community QA system are limited to the forum data, however, the answers are generally of high quality because it is written by humans.
- Community Question Answering has seen a spectacular increase in popularity in 2000s along with the popularity of QA forums like Yahoo! Answers, Stack Overflow, Quora, etc.

Community QA

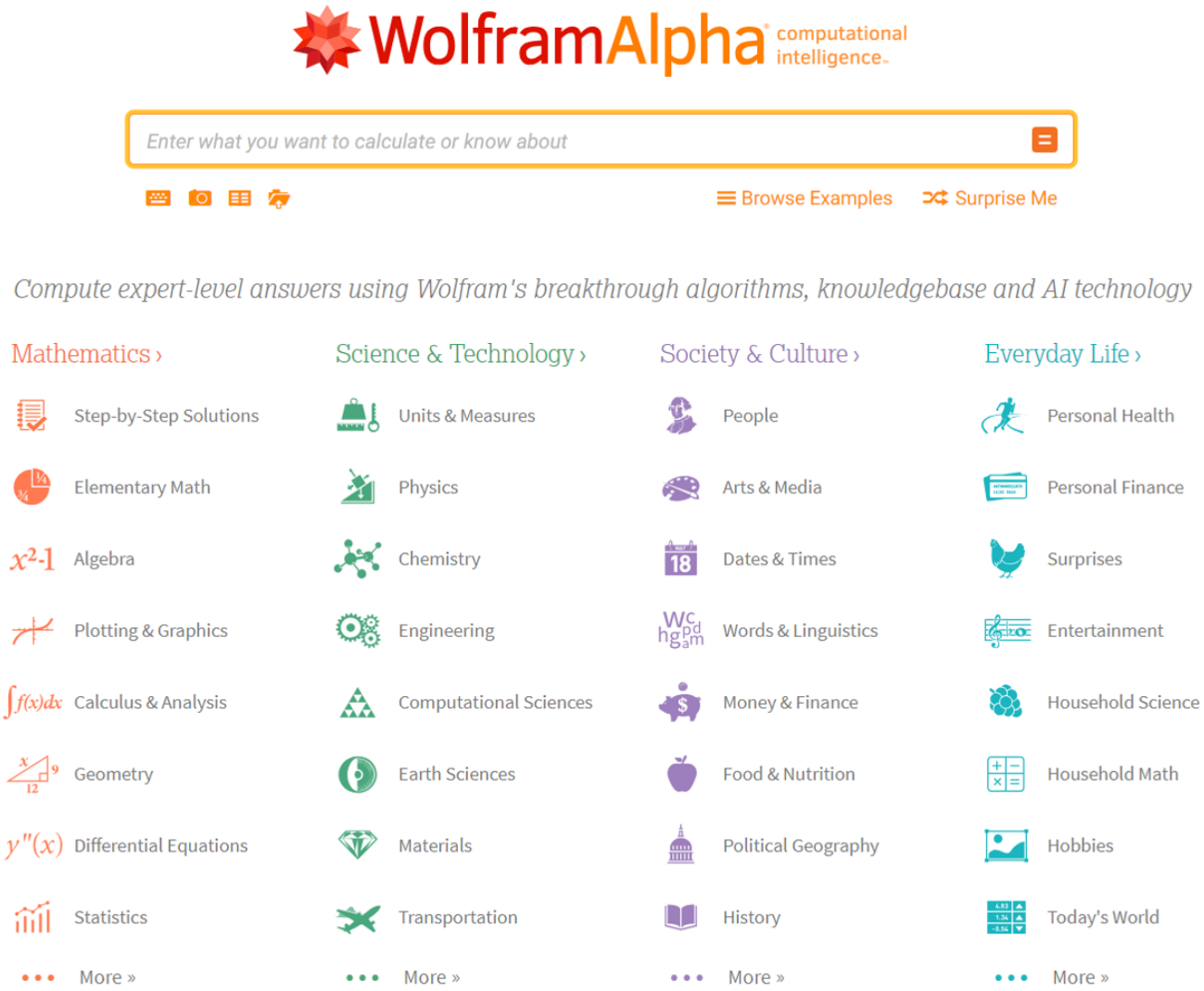


Community Question Answering System, NTU NLP Group

Wolfram Alpha (2009)

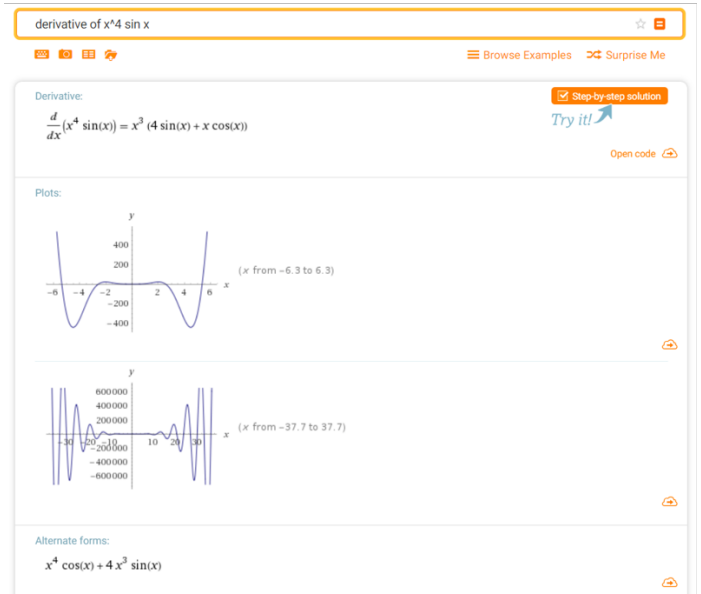
- WolframAlpha (also styled Wolfram|Alpha) is a computational knowledge engine or answer engine developed by Wolfram Alpha LLC, a subsidiary of Wolfram Research.
- It is an online service that answers factual queries directly by computing the answer from externally sourced "curated data", rather than providing a list of documents or web pages that might contain the answer as a search engine might.
- WolframAlpha can only provide robust query results based on computational facts, not queries on the social sciences, cultural studies or even many questions about history where responses require more subtlety and complexity.

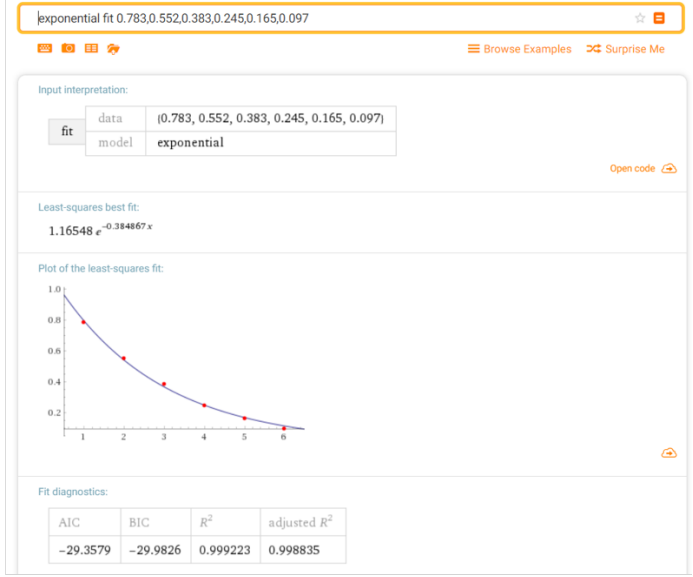
Wolfram Alpha (2009)



Wolfram Alpha (2009)

- WolframAlpha can not only answer factual questions with texts, but also run programs to compute the answer and give visualization of the answers (e.g. a curve of a function).





IBM Watson (2011)

- Watson is a question-answering computer system capable of answering questions posed in natural language, developed in IBM's DeepQA project by a research team led by principal investigator David Ferrucci.
- In 2011, the Watson computer system competed on Jeopardy! against champions Brad Rutter and Ken Jennings, winning the first place prize of $1 million.
- In February 2013, IBM announced that Watson software system's first commercial application would be for utilization management decisions in lung cancer treatment at Memorial Sloan Kettering Cancer Center, New York City, in conjunction with WellPoint (now Anthem).

IBM Watson in Jeopardy! (2013)

This is another breakthrough of an AI system to beat human competitors in real games after IBM's DeepBlue defeated the world chess champion Garry Kasparov on 10 February 1996.

Content

Neural era: machine comprehension and dialog systems

- Machine reading comprehension
- Siri(2011) and voice assistants
- Amazon Alexa(2015) and smart speakers
- XiaoIce(2014) and social chatbots
- Natural Language Interface to Databases
- Knowledge Base/Graph QA (KBQA/KGQA)
- Visual QA / Visual Dialog

Machine reading comprehension

- Machine Reading Comprehension (MRC), or Machine Reading (MC), or Machine Comprehension (MC), is the task to read and understand a piece of unstructured text and then answer questions about it.
- MRC is a growing field of research due to its potential in various enterprise applications.
- Although the idea of MRC emerged rather early, only in the past decade, a huge development has been witnessed in this field, including the soar of numbers of corpus (MSMARCO, SQuAD, NewsQA, etc.) and great progress in techniques.

Machine reading comprehension

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

Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." arXiv:1606.05250, 2016

Machine reading comprehension: a brief history

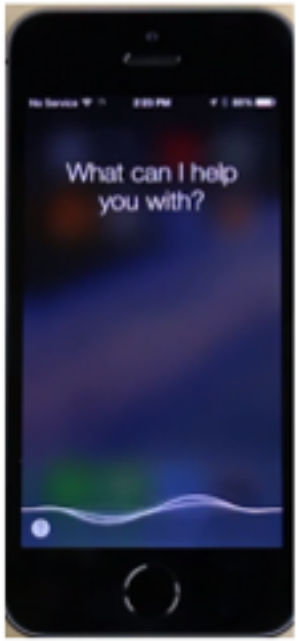
- Much early NLP work attempted reading comprehension
  - Schank, Abelson, Lehnert et al. c. 1977 – "Yale A.I. Project"
- Revived by Lynette Hirschman in 1999:
  - Could NLP systems answer human reading comprehension questions for 3rd to 6th graders? Simple methods attempted.
- Revived again by Chris 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, ...

Thomas Lukasiewicz, Advanced Machine Learning: Deep Learning for NLP: Lecture 11: Question Answering, 2019

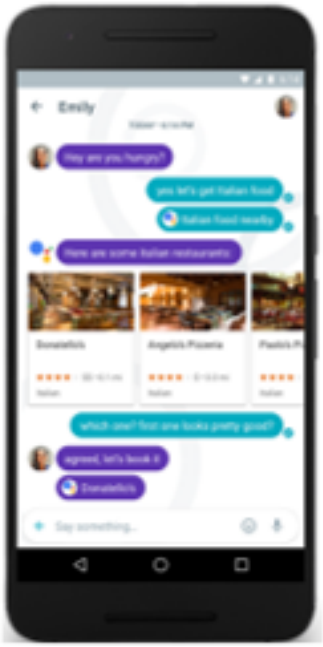
Siri (2011) and voice assistants

- Siri is a virtual assistant that is part of Apple Inc.'s operating systems.
- The assistant uses voice queries and a natural-language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Internet services.
- Siri became the first digital virtual assistant to be standard on a smartphone when the iPhone 4s came out on October 4, 2011, and followed by a number of similar products including Google Assistant and Microsoft Cortona.

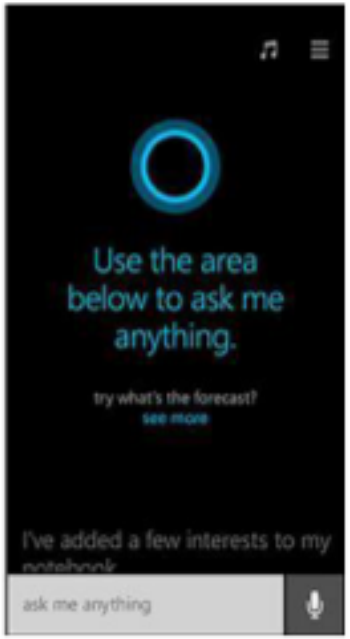
Siri (2011) and voice assistants



Apple Siri 2011



Google Assistant 2016



Microsoft Cortana 2014

Amazon Alexa (2015) and smart speakers

- Amazon Alexa, also known simply as Alexa, is a virtual assistant AI technology developed by Amazon, first used in the Amazon Echo smart speakers developed by Amazon Lab126.
- It is capable of voice interaction, music playback, making to-dolists, and many other real-time information. Alexa can also control several smart devices using itself as a home automation system.
- Users are able to extend the Alexa capabilities by installing "skills" (additional functionality developed by third-party vendors, in other settings more commonly called apps such as weather programs and audio features).

Amazon Alexa (2015) and smart speakers

- Amazon was successful in the market and has been followed by other smart speaker products like Google Home, etc.

Amazon Echo
2015

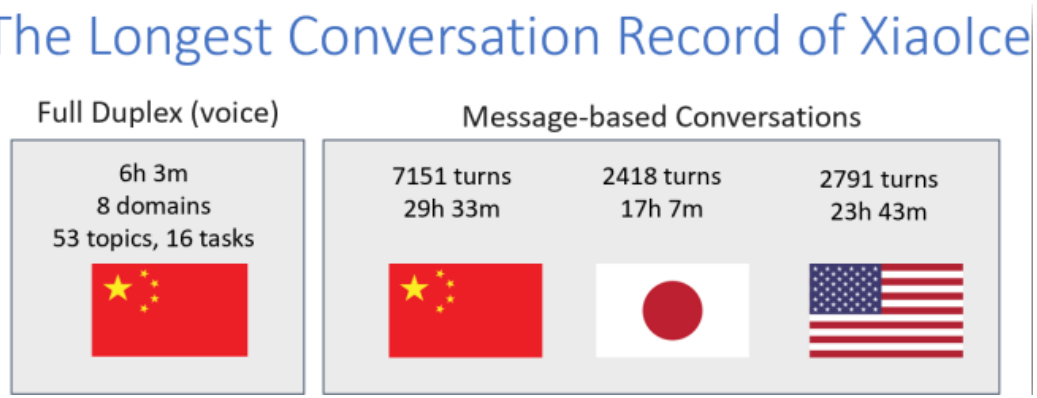
Google Home
2016

Microsoft Xiaoice (2014) and social chatbots

- Xiaoice (Chinese: 微软小冰) is the AI system developed by Microsoft STCA in 2014 based on emotional computing framework.
- Through the comprehensive application of algorithms, cloud computing and big data, Xiaoice adopts the intergenerational upgrade method to gradually form a complete artificial intelligence system to EQ.
- Microsoft Xiaoice has become one of the world's biggest interdisciplinary AI systems and taken various product forms including chat bot, intelligent voice assistant, AI content creation and production platform, etc.

Microsoft Xiaoice (2014) and social chatbots

- In many countries around the world, as a single brand, Microsoft Xiaoice has covered 660 million on-line users, 450 million third-party IoT devices and 900 million content viewers. It has kept 23 conversations per session (CPS) averagely with users while greatly increasing interactive scenarios.



Jianfeng Gao, Michel Galley, Neural Approaches to Conversational AI, ICML 2019

Natural Language Interface to Databases

- Convert a natural language question to a database query
- Execute the database query on the database and obtain the answer
- Example systems:
  - Luner (1960s) NLMENU(1980s)
  - PRECISE (2002) ThoughtSpot (2012) Arimo (2012)
  - NaLIR(2014) Poser BI (2015) SimpleQL (2013)

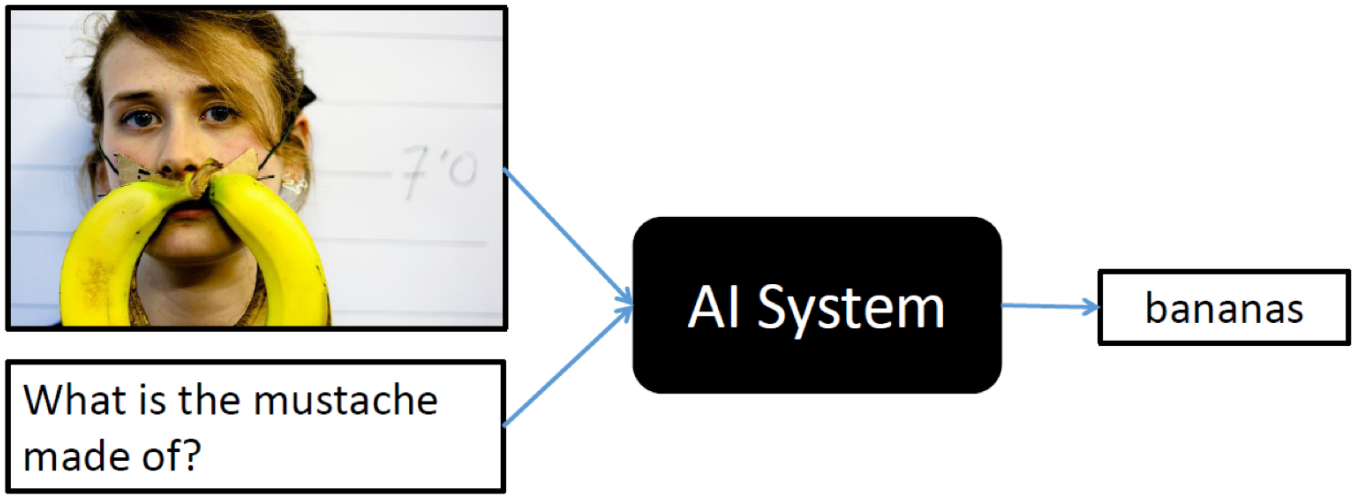
Jonas Chapuis, Natural Language Interfaces to Databases (NLIDB)

Knowledge Base/Graph QA (KBQA/KGQA)

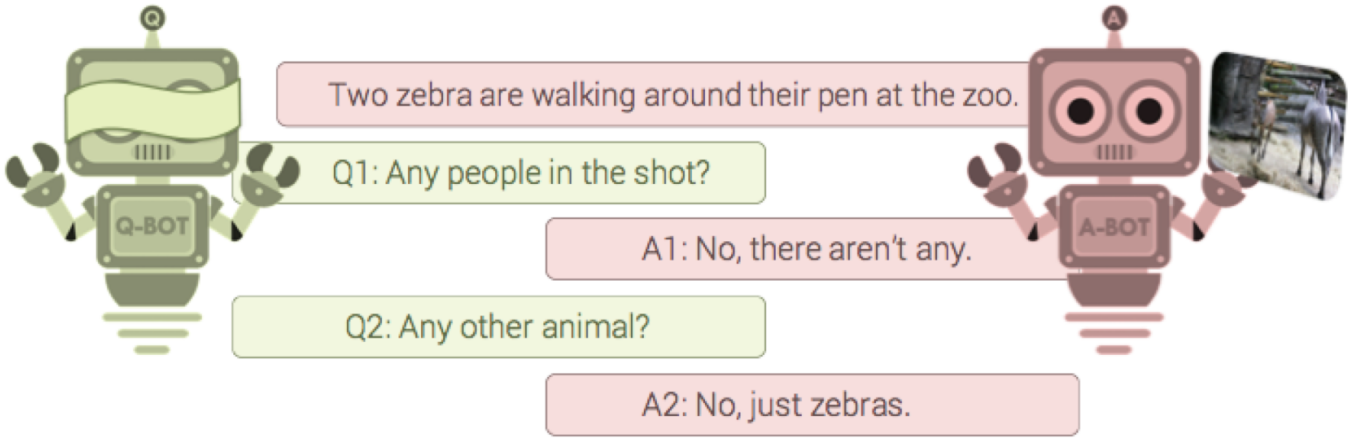
- QA over knowledge graphs
  - Freebase
  - DBPedia
  - etc.
- Large scale data
- Structured data
- Inference is needed for answering some questions

Visual QA / Visual Dialog

Visual QA



Visual Dialog



Visual Question Answering and Dialog Workshop

Content

Why we care about QA?

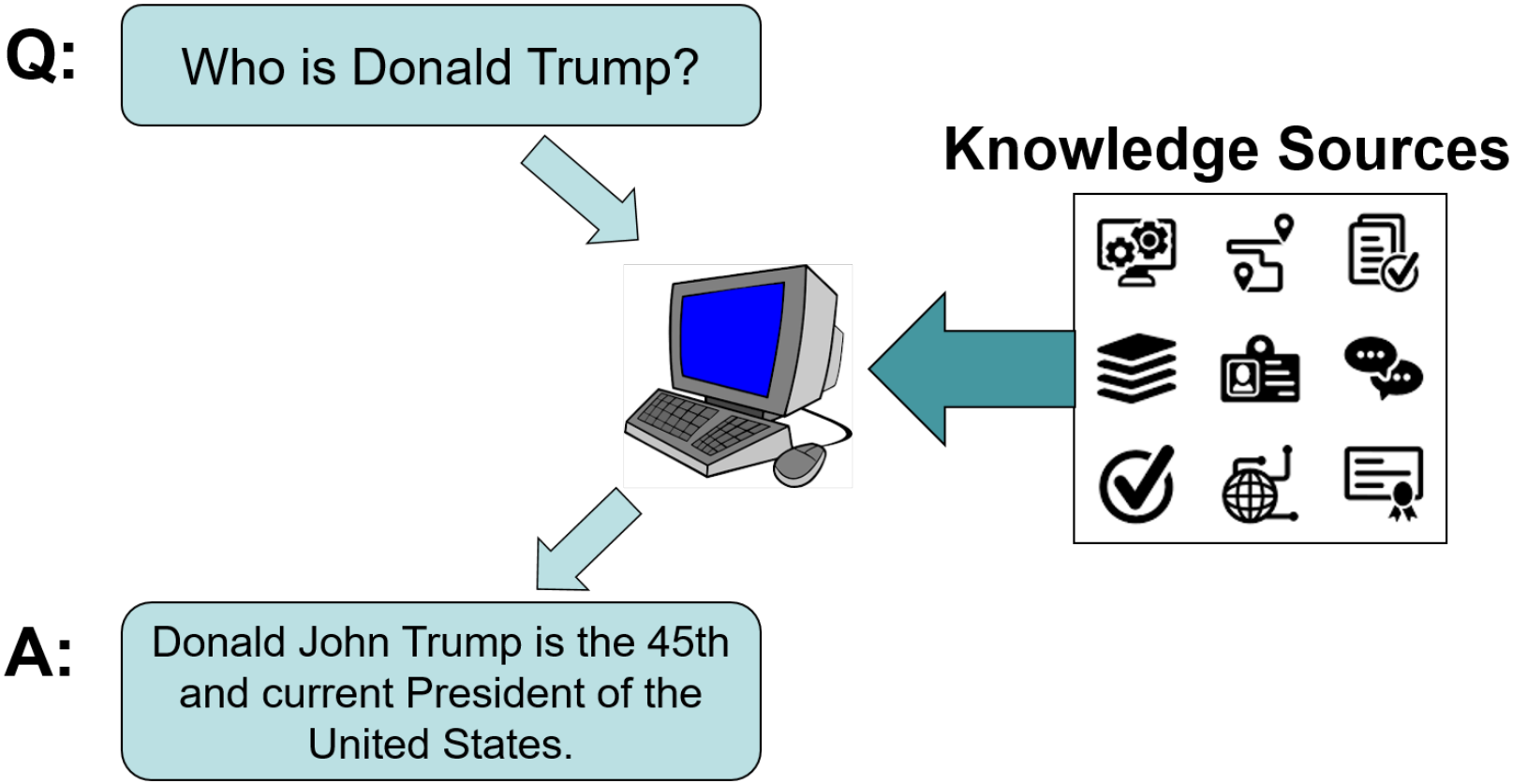
Because QA is awesome

1. QA is an AI-complete problem.
   If we solve QA, we have solved every other problem, too.
2. Many immediate and obvious applications
   Search, dialogue, information extraction, summarisation, ...
3. Some pretty nice results already
   IBM Watson and Jeopardy!, Siri, Google Search ...
4. Lots left to do!
   Plenty of interesting research and hard problems as well as low-hanging fruit.

Thomas Lukasiewicz, Advanced Machine Learning: Deep Learning for NLP: Lecture 11: Question Answering, 2019

Content

Question answering



Content

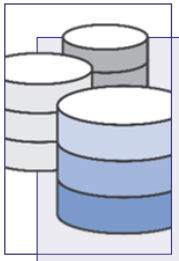
Knowledge sources

Text

Rules & Facts

Web

QA Archive



Database

Knowledge Graph