



Question Answering: Applications and Challenges

WDAqua R&D Week

*Contributions from Mohnish Dubey,
Konrad Höffner and Christina Unger*



2016-02-09

Jens Lehmann





QA Overview and Recent Developments



Definition of QA [3]

- ① Ask questions in natural language

Example (Natural Language)

Which books are written by Dan Brown?



Definition of QA [3]

- ① Ask questions in natural language
- ② People use their own terminology

Example

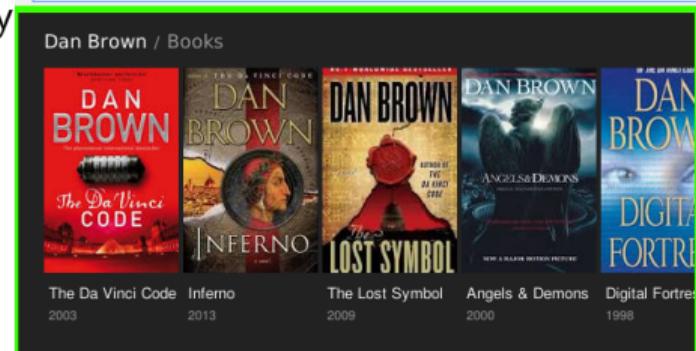
Which books are *written by* Dan Brown?

Which books have Dan Brown as their *author*?

What are *notable works* of Dan Brown?

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

Which books are written by Dan Brown?



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.
```

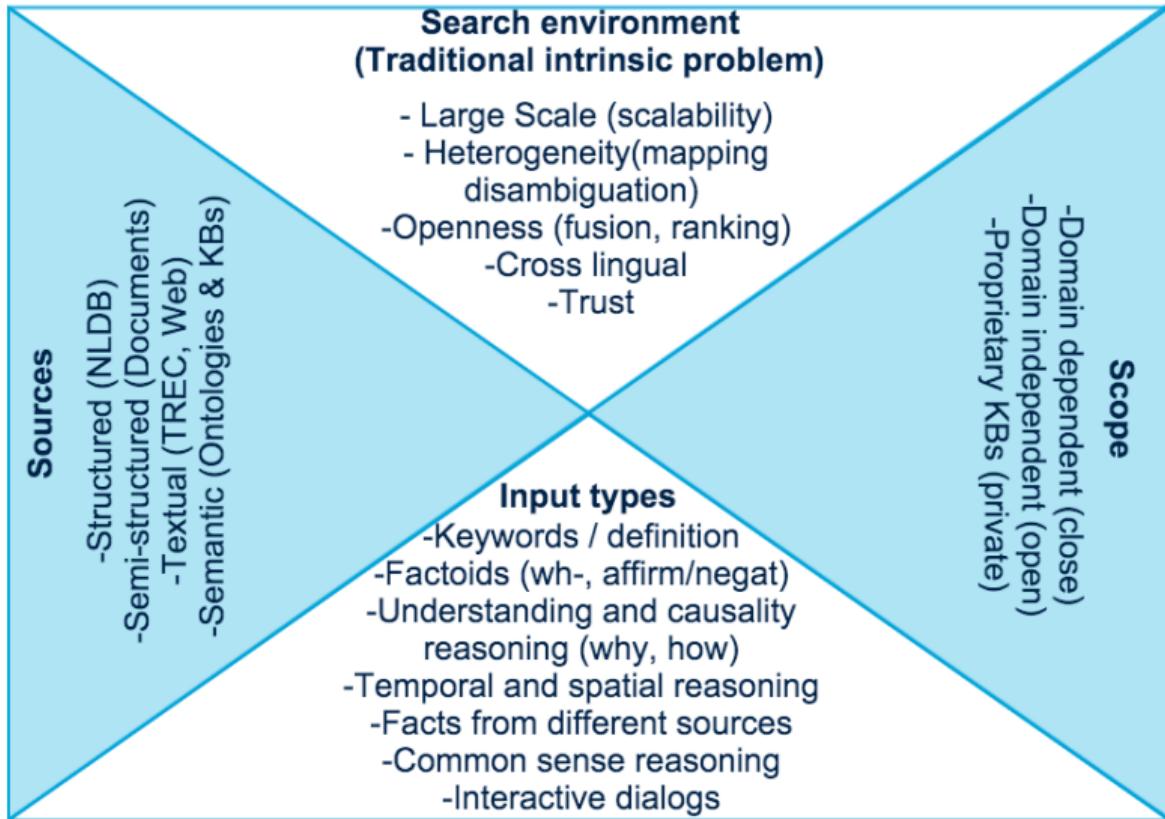
Motivation

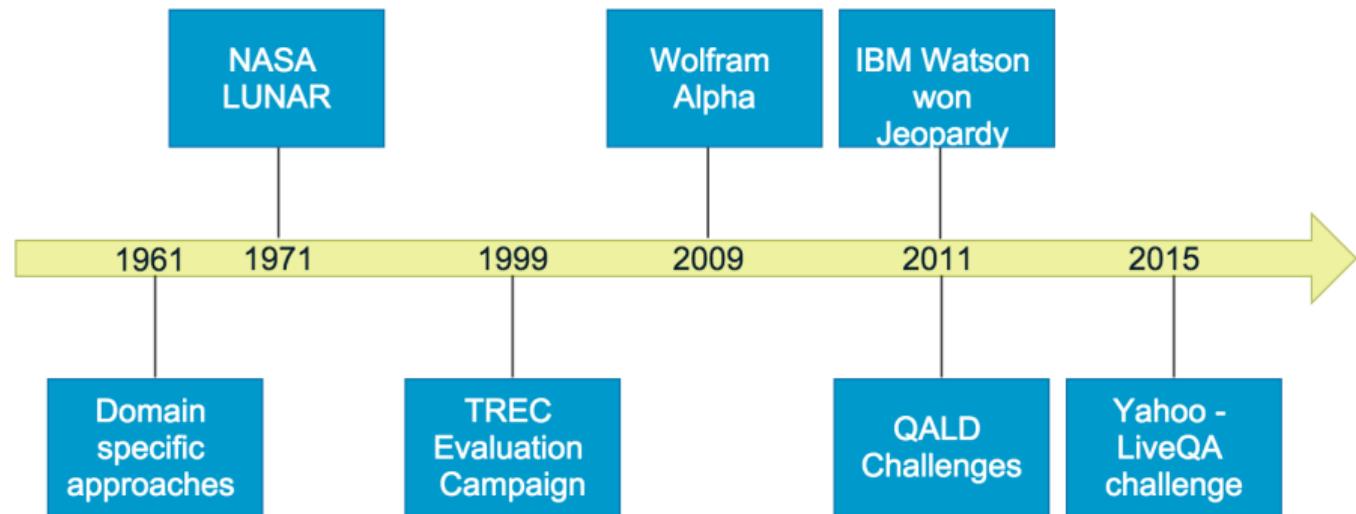
- Large amounts of knowledge modelled as RDF data
- Contrary to the Web of Documents, Semantic Web is made for machines
- Lay users cannot use SPARQL or do not know the schema

Approach	Expressiveness	Ease of Use	Ease of Setup
SPARQL Endpoint	++	--	++
Facetted Search	-	+	- ¹ /+ ²
Question Answering	+	++	-- ¹ /-- ²

¹new development

²existing approach, new domain







BASEBALL [2] 1961:

- First QA system (according to my research)
- Answered questions about the US baseball league over a period of one year

LUNAR [5] 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”



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



≡ Examples × Random



Input interpretation:



Result:

Djibouti

(\$1814 per year per person)



What was the population of Germany in 1950?



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



Input interpretation:

Germany	population	1950
---------	------------	------

Result:

68.4 million people (world rank: 7th) (1950 estimate)





- Developed by DeepQA research team at IBM
- Won Jeopardy against human experts
- Massively parallel, runs on a supercomputer
- Aggregates many different sources, including semantic
- Industry project: some algorithms are published, but not open source
- Brmson¹ platform approximates IBM Watson using open source technology
- QA engine Yoda QA² inspired by Watson runs on a normal server PC

¹<http://brmlab.cz/project/brmson>

²<http://ailao.eu/yodaqa/>



QA Challenges



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 .
```



Give me all **taikonauts**.

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



- 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 lgdo:Bank.
}
```

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



Give me a list of all **lakes in Denmark**.

```
SELECT DISTINCT ?uri
WHERE {
{ ?uri rdf:type yago:LakesOfDenmark . }
UNION
{ ?uri rdf:type dbpedia-owl:Lake .
?uri dbpedia-owl:country dbpedia:Denmark . }
}
```



Give me a list of all **lakes** in Denmark.

```
SELECT DISTINCT ?uri
WHERE {
{ ?uri rdf:type yago:LakesOfDenmark . }
UNION
{ ?uri rdf:type dbpedia-owl:Lake .
?uri dbpedia-owl:country dbpedia:Denmark . }
}
```



Who killed John Lennon?

```
SELECT DISTINCT ?uri
WHERE {
?uri rdf:type dbo:Person.
?uri dbp:conviction res:Death_of_John_Lennon.
```



Adjectives

Give me all **communist** countries.

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



Adjectives

Give me all animals that are **extinct**.

```
SELECT DISTINCT ?uri
WHERE {
?uri rdf:type dbpedia-owl:Animal .
?uri dbpedia-owl:conservationStatus 'EX' .
}
```



Clausal Query

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 .
}
```

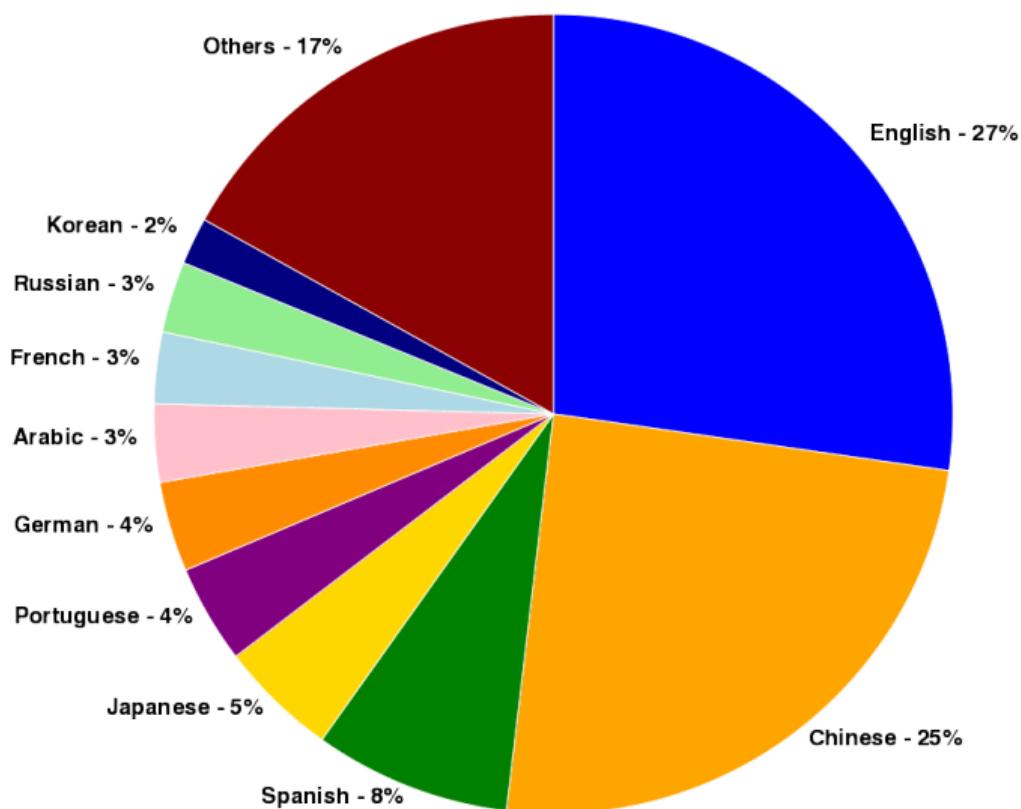


Crossover of QA Systems Required

What's the unemployment rate of females in European capitals in 2011 ?

- What are European Capitals?
 - QA (e.g. AskNow) over ontology answers e.g. "Berlin"
- What amount of the female population in Berlin were unemployed in 2011 ?
 - Now, the Data cube model and CubeQA could be used.

Ensemble learning and QA system combination.





¿Qué alto es el Hallasan?

Wie hoch ist der Hallasan?

Quelle est l'altitude du Hallasan?

한라산은 얼마나 높은가?

```
SELECT DISTINCT ?height
WHERE {
  dbpedia:Hallasan dbpedia-owl:elevation ?height .
```



Non-factoid questions

- What is the difference between impressionism and expressionism?
- How do I make cheese cake?
- What do most Americans think of gun control?
- What is the role of TRH in hypertension?
- How do histone methyltransferases cause histone modification?



Unanswerable Query

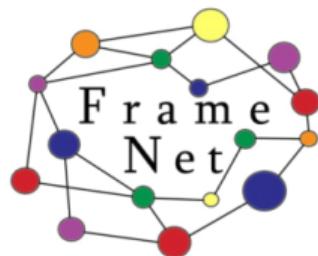
What is the best fruit to eat?

Who will be the next president of USA?

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



Approaching Challenges



a multilingual tree
encyclopedia

Wiktionary

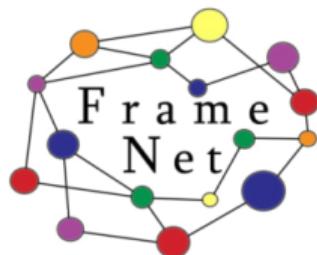
[ˈwɪkʃənri] *n.*,
a wiki-based Open
Content dictionary

Wiktionary



WordNet

A lexical database for English



a multilingual tree
encyclopedia

Wiktionary
[ˈwɪkʃənri] *n.*,
a wiki-based Open
Content dictionary

Wiktionary

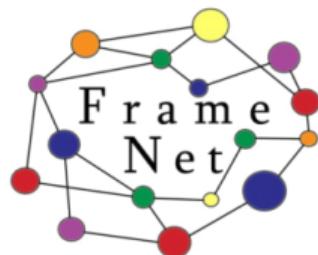
high

- POS: adjective
- forms: higher, highest
- related form: height
- POS: noun
- antonym: low



WordNet

A lexical database for English



a multilingual tree
encyclopedia

Wiktionary

[ˈwɪkʃənri] *n.*,
a wiki-based Open
Content dictionary

Wiktionary [ˈwɪkʃənri]

← ? →



OMEGAWIKI

WordNet

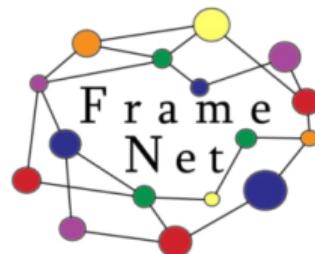
A lexical database for English



[Ontology lexica](#) capture rich linguistic information

- word forms
- part of speech
- subcategorization
- meaning

about how elements of a particular ontology are verbalized in a particular language.



a multilingual tree
encyclopedia

Wiktionary

[ˈwɪkʃənri] *n.*,
a wiki-based Open
Content dictionary

Wiktionary [ˈwɪkʃənri]



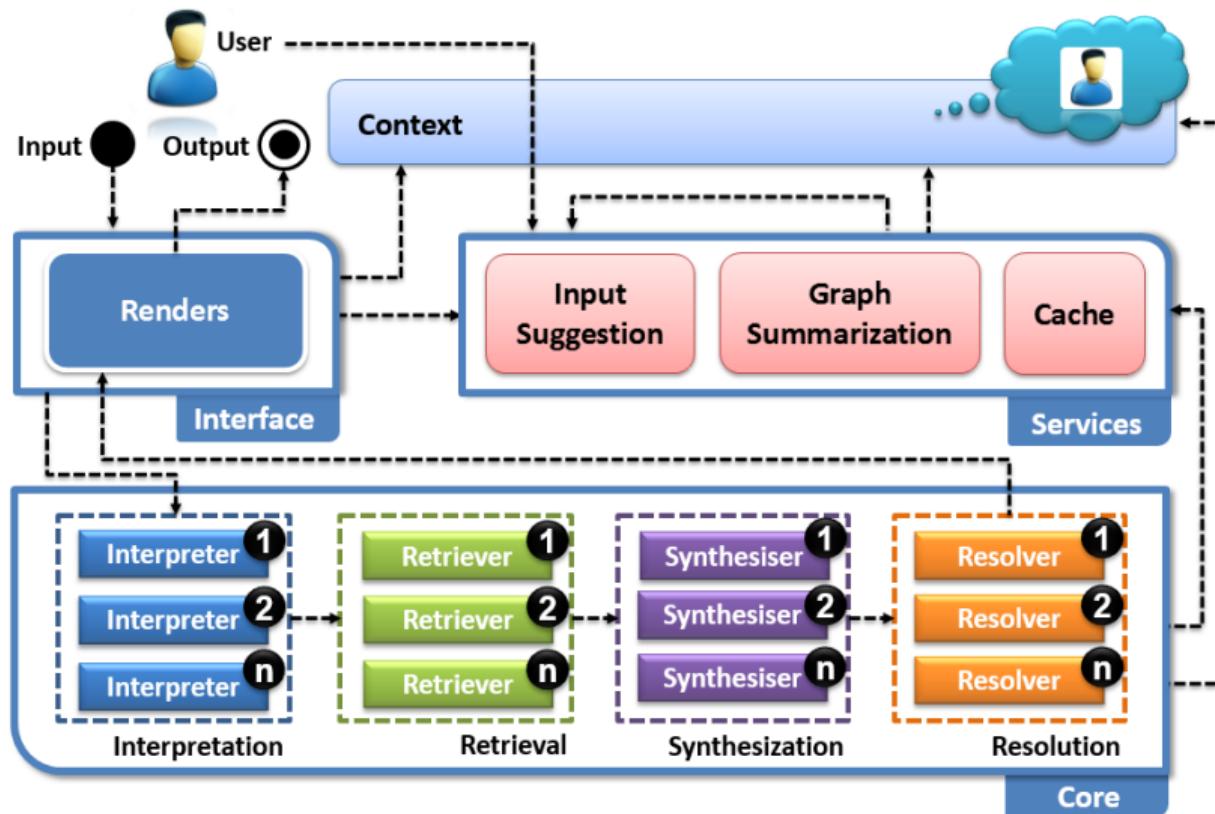
WordNet

A lexical database for English

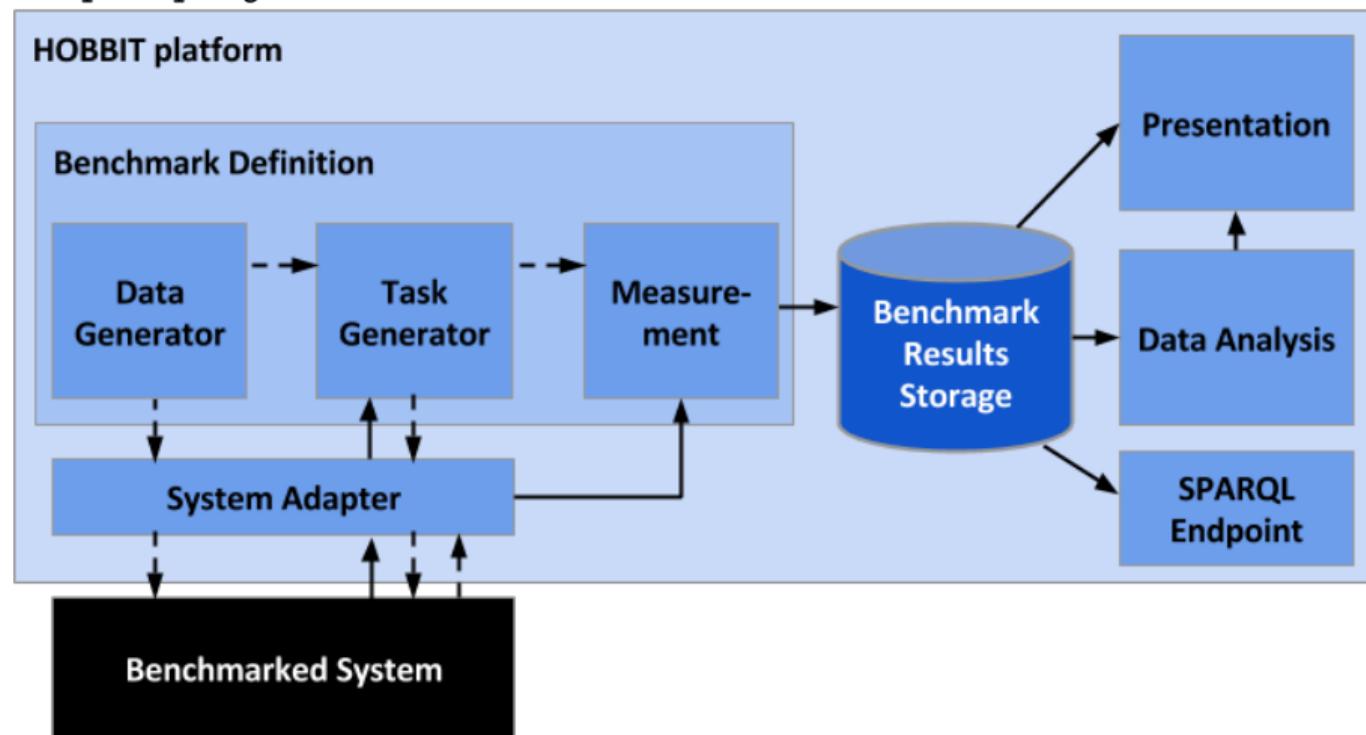


[lemon](#) provides a meta-model for describing ontology lexica with RDF.

- **Joint** building of meaning representations and instantiation w.r.t. dataset
 - Berant, Chou, Frostig, Liang: Semantic parsing on Freebase from question-answer pairs. EMNLP 2013.
- **Joint** instantiation w.r.t. dataset
 - Xu, Feng, Zhao: Xser.



EU Big Linked Data Benchmarking Project –
<http://project-hobbit.eu/>

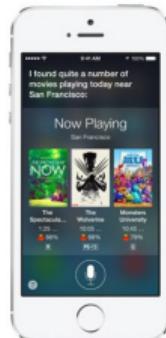




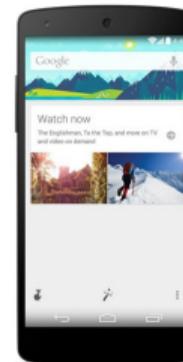
QA Applications

- Personal assistants on smartphones with voice interface
- Need small memory, CPU and battery footprint
- Calendar events
- Weather
- Route planning

Apple Siri

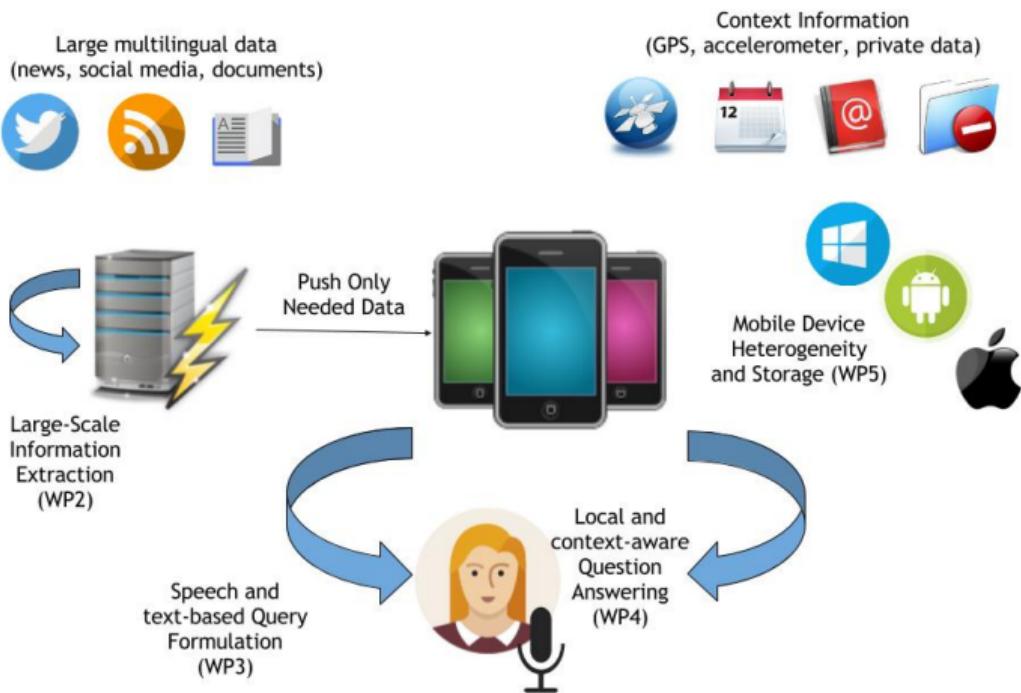


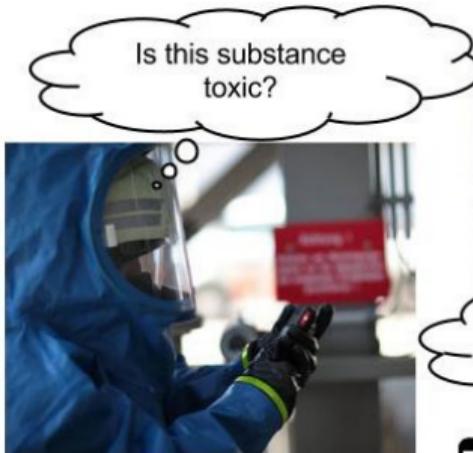
Google Now



Windows Cortana







- Crisis situations
- In-car QA systems
- Childcare ;-)



- Will get increasingly complex → people will ask devices more complex questions
- Networked devices (Internet of Things)

- QA / keyword search becoming part of main stream search engines
- Google:
 - Voice Search in 2008
 - Knowledge Graph in 2012
 - Question Intent Understanding in 2015
- Google can understand superlatives, ordered items, time e.g. "Who was the U.S. President when the Angels won the World Series?"





What can we do?

DON'T PANIC, ORGANIZE!



Join <https://www.w3.org/community/nli/>!



Prof. Dr. Jens Lehmann

<http://jens-lehmann.org>

<https://de.linkedin.com/in/jenslehmann82>

<https://twitter.com/JLehmann82>

lehmann@uni-bonn.de

jens.lehmann@iais.fraunhofer.de

University of Bonn

Fraunhofer IAIS

Thanks for your attention!

-  A. M. Gliozzo and A. Kalyanpur.
Predicting lexical answer types in open domain QA.
International Journal On Semantic Web and Information Systems,
8(3):74–88, July 2012.
-  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.

-  [V. Lopez, V. Uren, M. Sabou, and E. Motta.](#)
Is Question Answering fit for the Semantic Web?: A survey.
Semantic Web Journal, 2(2):125–155, April 2011.
-  [W. A. Woods.](#)
Semantics for a question-answering system, volume 27.
Garland Pub., 1979.

THÜCHE

by OTOM

