

Question Answering over Knowledge Graphs

Prof. Dr. Jens Lehmann
Smart Data Analytics Group
University of Bonn & Fraunhofer IAIS

The 2017 Voice Report Predicts More Than 24 Million Amazon Echo And Google Home Devices Will Be Sold This Year

Chatbots are revolutionizing customer support

ALEX DEBECKER, UBISEND @ALEXDEBECKER SEPTEMBER 5, 2016 12:10 PM

MARGARET RHODES DESIGN 02.11.16 7:01 AM

**WITH QUARTZ'S APP, YOU DON'T
READ THE NEWS. YOU CHAT WITH IT**

Chatbots Raised Over \$170M+ in 8 Months



Innovate

Baidu is bringing AI chatbots to healthcare

How Analytics, Big Data and AI Are
Changing Call Centers Forever



Definition of QA

Definition adopted from Hirschman and Gaizauskas [2]:

- 1 Ask question in natural language

Example (Natural Language)

Which books are written by Dan Brown?



Definition of QA

Definition adopted from Hirschman and Gaizauskas [2]:

- 1 Ask question in natural language
- 2 People use own terminology

Example

Which books are *written by* Dan Brown?

Which books have Dan Brown as one of their *authors*?

What are *notable works* of Dan Brown?




Definition of QA

Definition adopted from Hirschman and Gaizauskas [2]:

- 1 Ask question in natural language
- 2 People use own terminology
- 3 Receive a concise answer

Which books are written by Dan Brown?

Dan Brown / Books



Book Title	Year
The Da Vinci Code	2003
Inferno	2013
The Lost Symbol	2009
Angels & Demons	2000
Digital Fortress	1998

The Official Website of Dan Brown
www.danbrown.com/ ▼

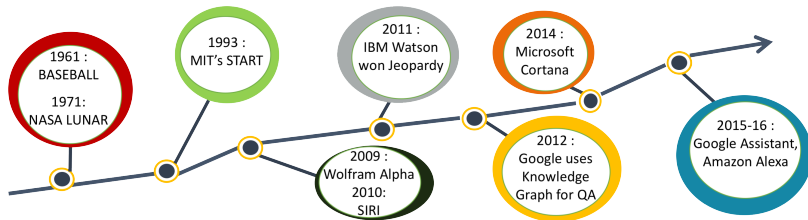
Inferno. In this riveting new thriller, Brown returns to his element and has crafted his highest-stakes novel to date. **The Lost Symbol**. Fact: In 1991, a document was locked in the safe of the director of the CIA. **The Da Vinci Code**. **Angels & Demons**. Deception Point. **Digital Fortress**.
[Inferno](#) - [The Lost Symbol](#) - [Deception Point](#) - [Digital Fortress](#)

Example (Formal Language)

```
select ?book {?book dbo:author dbr:Dan_Brown.}
```

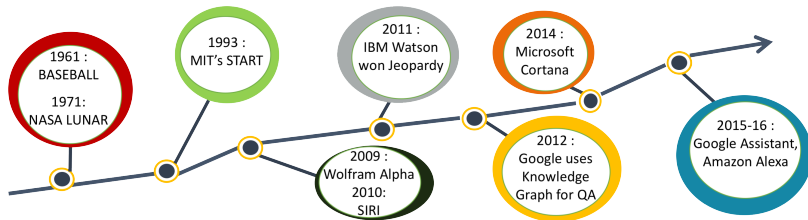



Question Answering Timeline





Question Answering Timeline

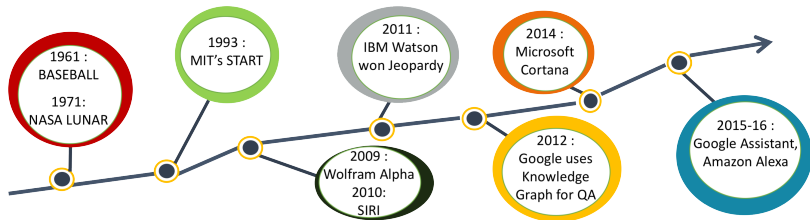


BASEBALL [1]:

- First QA system in 1961
- Answered questions about the US baseball league over a period of one year



Question Answering Timeline

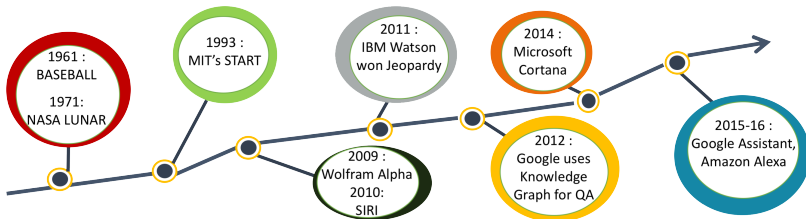


LUNAR [3] 1971:

- Developed for NASA to answer questions on lunar geology
- First evaluated QA system with 78 % correctly answered questions from lunar science convention visitors in 1971
- Compiles English into a “meaning representation language”



Question Answering Timeline



What is the fifty-second smallest country by GDP per capita? ☆ ☰



☰ Examples ↔ Random

Input interpretation:

52 nd smallest country	by	GDP	nominal per capita
-----------------------------------	----	-----	-----------------------

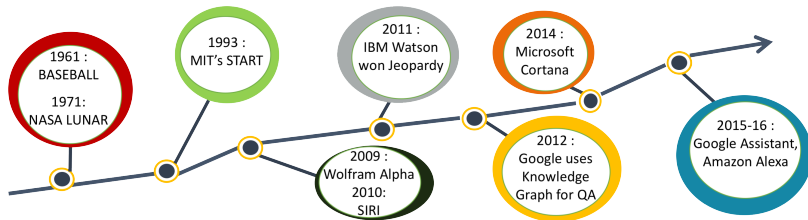
Result:

Djibouti

(\$1814 per year per person)

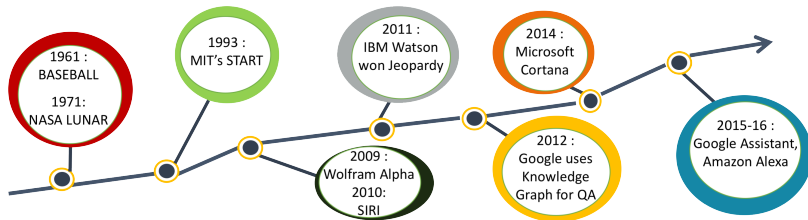


Question Answering Timeline





Question Answering Timeline



"Alexa, turn on
Welcome Home"

"Alexa, turn on my
Chill Time"



"Alexa, turn off my
Bedroom Sonos"

"Alexa, turn on
the TV"

Hi, how can I help?

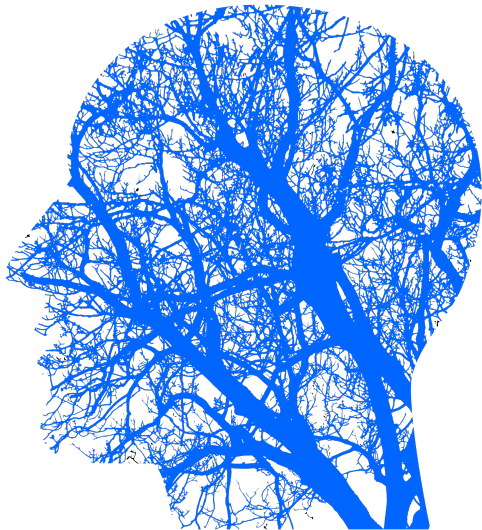
how is the weather in Bonn

Don't forget your sunglasses

31°C in Bonn
Sunny - High: 33° Low: 16°
Precip: 10%

32°	31°	30°	28°	24°	21
0%	0%	0%	0%	0%	2%
18	19	20	21	22	23

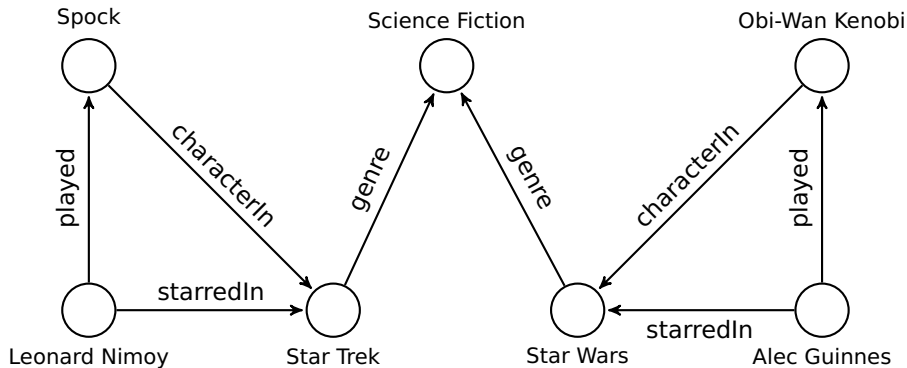
More on weather.com



1 Knowledge Graphs



Knowledge Graph Example



Labeled, directed multigraph



Popular Knowledge Graphs



2007



2007



2008



NELL 2010



Knowledge Graph 2012



Microsoft Satory 2012

facebook

2013

LinkedIn

2016



Knowledge Graph in Google Search



Alan Turing



All

Images

Videos

News

Books

More ▾

Search tools

About 499.000 results (0,65 seconds)

Alan Turing - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Alan_Turing ▾

Alan Mathison Turing OBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He was highly ...

[Joan Clarke](#) · [Turing machine](#) · [Enigma machine](#) · [Chemical castration](#)

Alan Turing – Wikipedia

https://de.wikipedia.org/wiki/Alan_Turing ▾ [Translate this page](#)

Alan Mathison Turing OBE, FRS [ˈælən ˈmæθɪsən ˈtjʊərɪŋ] (* 23. Juni 1912 in London; † 7. Juni 1954 in Wilmslow, Cheshire) war ein britischer Logiker, ...

[Enigma](#) · [Turing-Test](#) · [Turingmaschine](#) · [Turing-Bombe](#)

Alan Turing: the enigma

www.turing.org.uk/ ▾

Alan Turing (1912-1954). Large website by Andrew Hodges, biographer.

Alan Turing - a short biography

www.turing.org.uk/publications/dnb.html ▾

This short biography, based on the entry for the written in 1995 for the Oxford Dictionary of Scientific Biography, gives an overview of Alan Turing's life and work.

Alan Turing - Educator, Mathematician - Biography.com

www.biography.com/people/alan-turing-6512017 ▾



Alan Turing

Computer scientist



Alan Mathison Turing OBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. [Wikipedia](#)

Born: June 23, 1912, Maida Vale, London, United Kingdom

Died: June 7, 1954, Wilmslow, United Kingdom

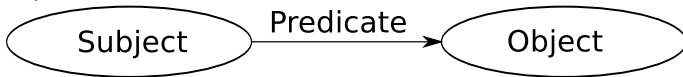
Education: [Princeton University](#) (1936–1938), [more](#)

Parents: [Julius Mathison Turing](#), [Ethel Sara Stoney](#)



RDF and the Linked Data Principles

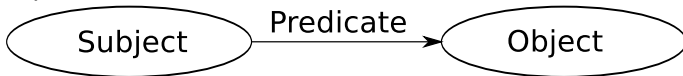
RDF Triple / Fact:





RDF and the Linked Data Principles

RDF Triple / Fact:



Example:

dbr:Germany dbo:hasCapital dbr:Berlin
Subject *Predicate* *Object*

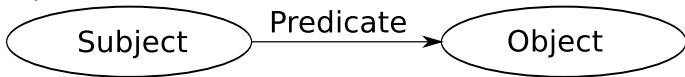
dbo : <http://dbpedia.org/ontology/>

dbr : <http://dbpedia.org/resource/>



RDF and the Linked Data Principles

RDF Triple / Fact:



Example:

dbr:Germany dbo:hasCapital dbr:Berlin
Subject *Predicate* *Object*

dbo : <http://dbpedia.org/ontology/>

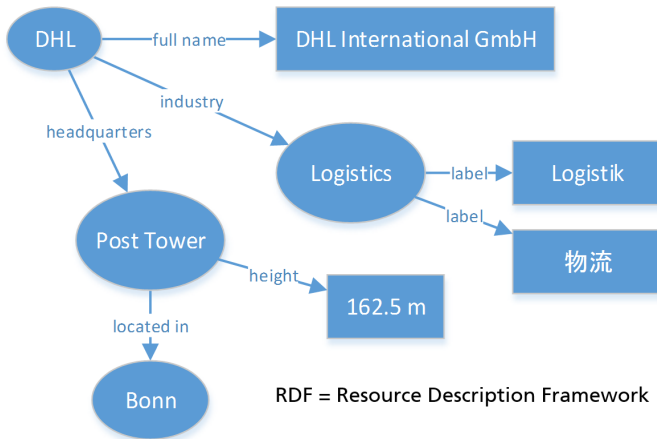
dbr : <http://dbpedia.org/resource/>

Linked Data principles (simplified version):

- 1 Use RDF and URLs as identifiers
- 2 Include links to other datasets



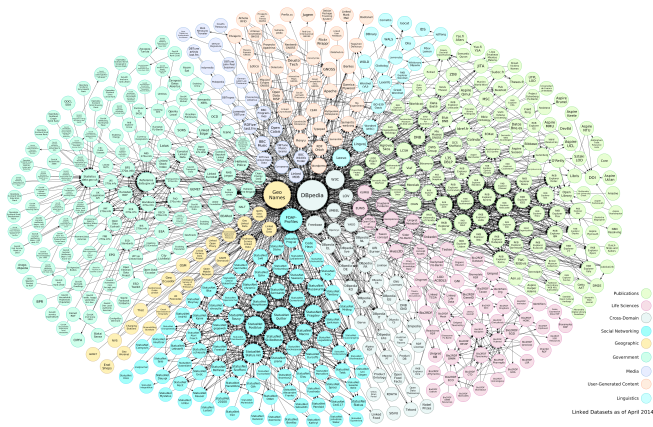
Linked Data / RDF Graphs



(URLs not shown; a set of triples is strictly not a graph but often labeled as such)



Linked Open Data Cloud



10000+ open datasets and 100+ billion facts



② QA over KG Applications

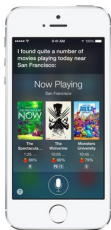


Personal Assistants

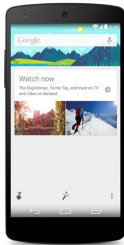
- Personal assistants on smartphones with voice interface
- Users expect devices to have **encyclopedic knowledge**
- Also becoming viable for other medium and large enterprises

“The world’s knowledge in your pocket”

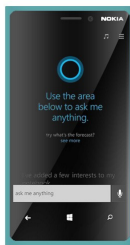
Apple Siri



Google Now



Windows Cortana





Entertainment & Smart Home



- Will get increasingly complex → people will ask devices more complex questions
- Example: QA on knowledge graphs plus images/videos
- Networked devices (Internet of Things)



When other devices are inconvenient . . .

Is this substance
toxic?

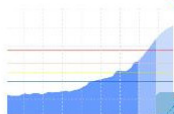


When will the wave
arrive here?



Integrated Overview

RDF / AMQP



- Crisis situations
- In-car QA systems (spatial and temporal QA)
- Childcare ;-)



Search

- QA is becoming part of main stream search engines
- Google:
 - Knowledge Graph in 2012
 - Question Intent Understanding in 2015
 - Can understand superlatives, ordered items, time e.g. "Who was the U.S. President when the Angels won the World Series?"
- Also relevant for **enterprise search in medium and large enterprises**



Google Alan Turing

All Images Videos News Books More Search tools

About 499,000 results (0.65 seconds)

Alan Turing - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Alan_Turing
Alan Mathison Turing OBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He is highly ...
Joan Clarke · Turing machine · Enigma machine · Chemical castration

Alan Turing - Wikipedia
https://de.wikipedia.org/wiki/Alan_Turing · Translate this page
Alan Mathison Turing OBE, FRS [ˈælən ˈmæθɪsən ˈtʃoʊrɪŋ] (* 23. Juni 1912 in London; † 7. Juni 1954 in Wilmslow, Cheshire) war ein britischer Logiker, ...
Enigma · Turing-Test · Turingmaschine · Turing-Bombe

Alan Turing: the enigma
www.turing.org.uk/ ·
Alan Turing (1912-1954). Large website by Andrew Hodges, biographer.

Alan Turing - a short biography
www.turing.org.uk/publications/dnb.html ·
This short biography, based on the entry for the written in 1995 for the Oxford Dictionary of Scientific Biography, gives an overview of Alan Turing's life and work.

Alan Turing - Educator, Mathematician - Biography.com
www.biography.com/people/alan-turing-9512017 ·
Jul 5, 2016 - Find out more about logician and mathematician Alan Turing, including what he proved in his paper, 'On Computable Numbers,' at ...

Enigma codebreaker Alan Turing receives royal pardon | Science ...
www.theguardian.com/science/alan-turing
Dec 24, 2013 - Alan Turing, the second world war codebreaker who took his own life after undergoing chemical castration following a conviction for ...

Alan Turing | Science | The Guardian
<https://www.theguardian.com/science/alan-turing> ·
Apr 16, 2016 - GCHQ chief apologises for 'horrifying' treatment of Alan Turing Alan Turing notebook sells for more than \$1m at New York auction.

8 things you didn't know about Alan Turing | PBS NewsHour
www.pbs.org/newshour/updates/8-things-didnt-know-alan-turing/ ·
Nov 28, 2014 - Alan Turing is hailed as the father of the computer, single-handedly halting the

Alan Turing
Computer scientist

Alan Mathison Turing OBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. He is highly ...

Born: June 23, 1912, Maida Vale, London, United Kingdom
Died: June 7, 1954, Wilmslow, United Kingdom
Education: Princeton University (1936–1938), more
Parents: Julius Mathison Turing, Ethel Sara Stoney

Quotes View 4+ more

We can only see a short distance ahead, but we can see plenty there that needs to be done.

A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.

Machines take me by surprise with great frequency.

Books

<https://www.google.de/search?client=ubuntu&channel=fs&biw=1366&bih=638&q=Ethel+Sa+nlVRVfEnYkkyANTmfakvAAAA&...>



alan turing date of birth



All Images News Videos Shopping More Search tools

About 565,000 results (0,79 seconds)

Alan Turing / Date of birth

June 23, 1912



Charles Babbage
December 26, 1791



Benedict Cumberbatch
July 19, 1976



Albert Einstein
March 14, 1879

Feedback

People also ask

Who is Alan Turing?



Why is Alan Turing famous?



Who cracked the Enigma machine?



Alan Turing



Computer scientist

Alan Mathison Turing OBE FRS was a pioneering English computer scientist, mathematician, logician, cryptanalyst and theoretical biologist. [Wikipedia](#)

Born: June 23, 1912, Maida Vale, London, United Kingdom

Died: June 7, 1954, Wilmslow, United Kingdom

Education: Princeton University (1936–1938), [more](#)

Parents: Julius Mathison Turing, Ethel Sara Stoney

Quotes

[View 4+ more](#)

We can only see a short distance ahead, but we can see plenty there that needs to be done.

A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.

Machines take me by surprise with great frequency.

Books



**Wendy Baker (Wendy Steger)**[Add Friend](#)

Ambassador at Plexus Worldwide

Lives in Fond du Lac, Wisconsin · From Theresa, Wisconsin

Likes Fishing, AMC and 93 others

5 mutual friends including Christine Ann and Keetra Baker

**Travis Sabel**[Friends](#)[1 new post](#)

Lives in Fond du Lac, Wisconsin · From Fond du Lac, Wisconsin

Likes Fishing, Drury Outdoors and 27 others

6 mutual friends including Chris Lembo Flitter and Troy Sabel

**Steven Davis**[Add Friend](#)

Fond du Lac, Wisconsin

Lives in Fond du Lac, Wisconsin · From Fond du Lac, Wisconsin

Likes Fishing, Tastemade and 3 others

12 mutual friends including Laura Baumann and Christine Ann

**Karyn Slater**[Add Friend](#)

Surgical Technologist at St Agnes Hospital

Lives in Fond du Lac, Wisconsin · From Fond du Lac, Wisconsin

Likes Fishing, ABC Television Network and 22 others

6 mutual friends including Mary Cruciani and Anna Marie

**Troy Sabel**[Friends](#)

Works at Ultratech Tool and Design

Lives in Fond du Lac, Wisconsin · From Fond du Lac, Wisconsin

Likes Fishing and 39 others

6 mutual friends including Travis Sabel and Anna Marie

**Tj Lefebber**[Add Friend](#)

Assembly Line at Alliance Laundry Systems

Lives in Fond du Lac, Wisconsin

Likes Fishing, NBC Olympics and 40 others

TRENDING

[iPhone 7](#)

1M people talking about this

[Teddy Bridgewater](#)

21K people talking about this

[McChicken](#)

14K people talking about this

[Ice Road Truckers](#)

50K people talking about this

[Dancing with the Stars](#)

13K people talking about this

[Tony Stewart](#)

16K people talking about this

[Bea Arthur](#)

16K people talking about this

[Arian Foster](#)

12K people talking about this

[Mary Shelley](#)

9.3K people talking about this

[Apple Pencil](#)

2.5K people talking about this

[Learn More](#)

SUGGESTED PAGES

[See All](#)**Adventures in Tail Wagging**

75 people like this.

[Like](#)**Rapalowicks**

89 people like this.

[Like](#)**What's That Paint?**

263 people like this.

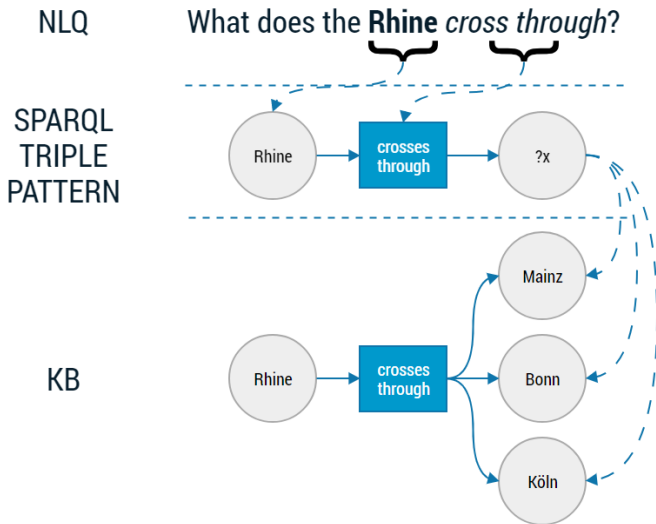
[Like](#)**A Gifted Amateur - Short**



3 Challenges for QA on KGs



Goal of QA over RDF Datasets





(Almost) No Challenge

Who developed Minecraft?

```
SELECT DISTINCT ?uri
WHERE {
  dbpedia:Minecraft dbpedia-owl:developer ?uri .
}
```



How **tall** is Michael Jordan?

```
SELECT DISTINCT ?num
WHERE {
  dbpedia:Michael_Jordan dbpedia-owl:height ?num .
}
```



Lexical Knowledge

Give me all taikonauts.

```
SELECT DISTINCT ?uri
WHERE {
  ?uri rdf:type dbpedia-owl:Astronaut .
  ?uri dbpedia-owl:nationality dbpedia:China .
}
```



Lexical Knowledge

- Who was the **last** princess of Joseon?
- !BOUND :successor
- Which of the Beatles is **still alive**?
- !BOUND :deathDate



Ambiguity

How many **banks** are there in London?

```
SELECT DISTINCT count(?bank)
WHERE {
  ?bank a dbo:Bank.
  ?bank dbp:location dbr:London.
}
```

```
SELECT DISTINCT count(?bank)
WHERE {
  ?bank a lgdo:Riverbank.
  ?bank dbp:location dbr:London.
}
```



Adjectives

Give me all **communist** countries.

```
SELECT DISTINCT ?uri
WHERE {
  ?uri rdf:type dbpedia-owl:Country .
  ?uri dbpedia-owl:governmentType dbpedia:Communism .
}
```



Complex Queries

Who are the parents of the wife of **Juan Carlos I** ?

```
SELECT DISTINCT ?uri
WHERE {
  dbpedia:Juan_Carlos_I_of_Spain dbpedia-owl:spouse ?x .
  ?x dbpedia-owl:parent ?uri .
}
```



Complex Queries

How many companies were founded in the same year as Google?

```
SELECT COUNT(DISTINCT ?c)
WHERE {
  ?c rdf:type dbo:Company .
  ?c dbo:foundingYear ?year .
  res:Google dbo:foundingYear ?year .
}
```



Scalability & Real World Problems

- Distributed data
- aggregating information from different datasets
(Question Answering over linked data)
- Multilingual data
- Conflicting data / versioning
- Missing and incomplete data
- Spatial data queries
- Temporal questions



Non-factoid Questions

- What is the difference between impressionism and expressionism?
- How do histone methyltransferases cause histone modification?
- Is honesty the best policy?
- Are donuts tasty?



Unanswerable Query

What is the best fruit to eat?

→ personal preference

Who will be the next president of USA?

→ unknown

If I ate myself, would I be twice the size as now or would I disappear completely?

Will QA systems be expected to detect unanswerable queries (and provide entertaining answers) in the future?

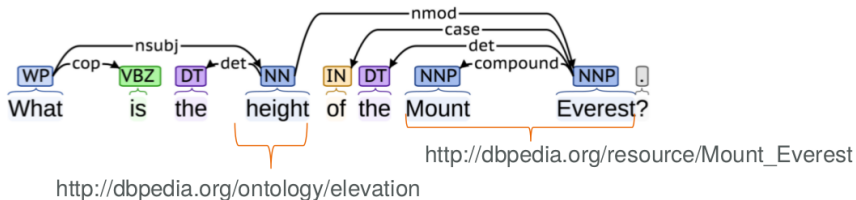


④ Question Answering Approaches



Semantic Parsing Approaches

- Explicitly represents the grammatical and semantical relationships between the words of the sentences
- Step 1: Forms a semiformal intermediate structure devoid of lexical and semantic variations
- Step 2: Informal structure is then converted into a logical form (see e.g. AskNow, XSER systems)





Addressing the Lexical Gap

- Add multiple labels to the knowledge graph
- String normalization (Stemming/Lemmatizing)
- String similarity functions
- Automatic query expansion e.g. via WordNet → improves recall, reduces precision
- More difficult: properties
 - No continuous substrings: In which **place** was Keira Knightley **born**? (relation *birthPlace*)
 - Expressed as nouns or verbs : Who **wrote** Braveheart? , Who was the **writer** of Braveheart? (relation *writer*)
 - "Solution": pattern libraries (BOA, Patty, Paralex)
- Word embeddings





Addressing Ambiguity

Who founded Apple?



Can I pickle apples?



Addressing Ambiguity

- Same phrase has different meanings (classical example: Jaguar)
- High lexical gaps reduce recall, ambiguity reduces precision
- Context can be used for disambiguation (neighbor words, POS tags, parse tree structure)
- Exploiting the structure of the knowledge graph i.e. maximizing semantic relatedness measures of entity candidates in the question
- User interaction e.g. inductive logic programming on given answers





Addressing Complex Queries

- IBM Watson:
 - Decomposes complex question into a set of simple question
 - Hypothesis generation and evidence scoring using multiple analysis algorithm.
 - Final confidence merging and ranking to generate answer
- AskNow:
 - Represent complex natural question as an intermediate canonical structure
 - Predicate and entity linking
 - Conversion of canonical queries into SPARQL using template fitting algorithm



Publications for Challenges (up to early 2015)

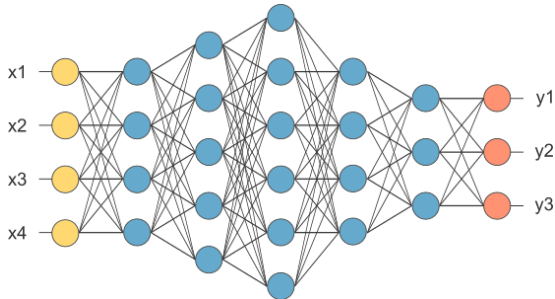
Year	Total Lexical Gap Ambiguity Complex Operators			
absolute				
2010	1	0	0	0
2011	16	11	12	3
2012	14	6	7	2
2013	20	18	12	5
2014	13	7	8	2
2015	6	5	3	0
all	70	46	42	12
percentage				
2011		68.8	75.0	18.8
2012		42.9	50.0	14.3
2013		85.0	60.0	25.0
2014		53.8	61.5	15.4
all		65.7	60.0	17.1



Motivation of End-to-End Systems

Problem of Semantic Parsing Approaches:

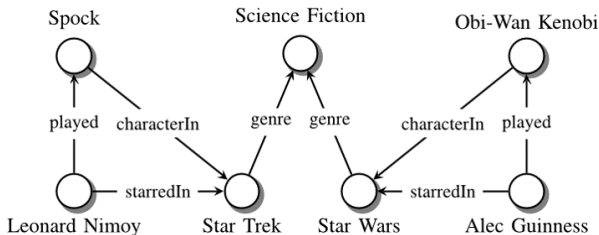
- Complex pipelines
- A lot of manual configuration
- Error propagation along the pipeline \rightarrow accuracy (a value between 0 and 1) after each step (roughly) multiplies
- \rightarrow Interest in end-to-end approaches





End-to-End Question Answering

- No traditional Natural Language Processing (NLP) – the QA System essentially **knows nothing about language**
- No intermediate logical structures – answers are generated directly from the questions („end-to-end“)
- Current approaches answer simple questions, i.e. involving one relation and one entity
 - Where is the southern cross travel insurance Australia located?
 - Which characters played in Star Trek?

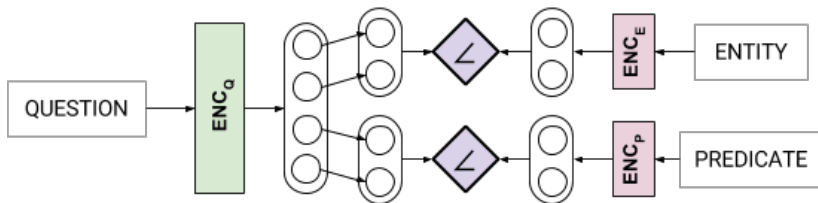




End-to-End Question Answering

Lukovnikov et. al (WWW'2017):

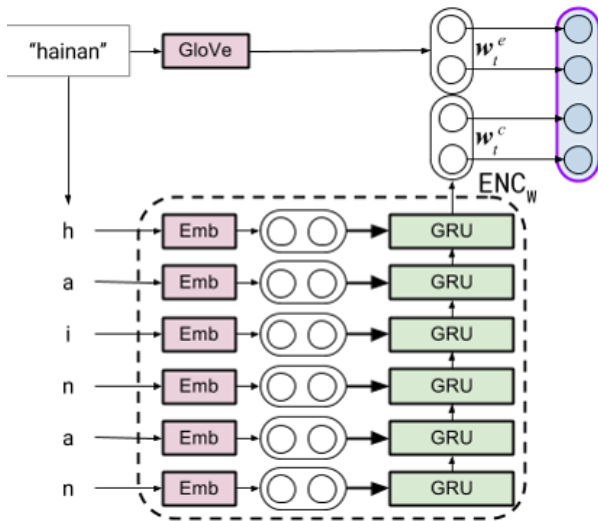
- Questions are mapped to vector representation (encoding)
- Entities (nodes) and predicates (edges) in the knowledge graph are encoded as well
- Representation of question and answer is compared using similarity measures





End-to-End Question Answering

- Each word represented via **word embeddings** (capturing the semantics of the word)
- Each **character** of the word is also passed through encoder (capturing the rare/unknown words)





Evaluation Facebook AI bAbI QA Dataset

Approach	Year	Accuracy %
Fraunhofer IAIS Dev	2017	74.8
Fraunhofer IAIS WWW	2017	71.2
Golub and He [A]	2016	70.9
Yin et al. [B]	2016	68.3
Bordes et al. [C]	2015	62.7
Dai et al. [D]	2016	62.6

- Comparisons with other end-to-end systems (pre-processing in other system achieves up to approx. 75% accuracy)

- A. D. Golub and X. He. Character-level question answering with attention. EMNLP, 2016.
- B. W. Yin et. Al. Simple question answering by attentive convolutional neural network. In COLING 2016, December 2016, Osaka, Japan, ACL 2016
- C. A. Bordes, N. Usunier, S. Chopra, and J. Weston. Large-scale simple question answering with memory networks. CoRR, abs/1506.02075, 2015.
- D. Z. Dai, L. Li, and W. Xu. Cfo: Conditional focused neural question answering with large-scale knowledge bases. arXiv preprint arXiv:1606.01994, 2016.



Semantic Parsing or End-to-End?

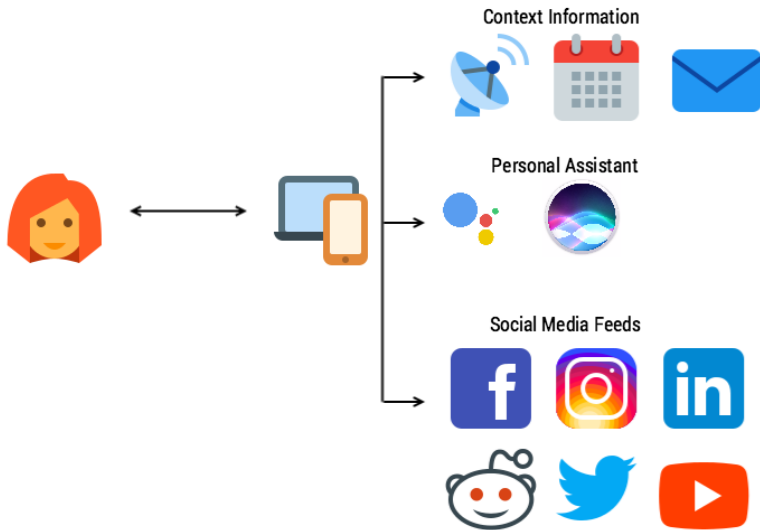
Type	Pros	Cons
Semantic Parsing	Less/no training data needed Can handle complex questions	More manual adaption work
End-to-End	Less manual adaption work Fast once the system is trained	More training data No robust theory for complex questions



⑤ Future Directions



Future Challenges: Context



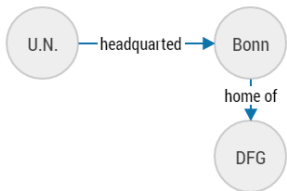


Future Challenges: Hybrid QA

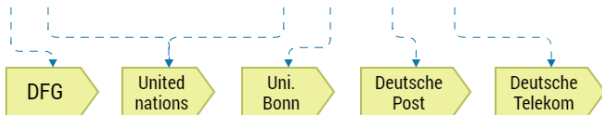
Name some institutions *located in Bonn*?

Structured

Unstructured

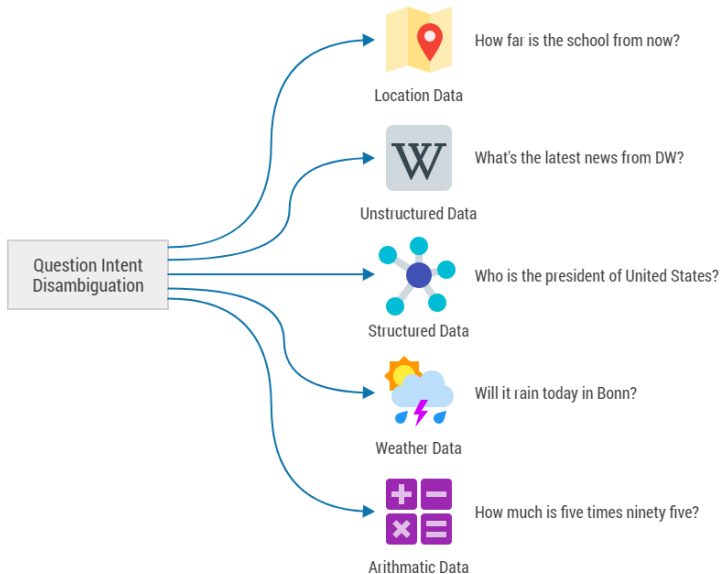


The headquarters of **Deutsche Post DHL** and **Deutsche Telekom**, both **DAX-listed** corporations, are in Bonn. The city is home to the **University of Bonn** and a total of 19 **United Nations** institutions.



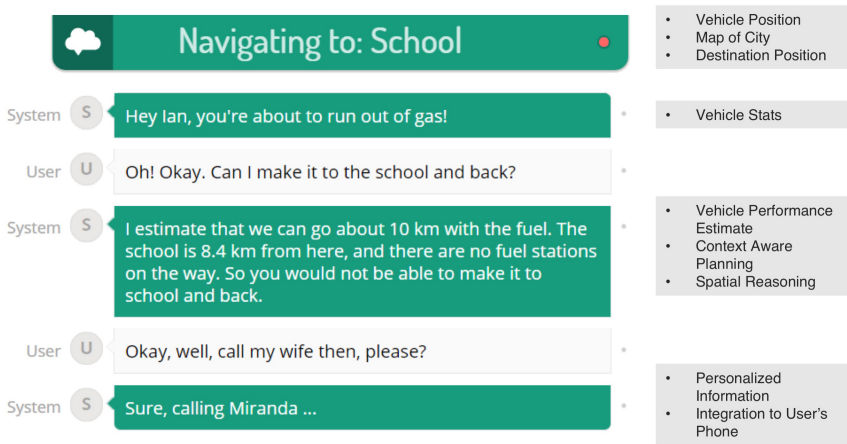


Future Challenges: QA from Several Sources





Future Challenges: APIs and QA



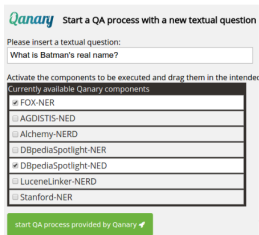


Future Challenges: Out-of-scope Questions

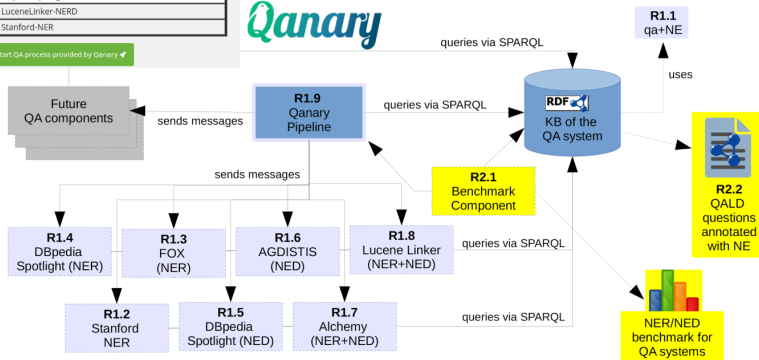
- QA system should know when question cannot be answered (very important in IBM Watson)
- Examples of OOS:
 - Knowledge base cannot answer it: Is my mother at home now?
 - Question refers to the future with no forecasts available: When will the world record for the fastest supercomputer be broken?
 - Unknown question type (e.g. asking non-factual question to factoid system): How honest is the government with me?
- Systems need to be benchmarked on OOS questions, otherwise they are encouraged to guess



Future Challenges: QA Architectures



Qanary - an Open Knowledge-based Architecture for Question Answering





Further Challenges

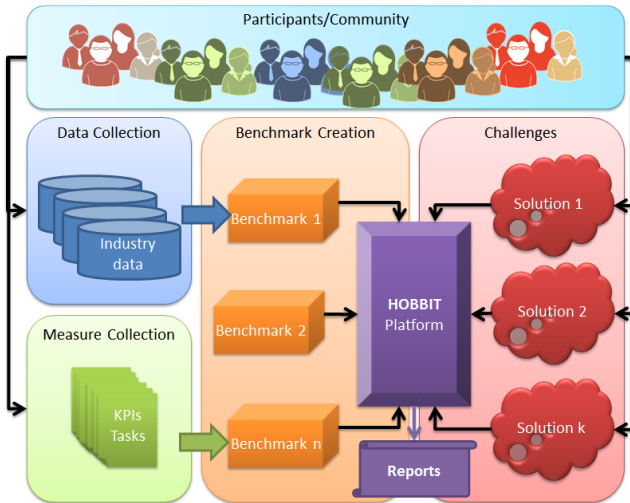
- Complex queries in end-to-end systems
- Better relation extraction
- Multilinguality
- Automatic template learning for semantic parsing systems
- Spatial and temporal questions
- User interaction requiring minimal user effort
- Integration of QA and dialogue systems





Want to build and test your own system?

HOBBIT Project





Want to build and test your own system?

QALD ("Question Answering over Linked Data")

- 7 editions so far
- Co-located with major conferences
- Datasets:
 - Multilingual question answering over DBpedia
 - Hybrid question answering (text and structure)
 - Large-scale QA → runtime performance
 - Wikidata question answering
- Coming up: 5000 question dataset including a number of interesting SPARQL query patterns!



Take Home Messages

- Knowledge graphs gaining popularity
- Wide range of applications for question answering over knowledge graphs
- Two main approaches: semantic parsing and end-to-end systems
- Active areas of research with many challenges – still many unsolved problems
- Benchmarks are available – you can explore and test your own ideas





The End



Jens Lehmann

<http://jens-lehmann.org>

lehmann@uni-bonn.de

<https://twitter.com/JLehmann82>

jens.lehmann@iais.fraunhofer.de

University of Bonn

Fraunhofer IAIS

Thanks for your attention!



References



B. F. Green Jr, A. K. Wolf, C. Chomsky, and K. Laughery.
Baseball: an automatic question-answerer.
In Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference, pages 219–224. ACM, 1961.



L. Hirschman and R. Gaizauskas.
Natural language question answering: the view from here.
natural language engineering, 7(4):275–300, 2001.



W. A. Woods.
Semantics for a question-answering system, volume 27.
Garland Pub., 1979.



TOUCHE

by ©Tom

