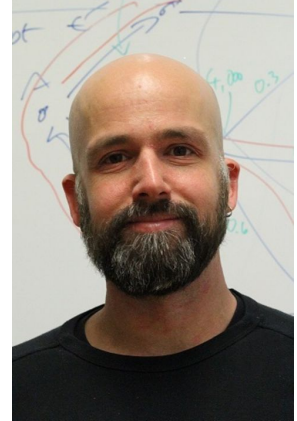


Knowledge Graphs Organization



Hello

- Prof. Dr. Heiko Paulheim
 - Chair for Data Science
- Research Interests:
 - Knowledge Graphs on the Web and their Applications
 - Data Quality and Data Cleaning on Knowledge Graphs
 - Using Knowledge Graphs in Data Mining
 - Societal Impact of Artificial Intelligence
- Room: B6 26, B0.22
- Consultation: Tuesdays 9-10
 - Please make an appointment with Bianca Lermer upfront
- Heiko will teach the lectures



Hello

- M.Sc. Sven Hertling
- Graduate Research Associate
- Research Interests:
 - Semantic Technologies / Semantic Web
 - Linked Data
 - Knowledge Graphs
- eMail: sven@informatik.uni-mannheim.de
- Sven will teach the exercises and co-supervise the projects.



Introduction and Course Outline

- Administration
- Introduction to Knowledge Graphs
- History of Knowledge Graphs
 - Vision of the Semantic Web
 - Building blocks of the Semantic Web
 - Technical foundations

Course Organization

- Lecture
 - Knowledge Graph standards and languages
 - Using public knowledge graphs
 - Creating knowledge graphs
- Exercise
 - Understand knowledge graphs and their principles, play with real data
- Project Work
 - teams of 3-4 students build a Knowledge Graph application
 - teams may choose their own data sets and tasks
(in addition, we will propose some pointers for ideas)
 - write summary about project, present project results
 - **not graded, but mandatory**
- Final exam
 - **final grades are only based on written exam**

Course Organization

- Registration
 - you have registered via Portal2
 - you should have access to ILIAS
 - the course is fully booked with a waiting list
 - if you decide not to attend, please deregister in Portal2
- Important: course replacement
 - This course replaces IE 650 Semantic Web Technologies
 - You cannot get credits for both courses

Course Contents and Schedule

you are here

Week	Tuesday	Friday
04.09.2023	Lecture: Introduction	Exercise: Introduction
11.09.2023	Lecture: Representing Knowledge in Graphs (RDF)	Exercise: Representing Knowledge in Graphs (RDF)
18.09.2023	Lecture: Lightweight Knowledge Graph Inference (RDFS)	Exercise: Lightweight Knowledge Graph Inference (RDFS)
25.09.2023	Lecture: Linked Data, Knowledge Graph Programming	Exercise: Linked Data, Knowledge Graph Programming, Kick Off Group Projects
02.10.2023	Holiday	No exercise, time to work on group projects
09.10.2023	Lecture: Querying Knowledge Graphs	Exercise: Querying Knowledge Graphs
16.10.2023	Lecture: Public Knowledge Graphs	Exercise: Public Knowledge Graphs
23.10.2023	Lecture: Labeled Property Graphs	Exercise: Labeled Property Graphs
30.10.2023	Lecture: Advanced Knowledge Graph Inference (OWL Part 1)	Exercise: Advanced Knowledge Graph Inference (OWL Part 1)
06.11.2023	<i>No lecture (time to work on projects)</i>	<i>No exercise (time to work on projects)</i>
21.11.2022	Lecture: Advanced Knowledge Graph Inference (OWL Part 2)	Exercise: Advanced Knowledge Graph Inference (OWL Part 2)
28.11.2022	Lecture: Knowledge Modeling	Exercise: Knowledge Modeling
05.12.2022	Lecture: Knowledge Graph Quality and Knowledge Integration	Exercise: Knowledge Graph Quality and Knowledge Integration
04.12.2023	Project Presentations	

Deadlines

- Submission of project work proposal
 - Sunday, October 1st, 23:59
- Submission of final project work report
 - Sunday, December 10th, 23:59

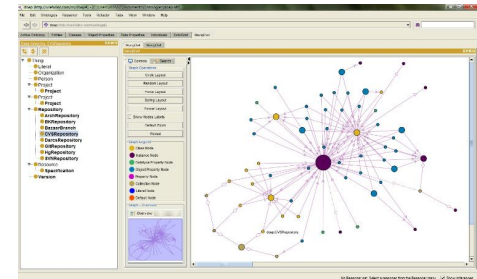


Course Organization

- Lecture Webpage: Slides, Announcements, Web Links
 - hint: look at version tags of slides!
- Additional Material
 - ILIAS eLearning System, <https://ilias.uni-mannheim.de/>
- Time and Location
 - Lecture: Tuesday, 3.30 – 5.00, Room B6 26, room A1.01
 - Exercise: Friday, 12.00 - 13.30, Room B6 26, room A1.04

Further Reading and Software

- Follow the links on the website
 - Most material is available online
- Programming environment
 - JENA framework (Java)
 - RDFlib (Python)
- Knowledge graph environment
 - Neo4j
- Ontology engineering environment
 - Protégé
 - <http://protege.stanford.edu/>



Warning

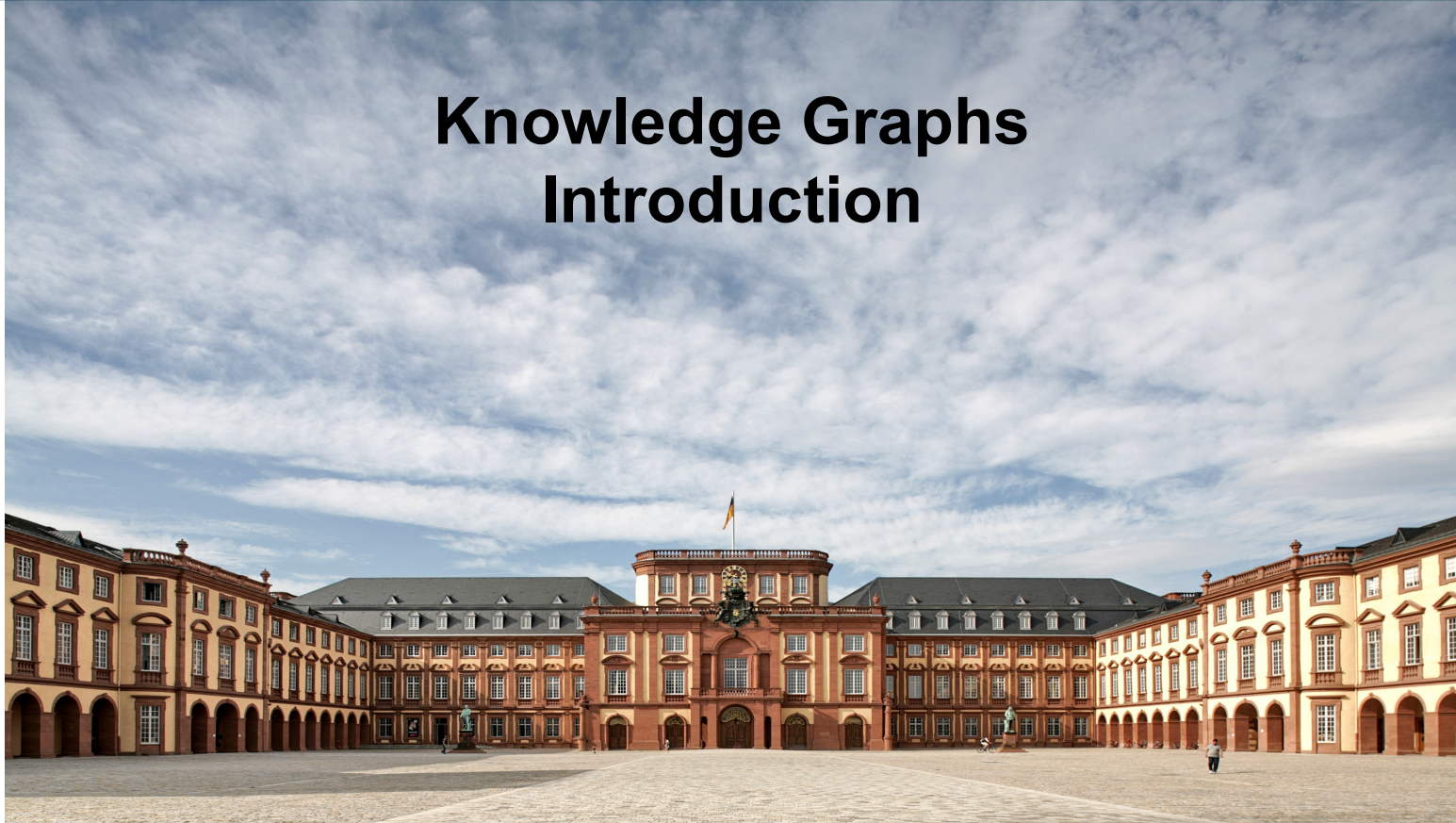
- This lecture contains
 - cartoons
 - Java and Python code
 - some digressions to philosophy
- Having said that:
 - have fun! :-)



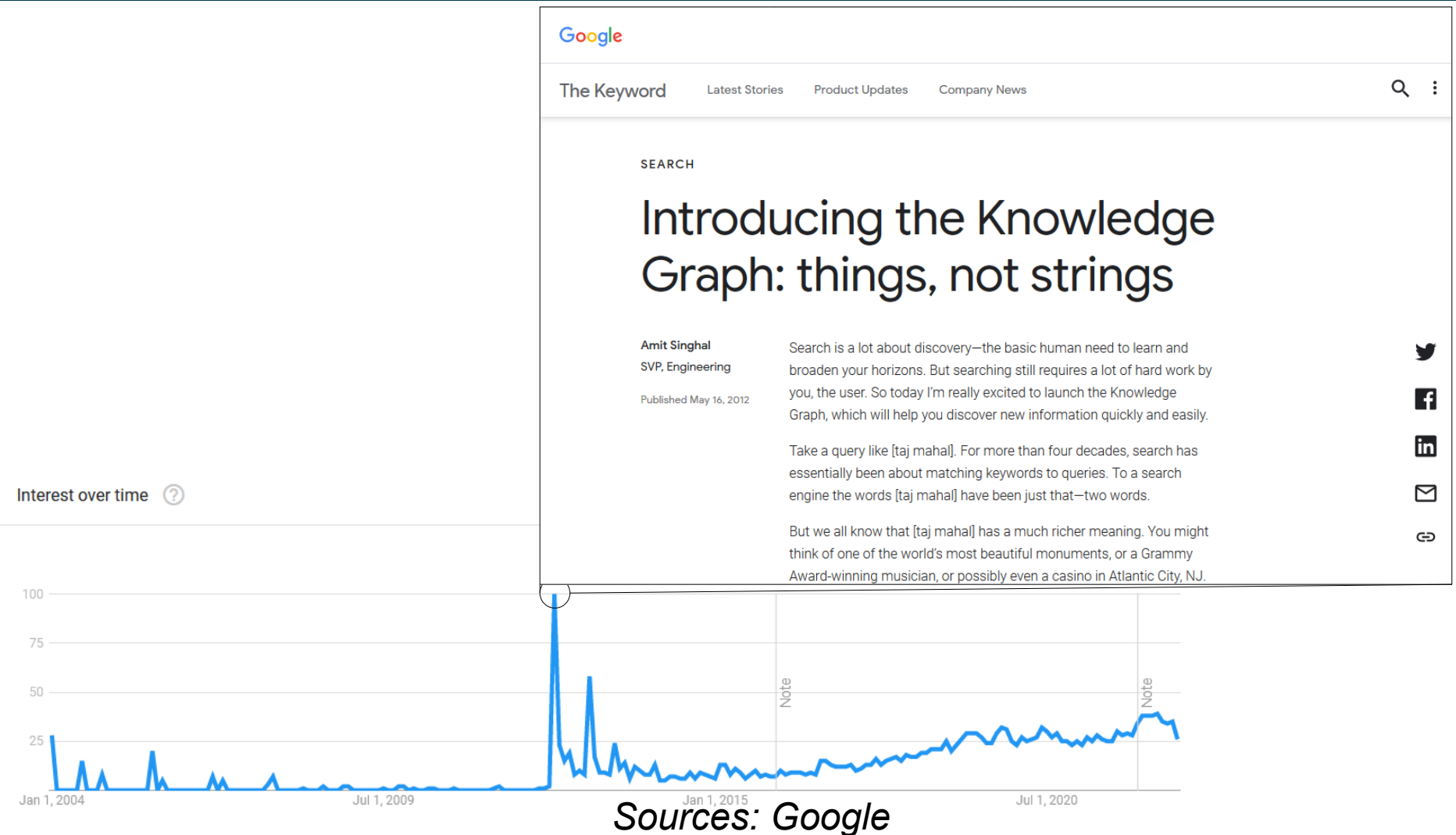
Questions?



Knowledge Graphs Introduction



The Birth of a Buzzword



Idea of the Google Knowledge Graph

- Web search in the pre-knowledge graph age
 - Documents
 - Keywords (not: disambiguated entities)



Hilton Paris Opera, 4-Sterne-Hotel im ...
hiltonhotels.de



Hilton Paris Opera, Paris ...
booking.com



Hilton Paris Opera, 4-Sterne-Hotel im ...
hiltonhotels.de



Paris Hilton – Wikipedia
de.wikipedia.org



Hilton Paris Opera - Fremdenverkehrsamt ...
de.parisinfo.com



Hilton Paris Opera, 4-Sterne-Hotel im ...
hiltonhotels.de



Hilton Paris Opera | Luxury hotel in Paris
hiltonparisoperahotel.com



Hotel Hilton Paris Opera in Paris ab 254 €
agoda.com



► Hotel Hilton Paris Opera, Paris ...
tui.com



Hilton Paris Opera, 4-Sterne-Hotel l...
hiltonhotels.de



Hotel Hilton Paris Opera in Paris ab 254 €
agoda.com



Hilton Paris Opera ab 247 €. Hotels in ...
kayak.de



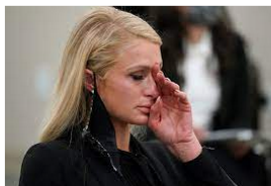
Hilton Paris Opera (1889), Paris ...
historichotelsthenandnow.com



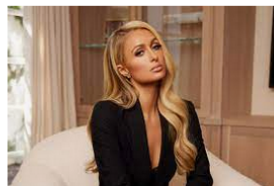
DIE 5 BESTEN Hilton Hotels in Paris ...
tripadvisor.de



Hilton to manage the Concorde Opéra ...
hospitality-on.com



Paris Hilton: »Ich bin der Beweis, dass ...
spiegel.de



Paris Hilton's 2022: Refreshing Her ...
bloomberg.com



Kathy Hilton, Paris Hilt...
alarmy.de



Zitate von Paris Hilton ...
beruhmte-zitate.de



Paris Hilton | Forum - D...
magazin-forum.de



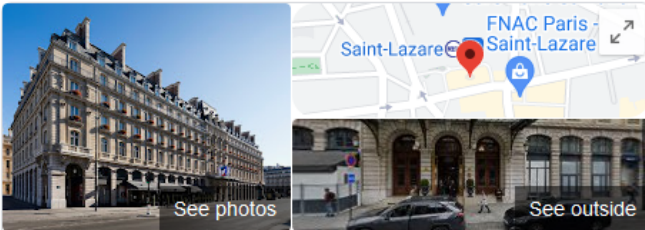
Barron Hilton: Bruder von Paris Hilton ...
gala.de



Hilton Paris Opera Hotel, Paris, Franc...
hotel-board.com

Idea of the Google Knowledge Graph

- Web search in the knowledge graph age
 - Backed by structured information
 - All entities are disambiguated
- Linked to other services
 - Ratings and reviews
 - Map information
 - External services (e.g., booking)
 - ...



The screenshot displays a Google search result for 'Hilton Paris Opera'. At the top, there are two images: a street view of the hotel building and a map showing its location near Saint-Lazare station. Below the images, the hotel name 'Hilton Paris Opera' is prominently displayed. Underneath the name are buttons for 'Website', 'Directions', and 'Save'. The rating is shown as 4.3 stars with 2,280 Google reviews. A blue button labeled 'CHECK AVAILABILITY' is present. The address '108 Rue Saint-Lazare, 75008 Paris, France' and phone number '+33 1 40 08 44 44' are listed. A 'Compare prices' section shows options for Monday, 7 Nov and Tuesday, 8 Nov for 2 people. Below this, a table lists featured booking options from Booking.com, Hotels.com, and Expedia.de, all showing a price of €349.

Hilton Paris Opera

Website Directions Save

4,3 ★★★★★ 2.280 Google reviews ⓘ

4-star hotel

CHECK AVAILABILITY




Address: 108 Rue Saint-Lazare, 75008 Paris, France

Phone: +33 1 40 08 44 44

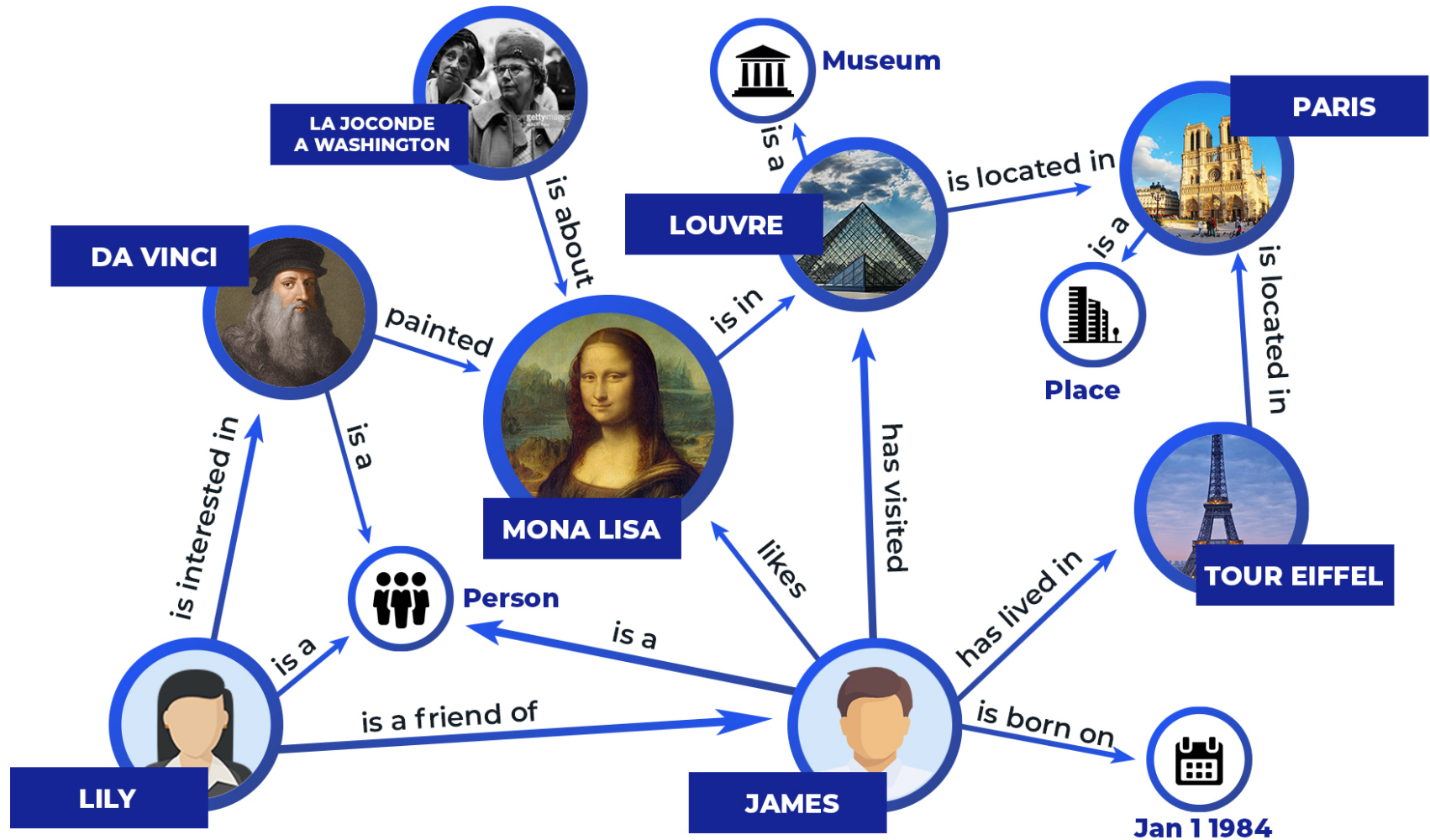
Compare prices

Mon, 7 Nov Tue, 8 Nov 2

Ads · Featured options

 Booking.com	€349 >
 Hotels.com	€349 >
 Expedia.de	€349 >

An Example for a Knowledge Graph



<https://yashueth.wordpress.com/2019/10/08/introduction-question-answering-knowledge-graphs-kgqa/>

From Google to the World

- Documented list of companies using knowledge graphs
 - Courtesy of Frank van Harmelen, VU Amsterdam



From Google to the World

Hype Cycle for Artificial Intelligence, 2023



gartner.com

Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2079794

Gartner.

Knowledge Graph Definitions

- Knowledge Graphs are a fairly new technology
 - Hence, there are few universally acclaimed definitions
- Some example definitions from the literature:
 - *Knowledge graphs could be envisaged as a **network of all kind things** which are **relevant to a specific domain or to an organization**. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets.*
(Blumauer, 2014)
 - *Knowledge graphs are **large networks of entities**, their semantic types, properties, and **relationships between entities**.*
(Kroetsch and Weikum, 2016)

Ehrlinger and Wöb: Towards a Definition of Knowledge Graphs. 2016

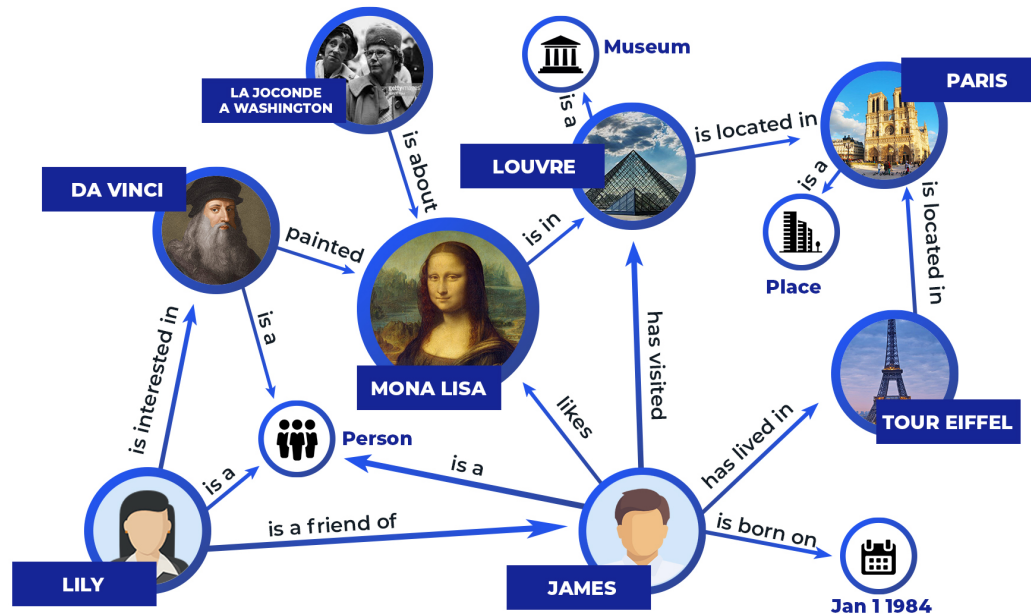
Knowledge Graph Definitions (ctd.)

- *[...] systems exist, [...], which use a variety of techniques to **extract new knowledge, in the form of facts**, from the web. These **facts are interrelated**, and hence, recently this extracted knowledge has been referred to as a knowledge graph.*
(Pujara et al., 2013)
- *A Knowledge Graph (1) **mainly describes instances and their relations in a graph**, (2) defines possible classes and relations in a schema or ontology, (3) allows for **interlinking arbitrary entities with each other**, and (4) covers various domains.*
(Paulheim, 2017)

Ehrlinger and Wöß: Towards a Definition of Knowledge Graphs. 2016

Knowledge Graph Definitions (ctd.)

- Common ground so far:
 - There are entities and relations that are connected and form a graph
 - There is a set of entity and relation *types*
 - those are often referred to as a *schema* or *ontology*
 - we will get back to this



Rewinding the Time Machine

- Google claim “things, not strings”
 - Entities instead of words in documents
 - Relations between entities explicitly modeled
 - Accessible to humans *and machines*
 - think: computers cannot read text



The Vision of the Semantic Web (2001)

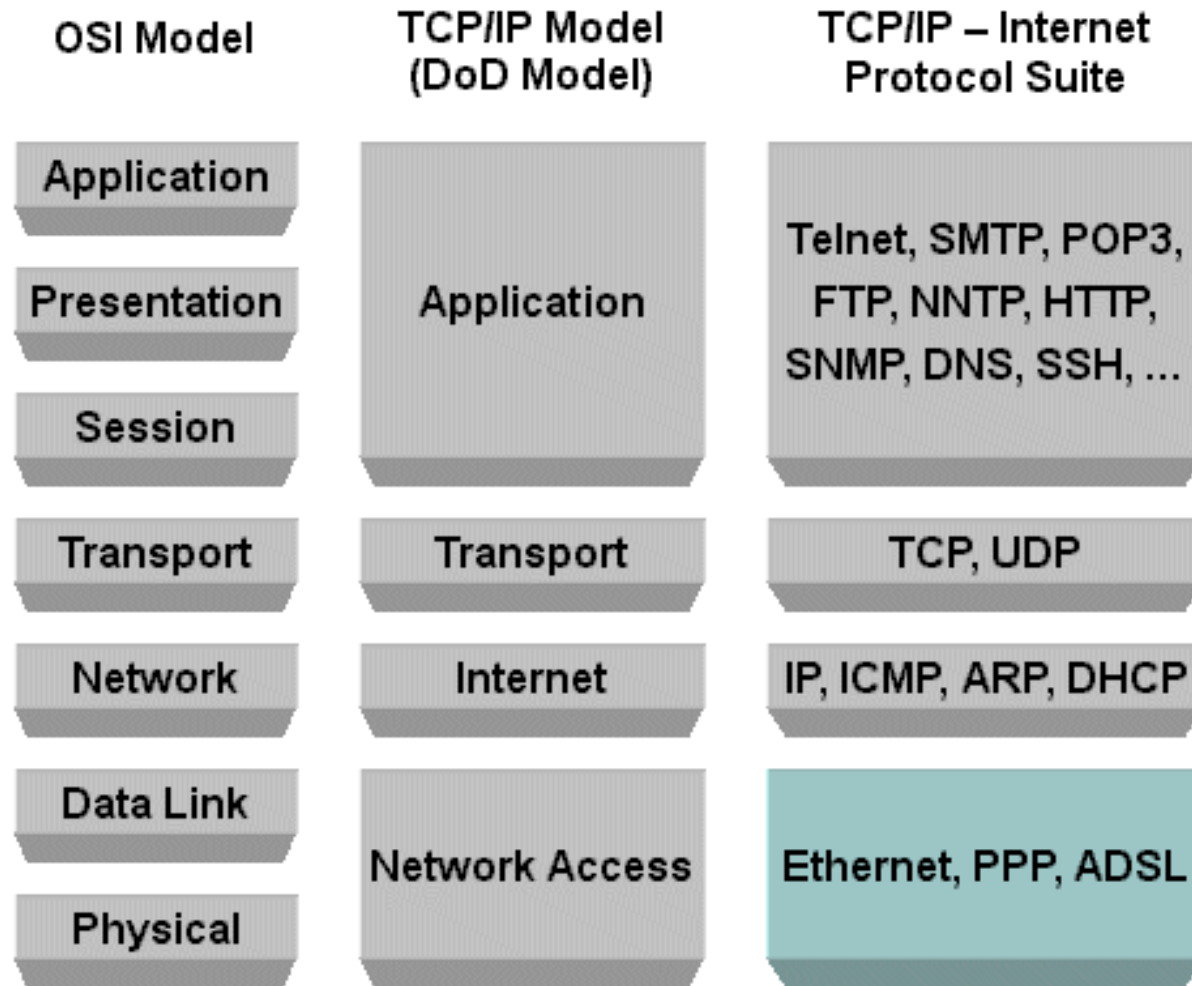
- 2001 article by Tim Berners-Lee, Jim Hendler, and Ora Lassila:

„The Web is the killer app of the Internet.
The Semantic Web is another killer app
of that magnitude.“



Berners-Lee et al. (2001): *The Semantic Web*. In: Scientific American, Mai 2001.

Web vs. Internet?



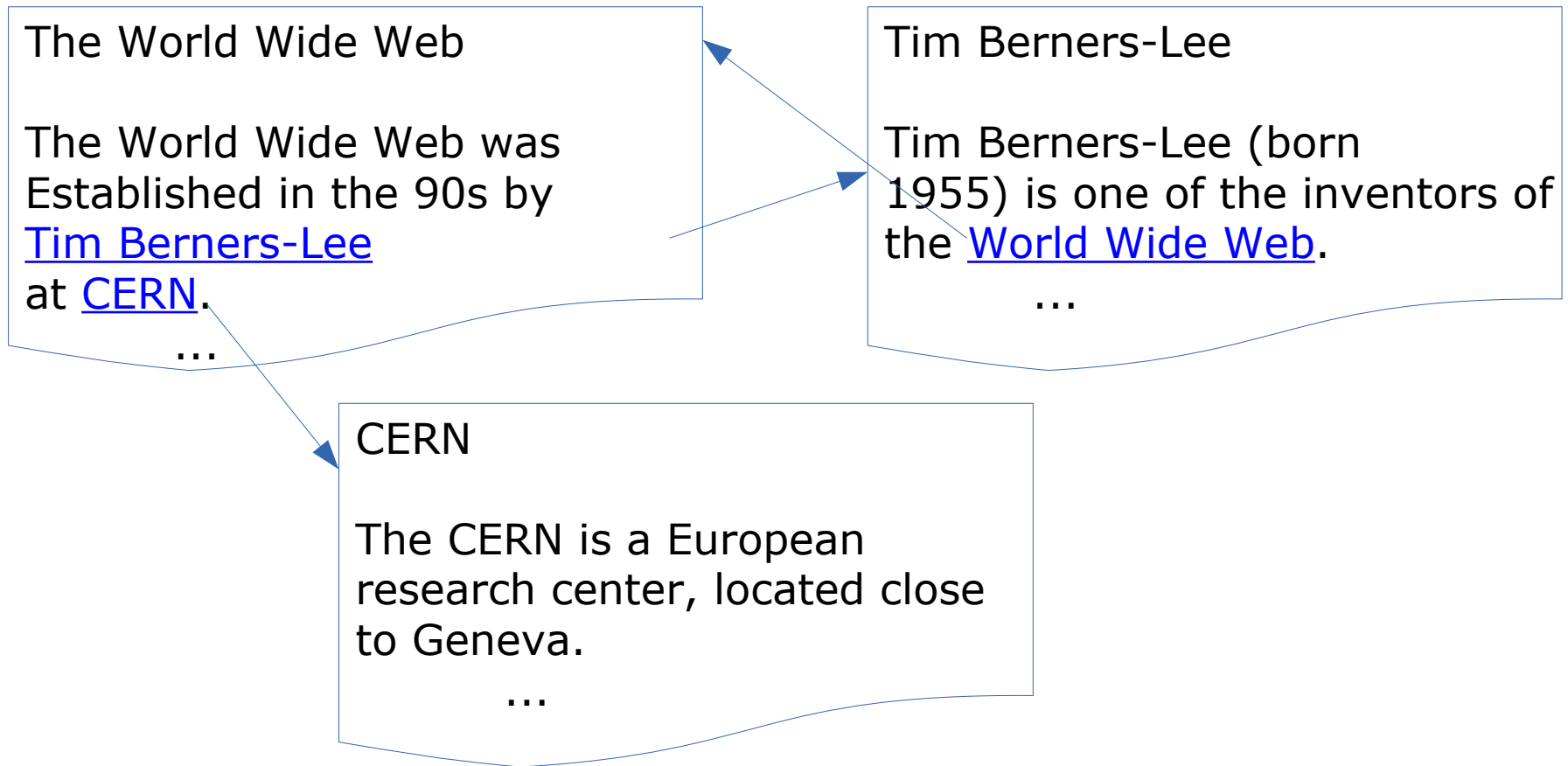
Chin-Shiuh Shieh (2000): *TCP/IP - Internet Protocol Suite and Ethernet*. <http://bit.kuas.edu.tw/~csshie/teach/np/tcpip/index.html>

The “Classic” Web

- a.k.a. “World Wide Web”, “Document Web”
- Uses HTTP protocol and URLs
- HTML as a markup language
 - plus CSS, JavaScript, ...
 - plus a few other, more or less standardized formats (GIF, JPEG, Flash, ...)
- Browser as a universal client
- Information is accessible to humans, but not to machines

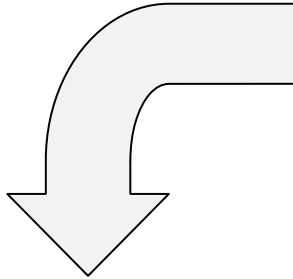
The “Classic” Web

- Hypertext: linked documents



The “Classic” Web

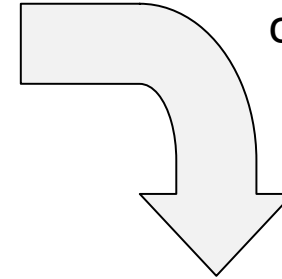
In the eyes of a
human



```
<html>
...
<b>Dr. Mark Smith</b>
<i>Physician</i>
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm
...
</html>
```

Dr. Mark Smith
Physician
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm

in the eyes of a
computer



Print in bold: „hmf298hmmhudsä“
Print in italics: „mj2i9ji0“
Print normal: „fdsah
02hfadsh0um2m0adsmf0ihm
asdfjkköfdsa298ndsfmij32mio
lk2mjpoimjiofdpmsajiomjm“

Searching for Information on the Web

Full text search by keywords (e.g., Google):

- „Mark Smith“
- „Physician in Smalltown“
- „Doctor in Smalltown“
- „Doctor in Smalltown with opening hours on Wednesday afternoon“
- „Somebody in Smalltown who can fix a broken leg“

```
<html>
...
  <b>Dr. Mark Smith</b>
  <i>Physician</i>
  Main St. 14
  Smalltown
  Mon-Fri 9-11 am
  Wed 3-6 pm
  ...
</html>
```

→ “classic” Web is too inflexible for useful search

→ hard to use for intelligent agents

Problems of the “Classic” Web

- Finding information
 - Keyword based search instead of natural language questions
 - Different natural languages
 - Synonyms, homonyms and polysemous words
 - Ambiguity of natural language
- Processing information
 - Formats and encodings
- Making use of information
 - Distributed across pages
 - e.g., a book's author on the publishers site, address on his/her personal page



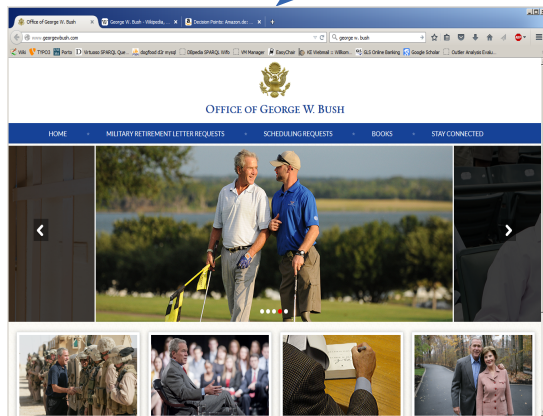
<http://geekandpoke.typepad.com/geekandpoke/2011/08/coders-love-unicode.html>

Homonyms and Polysemous Words



Untyped Links

Bush Era Law Could Get You
20 Years in Prison For
Clearing Your Browser History



Example: Wolfram Alpha (2009)

The screenshot shows the Wolfram Alpha interface with the query "distance from mannheim to karlsruhe" entered in the search bar. Below the search bar, there are links for "Extended Keyboard", "Upload", "Examples", and a settings icon. A blue box highlights the input interpretation section, which shows the system's assumptions: "Assuming 'mannheim' is a city | Use as an airport instead", "Assuming 'karlsruhe' is a city | Use as a ship instead", and "Assuming Karlsruhe (Germany) | Use Karlsruhe (United States) instead". Below this, the "Input interpretation:" section shows a table with "distance" as the unit, "from" as Mannheim, Baden-Wuerttemberg, and "to" as Karlsruhe, Baden-Wuerttemberg. The "Result:" section shows "55.84 km (kilometers)". The "Unit conversions:" section lists "34.7 miles", "55.84 km (kilometers)", "55840 meters", "5.584 x 10^6 cm (centimeters)", and "30.15 nmi (nautical miles)". The "Direct travel times:" section shows a table with travel times for car, sound, light in fiber, and light in vacuum. A blue callout points to the input interpretation section, and another blue callout points to the direct travel times table.

WolframAlpha[®] computational intelligence.

distance from mannheim to karlsruhe

Extended Keyboard Upload Examples

Assuming "mannheim" is a city | Use as an airport instead
Assuming "karlsruhe" is a city | Use as a ship instead
Assuming Karlsruhe (Germany) | Use Karlsruhe (United States) instead

Input interpretation:

distance	from	Mannheim, Baden-Wuerttemberg
	to	Karlsruhe, Baden-Wuerttemberg

Open code

Result:

55.84 km (kilometers)

Unit conversions:

34.7 miles

55.84 km (kilometers)

55840 meters

5.584×10^6 cm (centimeters)

30.15 nmi (nautical miles)

Direct travel times: More

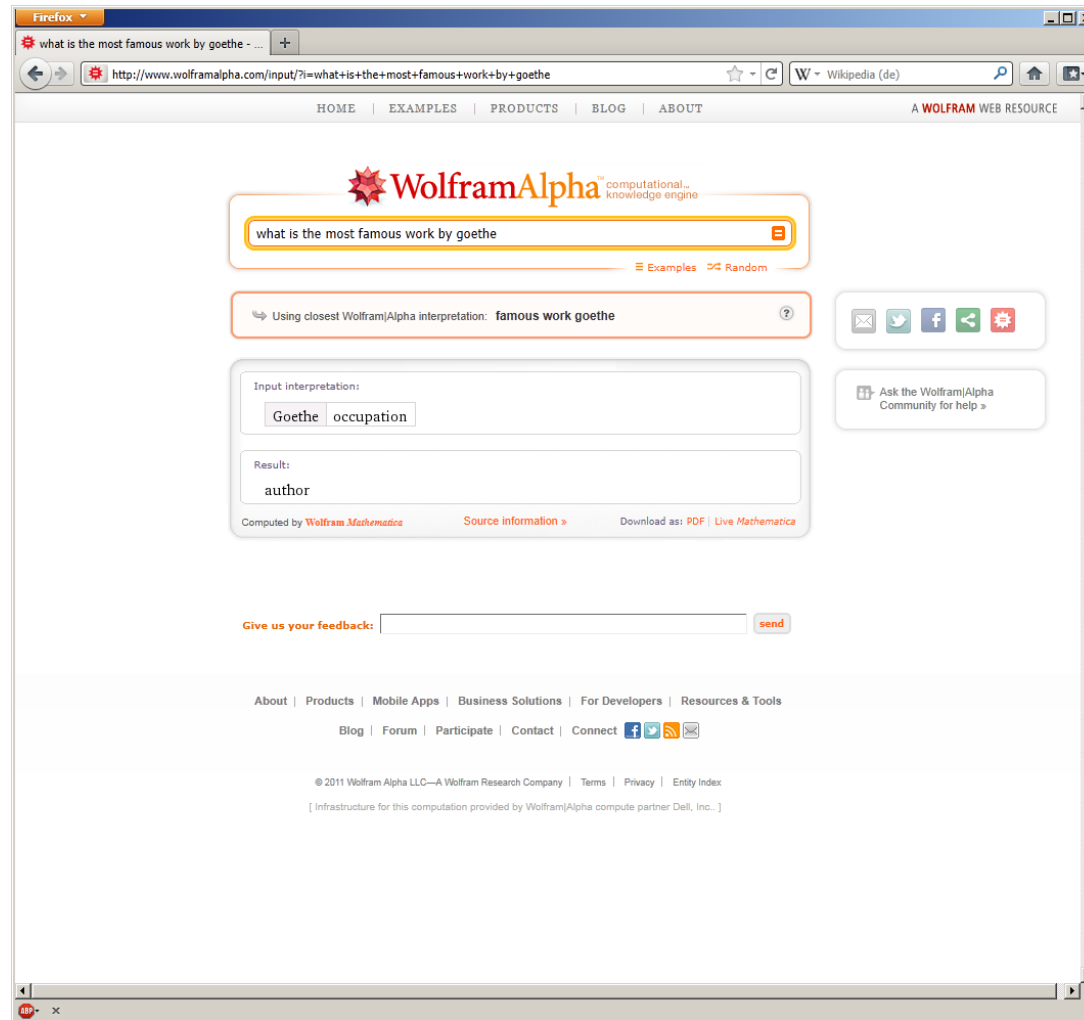
car (55 mph)	38 minutes
sound	2 minutes 44 seconds
light in fiber	261 μ s (microseconds)
light in vacuum	186 μ s (microseconds)

(straight-line path)

Multiple interpretations of "Mannheim" and "Karlsruhe"

Multiple interpretations of "distance"

Example: Wolfram Alpha (2009)



Example: ChatGPT (2022)



What is the distance from Mannheim to Karlsruhe?



The distance between Mannheim and Karlsruhe can vary depending on the specific route you take, but they are relatively close cities in the state of Baden-Württemberg, Germany. The approximate driving distance between Mannheim and Karlsruhe is around 70 to 80 kilometers (43 to 50 miles) depending on the exact starting and ending points and the route you choose to take. The travel time by car typically ranges from 45 minutes to 1 hour, depending on traffic conditions. Additionally, there are train connections available between Mannheim and Karlsruhe, which can be a convenient way to travel between these two cities.

Example: ChatGPT (2022)

HE

What is the most famous work by Goethe?



Johann Wolfgang von Goethe, one of Germany's most famous literary figures, wrote numerous works that have had a significant impact on literature and culture. His most famous work is arguably "Faust." "Faust" is a tragic play in two parts, with "Faust, Part One" being particularly well-known. It tells the story of a scholar named Heinrich Faust who makes a pact with the devil, Mephistopheles, in exchange for worldly pleasures and knowledge. The work is considered a masterpiece of German literature and is often studied and adapted in various forms. Goethe's "Faust" is a classic of world literature and has had a profound influence on literature, philosophy, and the arts.

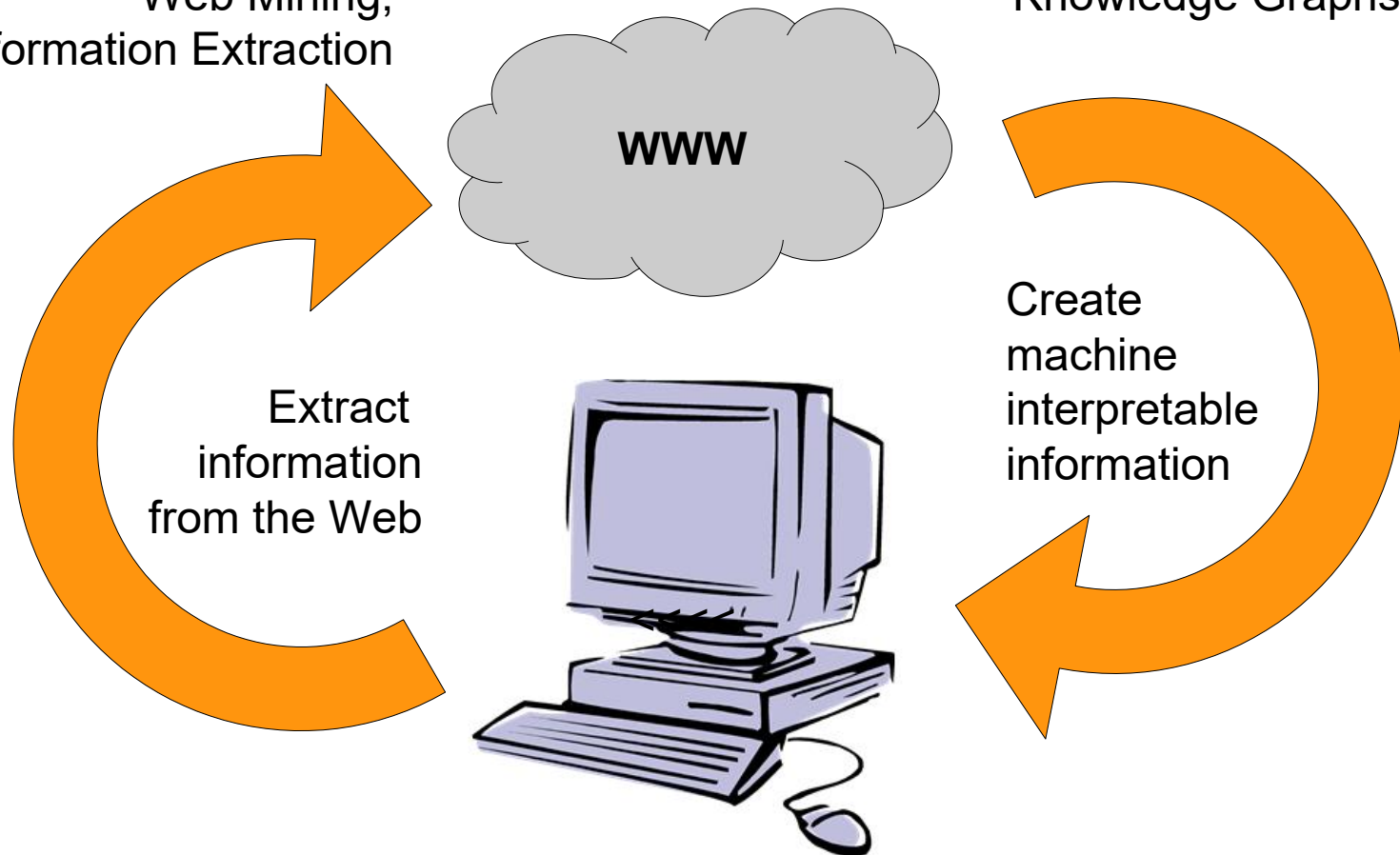
Observations

- Wolfram Alpha
 - Structured knowledge
 - Explicit inference rules
 - Good at precise factual questions
 - Not good at common knowledge
- ChatGPT
 - Unstructured knowledge
 - Heuristic inference
 - Good at common knowledge
 - Not good at precise factual questions

Solutions

Lectures:
Web Mining,
Information Extraction

Lecture:
Knowledge Graphs



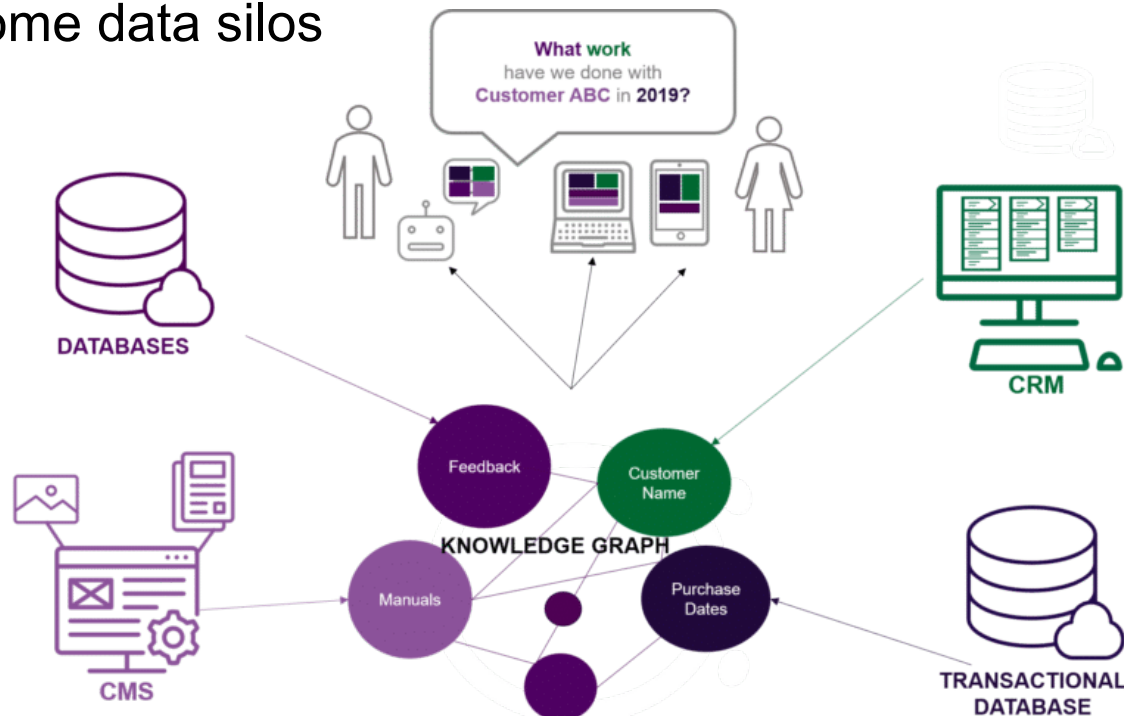
Semantic Web Vision

- Provide information in machine interpretable form
- Make (semantic) links between (data) documents us
- Facilitate useful (!) complex queries
- Allow logical reasoning



(Enterprise) Knowledge Graph Vision

- Integrate data from different sources
- Make connections between entities in those sources
- Facilitate cross data source queries
- Overcome data silos



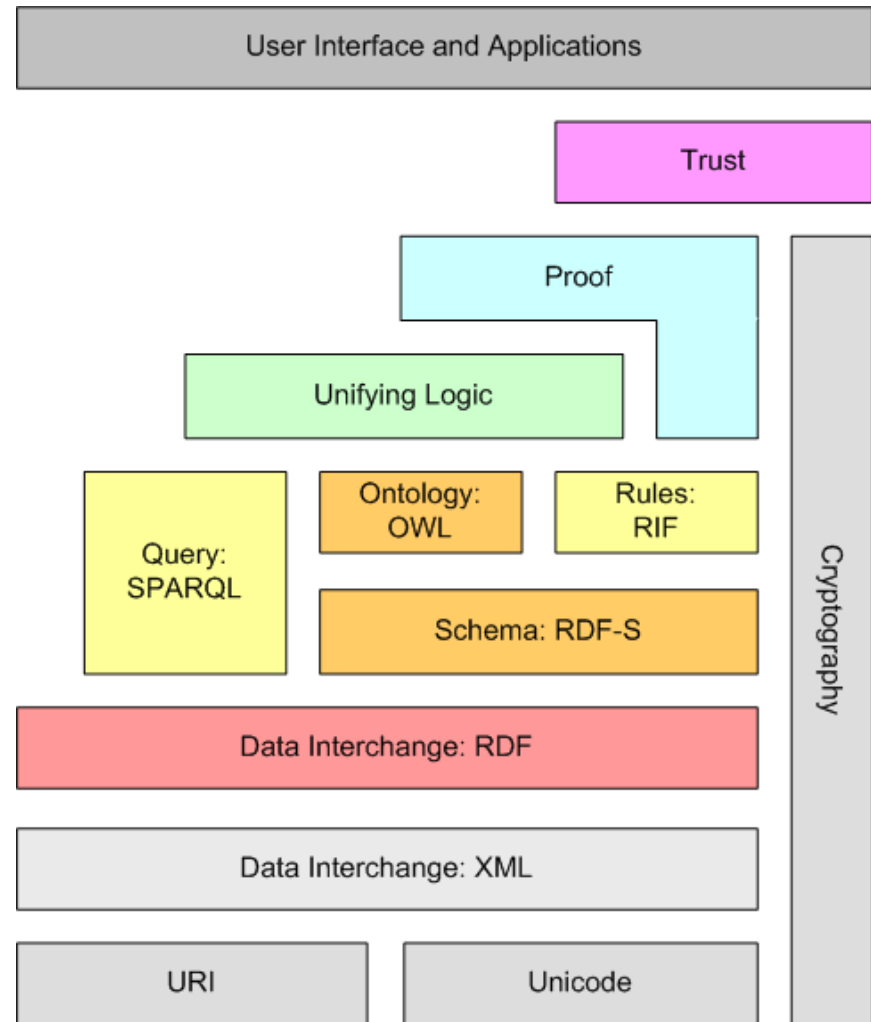
Semantic Web – Architecture



here be dragons...

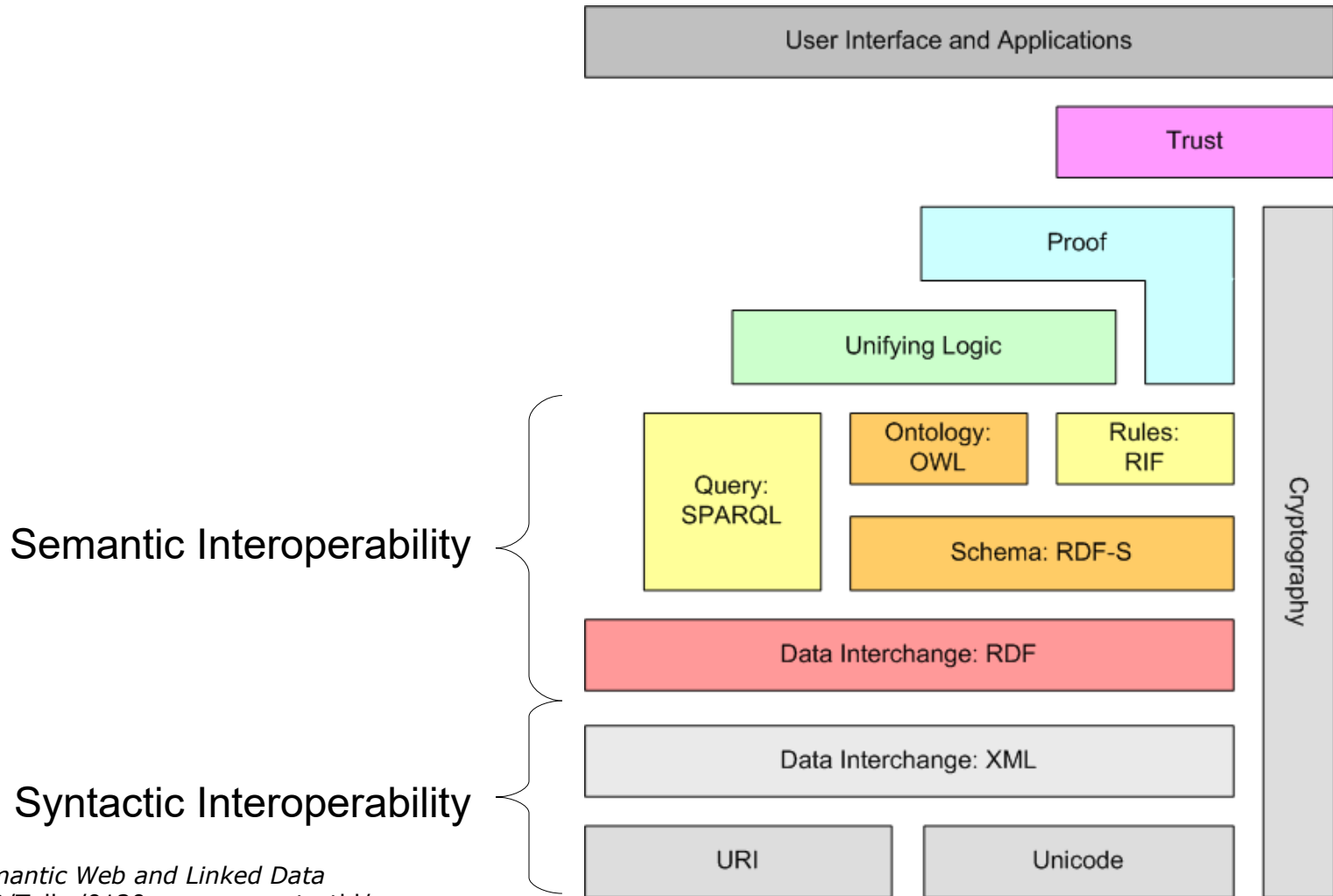
Knowledge Graph
Technologies
(This lecture)

Technical
Foundations



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

Data Interoperability with Knowledge Graphs



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

Syntactic Interoperability: Character Sets

- ASCII („American Standard Code for Information Interchange“) ISO 646 (1963), 127 characters, 95 of which are printable:

!"#\$%&'()*+,-./0123456789:;<=>?
@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_
`abcdefghijklmnopqrstuvwxyz{|}~

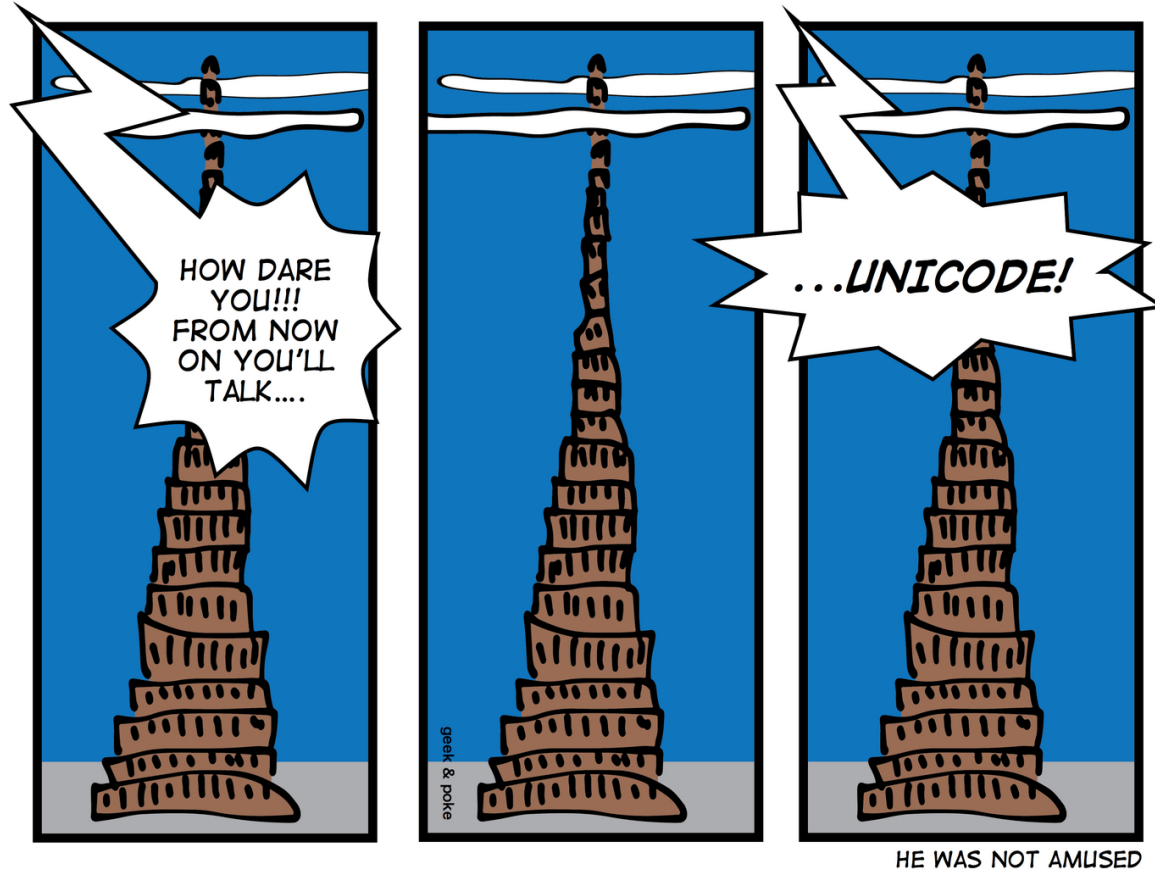
- Extension to 8 Bit: ISO 8859-1 to -16 (1998)
 - covers major European languages
 - most well known: 8859-1 („Latin-1“)
- The Web, however, speaks many more languages...

وللحب علامات يتفقوها الف
فأولها إدمان النظر، والعج
سراؤها، والمعبرة لضمائرها
بر لا يطرف، يتنقل بتنقل
ن مال، كالحرباء مع الشمس

我爱中国
国中爱我

Syntactic Interoperability: Multilinguality

TOWER OF BABEL



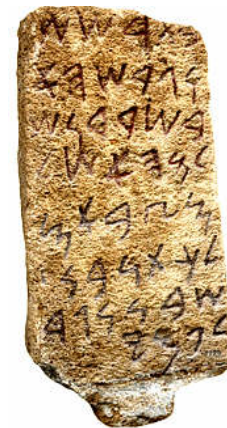
<http://geek-and-poke.com/geekandpoke/2013/8/29/when-it-all-began>

Syntactic Interoperability: Unicode

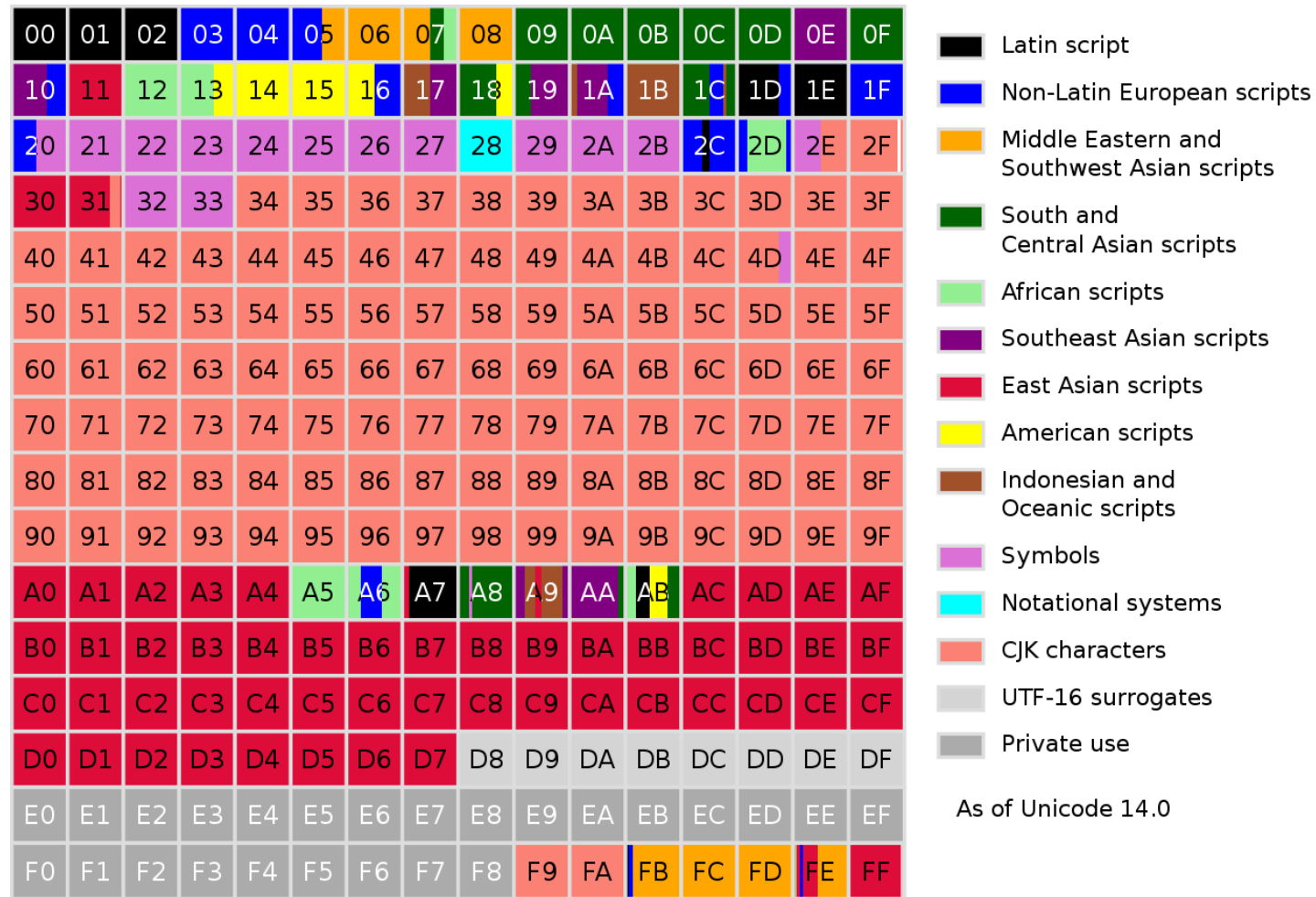
- ISO 10646
 - first version 1991 (Europe, Near East, India)
 - Unicode 14.0 (September 2021)
 - defines ~144,000 characters
 - covers even very exotic languages
 - Plus: currency symbols, emojis, sign language, music notation...



Klingon
is still
missing!!!



Syntactic Interoperability: Unicode



Source: Wikimedia Commons

Information Representation in XML

XML (eXtensible Markup Language)

- A W3C standard since 1998
- Universal format for data exchange and integration



```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

XML: Basic Concepts

- Tags (arbitrarily definable):
 - Form pairs:
`<physician> ... </physician>`
 - ...or empty element tags
`<young />`
- Attributes:
`<physician location="Smalltown">`
- Tags are nested (with *exactly one* root element):
`<physician>`
 `<address> ... </address>`
`</physician>`

XML: Well-formed Documents

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

The diagram highlights several errors in the XML document on the right using colored annotations:

- Green circles:** Surround the opening `<address>` tag and the closing `</address>` tag, which are not properly nested with the `<telephone>` tag.
- Blue circles:** Surround the opening `<telephone>` tag and the closing `</telephone>` tag, which are not properly nested with the `<address>` tag.
- Red circles:** Surround the `<monday>` element and its content `9-11 am`, which is not properly closed with `</monday>`.

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  <telephone>
    <number>+44 123 456789</number>
  </address>
</telephone>
  <hours>
    <monday>9-11 am
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

HTML and XML

- HTML documents look like XML documents
 - ...but they are usually not well-formed!

```
<p>Look at this!<img src=smiley.gif> <br>
```

- XHTML: HTML as well-formed XML documents
- A W3C standard since 2000

```
<p>Look at this! <br/> </p>
```



XPath: Accessing Information in XML

- Query language for XML
- A W3C standard since 1999 (Version 2.0: 2010)

`/physician[name='Dr. Mark Smith']/telephone/number`

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

Namespaces in XML

- Elements with the same name can occur in different places
 - ...but the contents and semantics may differ
- How can we tell them apart?

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

Namespaces in XML

- Namespace definition using prefixes (Notation: `prefix:name`)
- Each namespace itself is a URI
- Default namespaces may be defined

```
<physician xmlns      ="http://www.med.com/physician"
           xmlns:addr="http://www.med.com/addr">
  <name>Dr. Mark Smith</name>
  <addr:address>
    <addr:street>Main St.</addr:street>
    <addr:number>14</addr:number>
    <addr:city>Smalltown</addr:city>
  </addr:address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```


XML: Document Type Definition (DTD)

- Defines valid elements for a class of XML documents
 - Names
 - allowed attributes
 - allowed nested child elements
- DTD is a part of the W3C's XML specification
- XML documents matching a DTD are called “valid”

XML: Document Type Definition (DTD)

```
<!DOCTYPE physician [  
  <!ELEMENT physician (  
    name,  
    address*,  
    telephone?,  
    fax?,  
    hours)>  
  
  <!ELEMENT address (  
    street,  
    number,  
    city)>  
  
  <!ELEMENT street (#PCDATA)>  
  
    ...  
>]
```

```
<!DOCTYPE physician SYSTEM  
  "physician.dtd">  
  
<physician>  
  <name>Dr. Mark Smith</name>  
  <address>  
    <street>Main St.</street>  
    <number>14</number>  
    <city>Smalltown</city>  
  </address>  
  <telephone>  
    <number>+44 123 456789</number>  
  </telephone>  
  <hours>  
    <monday>9-11 am</monday>  
    <tuesday>9-11 am</tuesday>  
    ...  
  </hours>  
</physician>
```



XML: Document Type Definition (DTD)

- Definition of child elements and their order

```
<!ELEMENT address (street,no,line*,zip,city,state?)>
```

- ?, + and * mark optional and possible multiple elements

- Definition of attribute lists

```
<!ATTLIST person title CDATA>
```

- Allowed modifiers: #REQUIRED, #FIXED, #IMPLIED, “...”
- Enumerating allowed values: (dr|prof)

- Definition of entities:

```
<!ENTITY sw “Semantic Web”>
```

- May be used as shortcuts in the XML document: &sw;

XML Schema

- W3C-Standard (since 2004)
- XML schemas are XML files themselves
- More flexible than DTDs:
 - Minimum and maximum number of elements
 - Combinations of elements (either/or, combinations w/out fixed order, ...)
 - Data types (Numbers, dates, ...), own types may be defined
 - Support for namespaces
 - Possibility to create modular schemas

XML Schema

```
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
  <xs:element name="physician">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name"
          type="xs:string">
        <xs:element name="address">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="street"
                type="xs:string">
              ...
            </xs:sequence>
          </xs:complexType>
        </xs:element>
        ...
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```
<physician xmlns:xsi=
  "http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation=
    "physician.xsd">
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

XML Schema – Modular Schemas

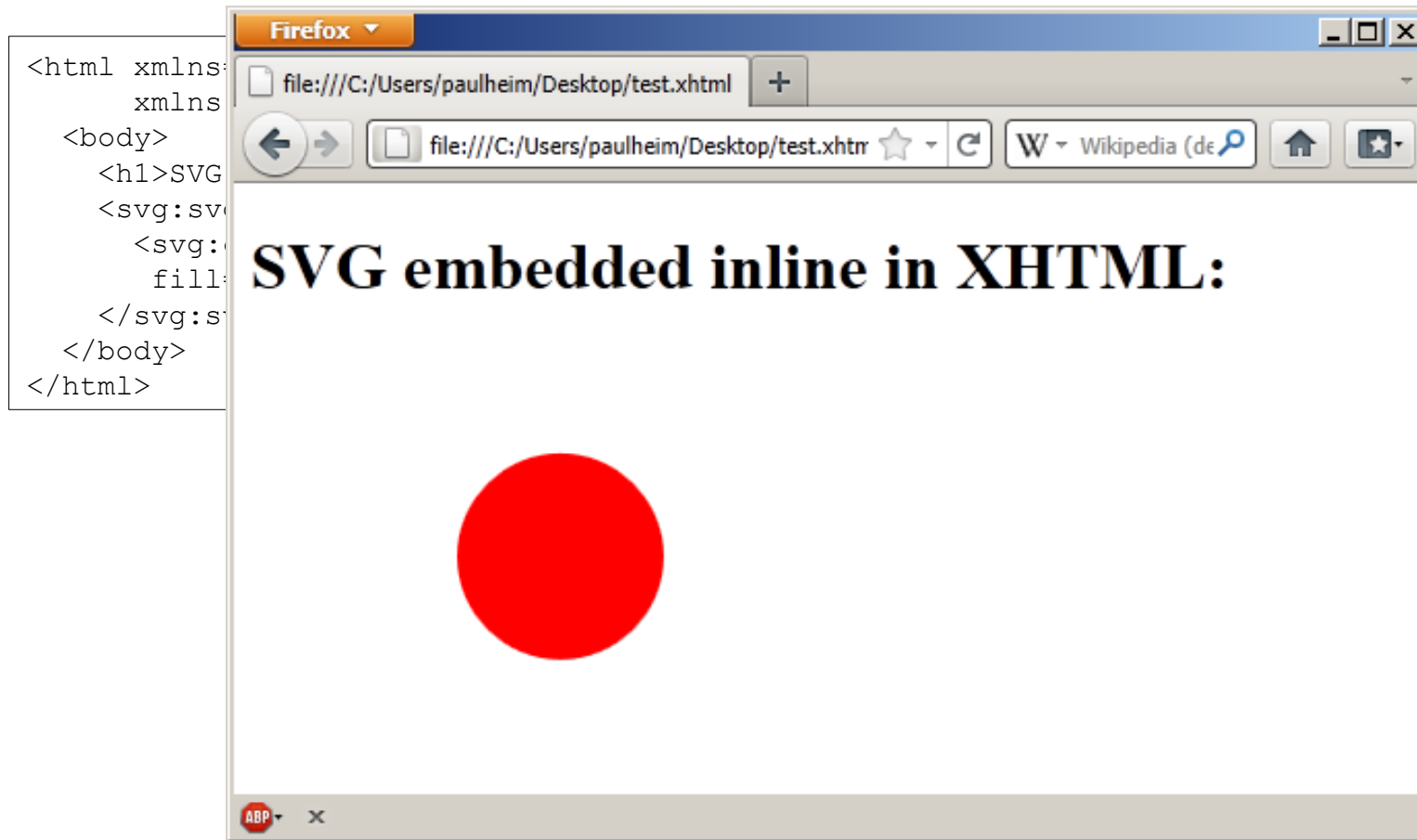
```
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:addr="http://www.address.com/">
```

```
  <xs:import
    namespace="http://www.address.com/"
    schemaLocation="address.xsd"/>
  <xs:element name="physician">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name"
          type="xs:string">
          <xs:element ref="addr:address" />
          ...
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:schema>
```

```
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
  <xs:element name="address">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="street"
          type="xs:string">
          ...
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:schema>
```

Example: Modular Schemas in XHTML



https://developer.mozilla.org/En/SVG:Namespaces_Crash_Course

So, what does a DTD/Schema Define?

- Syntax – σύνταξις („together“ + „order“)
 - Which elements are there?
 - How are they arranged?
 - Which combinations are allowed?
- ...as opposed to: Semantics - σημαίνει („denote“)
 - How to interpret the contents of an element?
 - What is their relation?

Syntax and Semantics: The Linguists' View


- Syntax: how are correct sentences formed?
 - „This sentence no verb.“
 - „The dreaming lamp ~~give~~ gives a freshly cut ~~juices~~ juice to the ~~tire~~ tired sink.
- Semantics: what does a word and sentence *mean*?
- Notes
 - syntactic correctness does not guarantee semantic interpretability
 - semantic interpretability does not require syntactic correctness (for humans)



Syntax and Semantics: The Linguists' View


syntax


semantics

Definition of **knowledge** noun from the Oxford Advanced Learner's Dictionary

 **knowledge** noun


BrE /'nɒlɪdʒ/ ; NAmE /'nɑ:lɪdʒ/ 

 Add to my wordlist

1  [uncountable, singular] the information, understanding and skills that you gain through education or experience

- *practical/medical/scientific knowledge*
- **knowledge of/about something** *He has a wide knowledge of painting and music.*
- *There is a lack of knowledge about the tax system.*

→ SEE RELATED ENTRIES: **Teaching and learning**

2  [uncountable] the state of knowing about a particular fact or situation

- *She sent the letter **without my knowledge**.*
- *The film was made with the Prince's **full knowledge** and approval.*
- *She was impatient **in the knowledge** that time was limited.*
- *I went to sleep **secure in the knowledge** that I was not alone in the house.*
- *They could relax **safe in the knowledge** that they had the funding for the project.*
- *He **denied all knowledge of** the affair.*

3 **knowledge economy/industry/worker** working with information rather than producing goods

- *the emergence of consultancy as a knowledge industry*
- *the shift toward a knowledge economy*

So, what does a DTD/Schema Define?

Employee catalog
of the hospital

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

(probably)
the private address

Yellow Pages

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

(probably)
the work address

?
=

So, what does a DTD/Schema Define?

- XML Schema / DTD defines the *syntax* of an XML document, but not its *semantics*
- Tag names are not interpretable by machines
 - i.e., they do not ease the information retrieval process...
 - Semantics of the data is hidden – usually hard wired in the application
- The Semantic Web is meant as a remedy to that problem
 - *Semantic Web is/can do more than XML!*

```
<2nf3oiü*>
  <34f0>Dr. Mark Smith</34f0>
  <rm4935r>
    <e2m4>Main St.</e2m4>
    <dur3>14</dur3>
    <jfa34>Smalltown</jfa34>
  </rm4935r>
  <d24r3fmö>
    <deß5>+44 123 456789</deß5>
  </d24r3fmö>
  <vsfif>
    <f02>9-11 am</f02>
    <fj9>9-11 am</fj9>
    ...
  </vsfif>
</2nf3oiü*>
```


Uniform Resource Identifiers (URIs)

- “Things, not strings” requires identifiers for *things*
 - URIs: Proposed by Tim-Berners-Lee as „Universal Resource Identifier“ (IETF RFC 1630)
 - Standardized: IETF RFC 3986 (2005)
- Used for naming and finding resources on the Web

URI = scheme ":" hier-part ["?" query] ["#" fragment]

The diagram illustrates the components of a URI. The text 'URI = scheme ":" hier-part ["?" query] ["#" fragment]' is shown. Below 'scheme' is a bracket pointing to 'http' in the example. Below 'hier-part' is a bracket pointing to 'example.com:8042/over/there'. Below 'authority' is a bracket pointing to 'example.com:8042'. Below 'path' is a bracket pointing to '/over/there'. Below 'query' is a bracket pointing to '?name=ferret'. Below 'fragment' is a bracket pointing to '#nose'.

http://example.com:8042/over/there?name=ferret#nose

URIs vs. URLs

- Uniform Resource Locators (IETF RFC 1738, 1994) are a *subset* of URIs
- URIs can refer to *arbitrary* things
- A URL refers to a resource on the Web
- Typical URL prefixes
 - http
 - ftp
 - mailto
 - telnet
 - file
 - ...

URLs on the Web

- Most common usage:
Hyperlinks in HTML documents
- Links usually do not carry
any meta information

Tim Berners-Lee

Tim Berners-Lee (born 1955) is one of the inventors of the [World Wide Web](#).

...

<http://www.w3.org/WWW/>

The World Wide Web

The World Wide Web was initiated in the 90s by [Tim Berners-Lee](#) at [CERN](#).

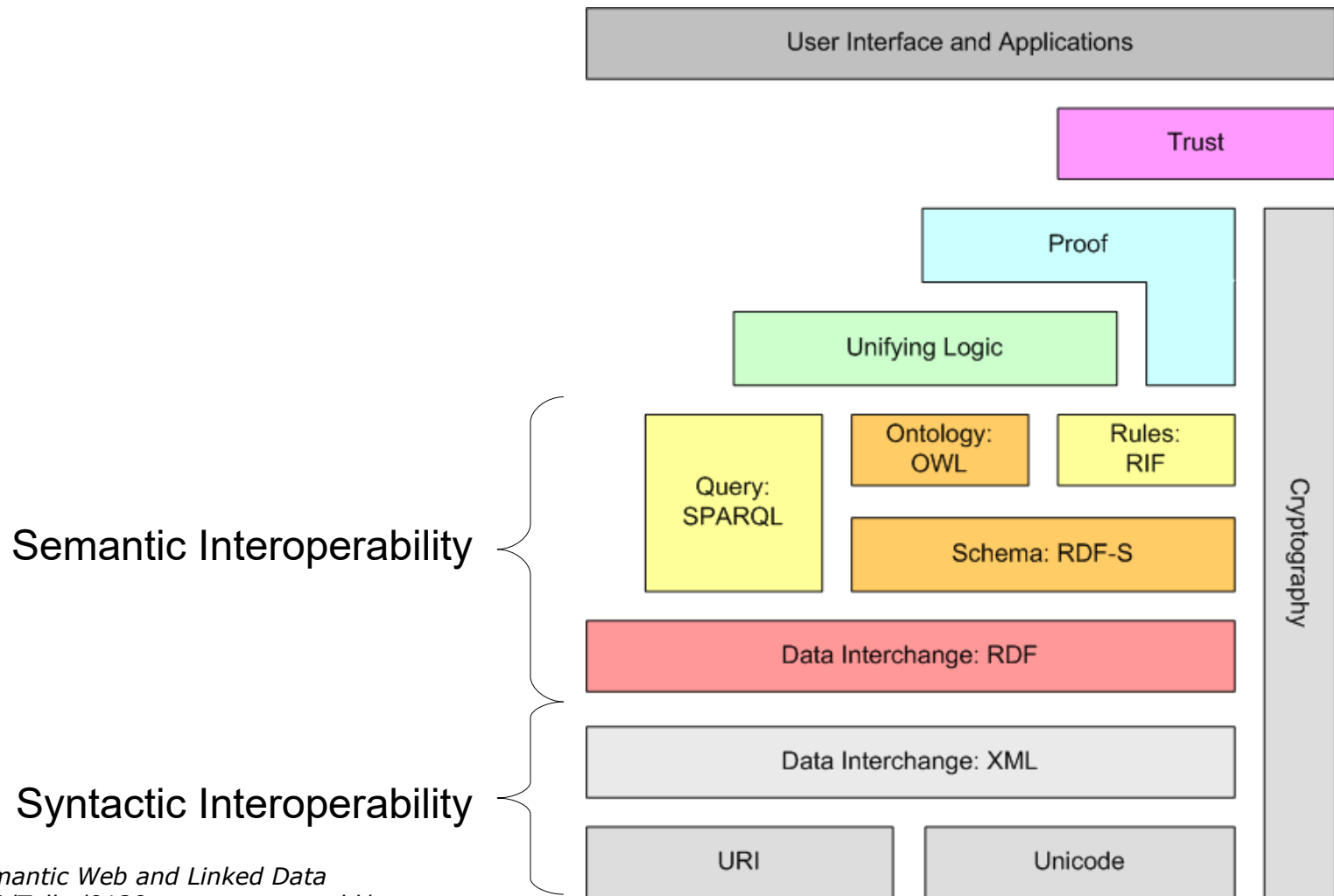
...

<http://www.w3.org/People/Berners-Lee/>

Wrap Up

- Knowledge Graphs
 - Facilitate syntactic and semantic data interoperability
- Today, we have seen syntactic interoperability
 - Unicode: a character set for all languages
 - XML: a universal data exchange format
 - XPath
 - DTD
 - XML Schema
 - URIs
 - Unique identifiers for *things* (entities, resources, ...)
 - On the Web, URLs are dereferencable

Data Interoperability with Knowledge Graphs



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

Questions?

