



## **Open Data on the Web as the fuel for Cognitive Computing**Das Potenzial von Open Data für Cognitive Computing

**Axel Polleres** 

web: http://polleres.net twitter: @AxelPolleres





23.3.2017 16:30 bis 23.03.2017 18:30

## KICK-OFF: DIGITAL INFORMATION MANAGEMENT COMMUNITY AUSTRIA

Datum 23.3.2017 16:30 - 23.03.2017 18:30

Ort Impact Hub Vienna

künstliche Intelligenz virtuelle Assistenten werden im privaten und geschäftlichen Alltag immer präsenter. Digitale Lösungen nutzen Cognitive Computing ifbereitete Informationen und Prozessautomatisierung. Disruptive digitale Businessmodelle revolutionieren ganze Branchen.

Mit welchen Themen sollten wir uns befassen?

Mit welchen Themen sollten wir uns befassen?

Mit welchen Themen sollten wir uns befassen?

Welche Lösungen sind im Einsatz?

bereits im Einsatz: In uns dur

spannende Diskussionen sowie einen projektorientierten Erfahrungsaustausch

zwischen Experten aus dem Public Sector und der Privatwirtschaft.

## 1. What is Cognitive Computing?

### Cognitive computing

From Wikipedia, the free encyclopedia

Cognitive computing (CC) describes technology platforms that, broadly speaking, are based on the scientific disciplines of artificial intelligence and signal processing. These platforms encompass machine learning, reasoning, natural language processing, speech and vision, human-computer interaction, dialog and narrative generation, among other technologies.[1][2]

#### Contents [hide]

- 1 Definition
- 2 Use cases
- 3 See also
- 4 References
- 5 Further reading

### Definition [edit]

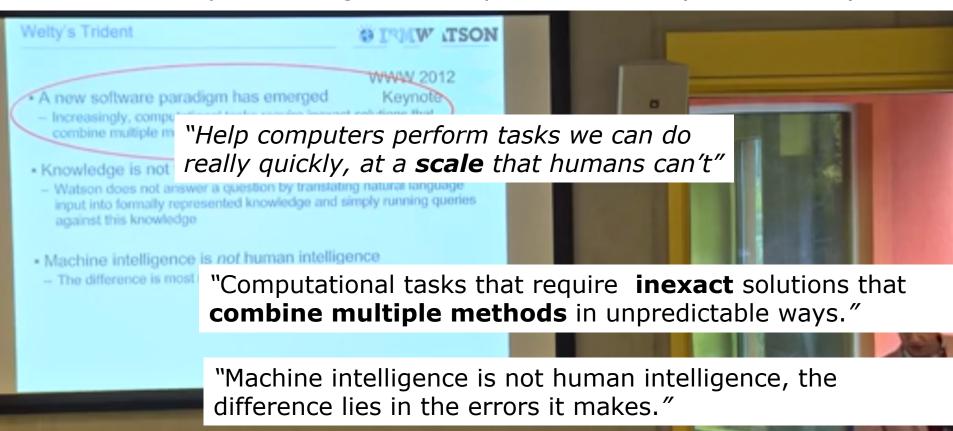
At present, there is no widely agreed upon definition for cognitive computing in either academia or industry. [3][4]



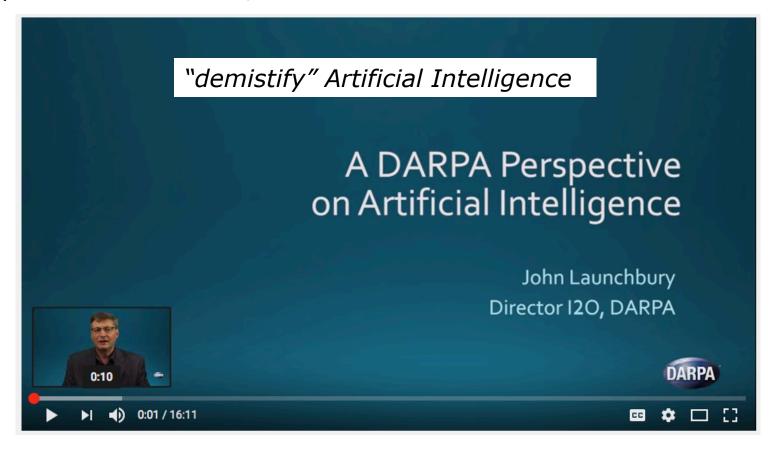
#### WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND RUSINESS

## 1. What is Cognitive Computing?

Dr. Chris Welty, now Google, formerly IBM Research (Watson team)



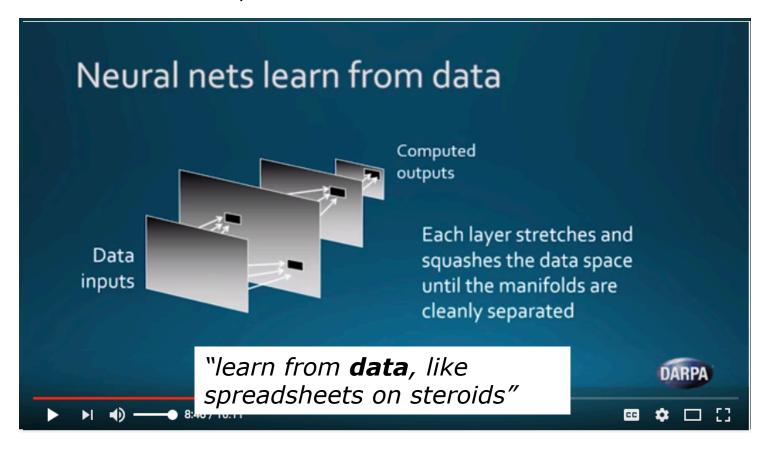
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BIJSINESS



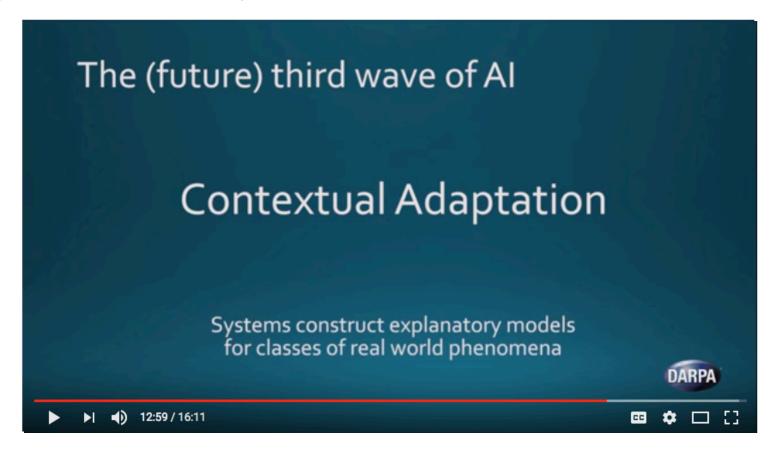
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS



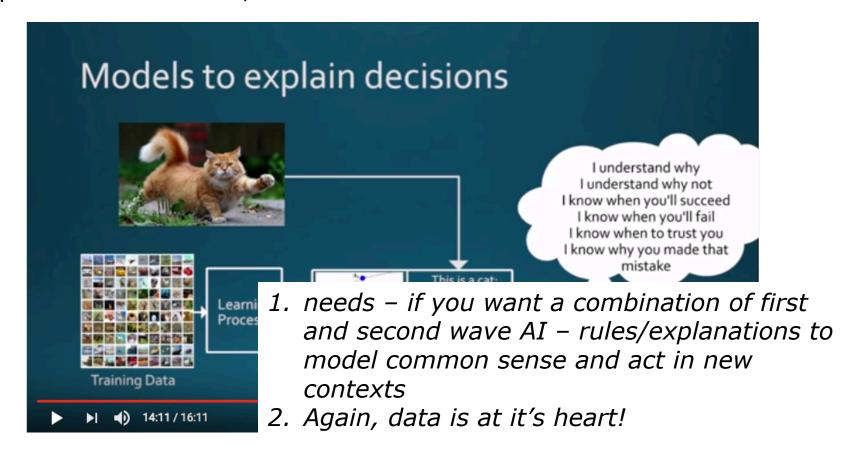
WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS



WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



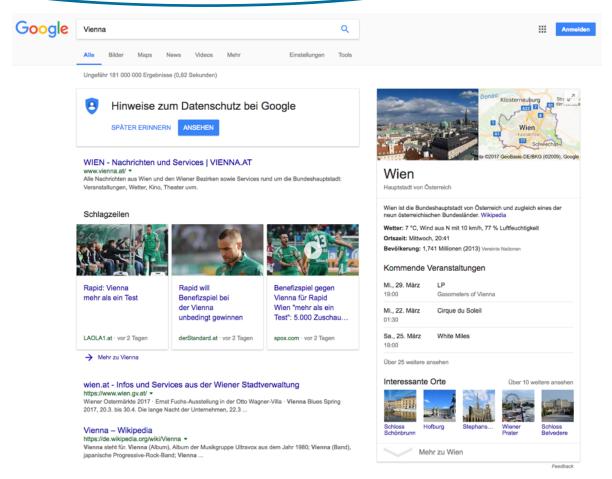
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS



## Which solutions exist now on the Web?



### Welche Lösungen sind im Einsatz?



## **Example 1:**Google's Knowledge Graph

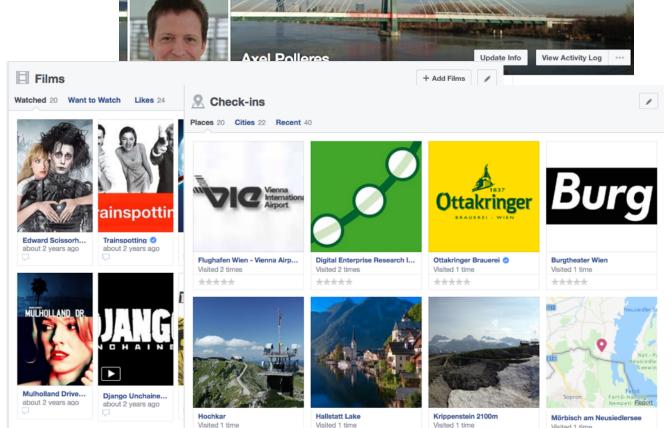
### Which solutions exist now on the Web?



Welche Lösungen sind im Einsatz?

\*\*\*\*

Axel Polleres



\*\*\*\*

\*\*\*\*

Example 2: FB's Social Graph & News Recommendations

1 0 0

Visited 1 time

\*\*\*\*

Also uses a knowledge graph...

## Which solutions exist now on the Web?



Welche Lösungen sind im Einsatz?



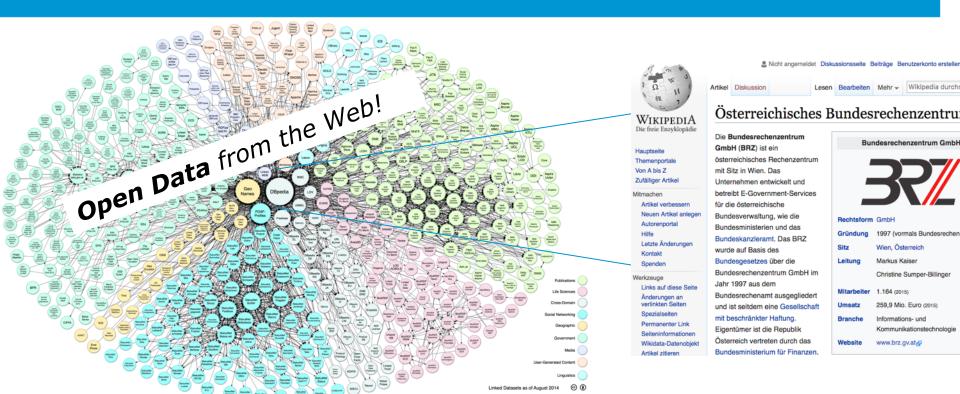
Example 3: IBM Watson!

Also uses a knowledge graph...

https://youtu.be/P00bm0DBvwI?t=951







Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak.

http://lod-cloud.net/



### 3. What is Open Data?

- **Availability and Access**: data must also be available in a convenient and modifiable form, at no more than a reasonable reproduction cost.
- **Reuse and Redistribution**: data must be provided under terms that permit reuse and redistribution and must be <u>machine-readable</u>.
- Universal Participation: everyone must be able to use, reuse and redistribute it (no discrimination)

See more at: <a href="http://opendefinition.org/okd/">http://opendefinition.org/okd/</a>





### Open Data is a global trend:

Cities, International Organizations, National and European Portals, Int'l. Conferences:











**london**.gov.uk



**European Union Open Data Portal** 













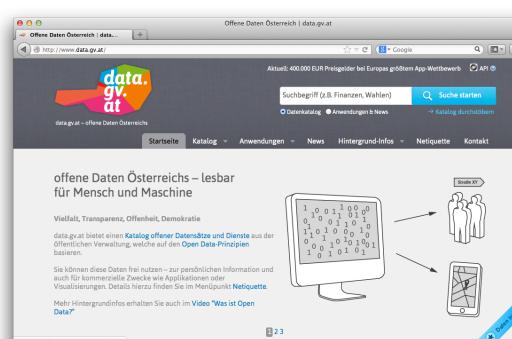




### CKAN ... <a href="http://ckan.org/">http://ckan.org/</a>

- almost "de facto" standard for Open Data Portals
- facilitates search, metadata (publisher, format, publication date, license, etc.) for datasets
- http://opendataportal.at/
- http://data.gv.at/

machine-processable? ...... partially



## Our research: data.wu.ac.at



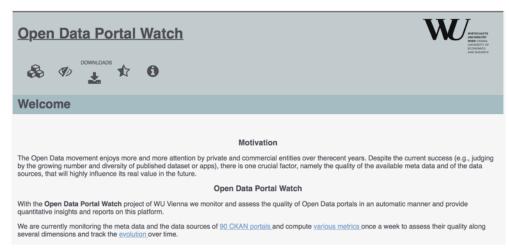
- What is the status of Open Data and what are the challenges using Open Data?
  - OpenData PortalWatch a project at WU
- How can Open Data be used by enterprises?
  - Open City Data Pipeline a joint project with Siemens on using Open Data in an Enterprise context!
- What's next?
  - Improving Open Data Quality and Access: ADEQUATE (FFG project)
  - Making Open Data Searchable
  - Building an Open Data Knowledge Graph!
- A striving Data Economy needs no silos... re-democratise the Web by Congitive Intelligence based on Open Data?





### http://data.wu.ac.at/portalwatch/

- Periodically monitoring a list of Open Data Portals
  - 260 CKAN powered Open Data Portals worldwide
- Quality assessment
- Evolution tracking
  - Meta data
  - Data
  - Formats, growth





## What have we learnt? Open Data also has the "Vs"





### Volume:

 It's growing! (we currently monitor 90 CKAN portals, 512543 resources/ 160069 datasets, at the moment (statically) ~1TB only CSV files...



### Variety:

- different datasets (from different cities, countries, etc.), only partially comparable, partially not.
- Different metadata
- Different data formats



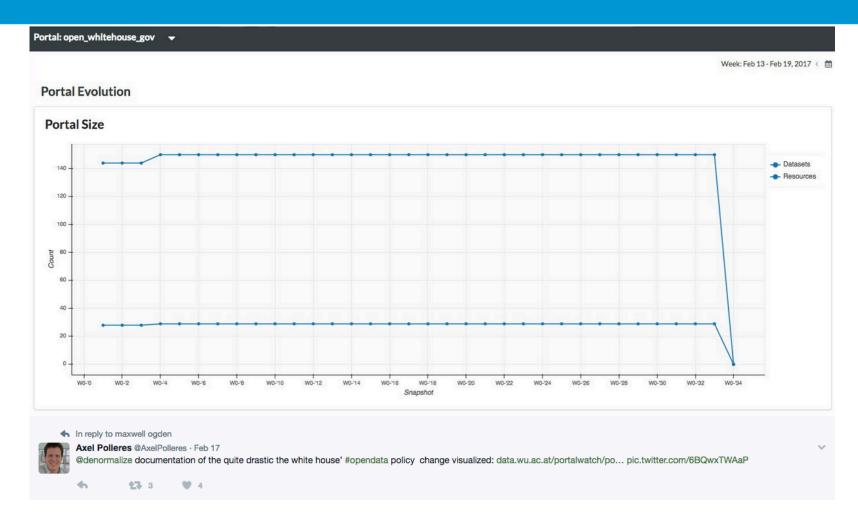
### Velocity:

- Open Data changes regularly (fast and slow)
- New datasets appear, old ones disappear



### **Portalwatch Example:**







### Our research: data.wu.ac.at



- What is the status of Open Data and what are the challenges using Open Data?
  - OpenData PortalWatch a project at WU
- How can Open Data be used by enterprises?
  - Open City Data Pipeline a joint project with Siemens on using Open Data in an Enterprise context!
- What's next?
  - Improving Open Data Quality and Access: ADEQUATE (FFG project)
  - Making Open Data Searchable
  - Building an Open Data Knowledge Graph!
- A striving Data Economy needs no silos... re-democratise the Web by Congitive Intelligence based on Open Data?

### Use Case: City Data - Important for **Infrastructure Providers & for City Decision Makers**



- City Assessment and Sustainability reports
- Tailored offerings by Infrastructure Providers















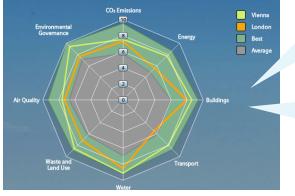




US Mayors

European Green City Index

- → Needs up-to-date City Data and calculates City KPIs in a way that allows to display the current state and run scenarios of different product applications.
- e.g. towards a "Dynamic" Green City Index:



... however, these are often **outdated** before even published!

### Goal (short term):

 Leverage Open Data for calculating a city' performance from public sources on the Web automatically

#### Goal (long term):

Define and Refine KPI models to assess specific impact of infrastructural investments and gather/check input automatically



### City Data Pipeline (started 2012)



http://citydata.wu.ac.at/



### **SIEMENS**

### Open City Data Pipeline

We present the City Data Pipeline – a system for gathering city performance indicators published as Open Data in order to ease the compilation of studies and reports used within Siemens. Under the assumption that Open Data provides means to automatise tedious data research tasks, we have built a system that integrates basic indicators for cities from various Open Data sources. The architecture is flexible, extensible, and natively based on RDF & SPARQL.

**Launch Open City Data Pipeline** 

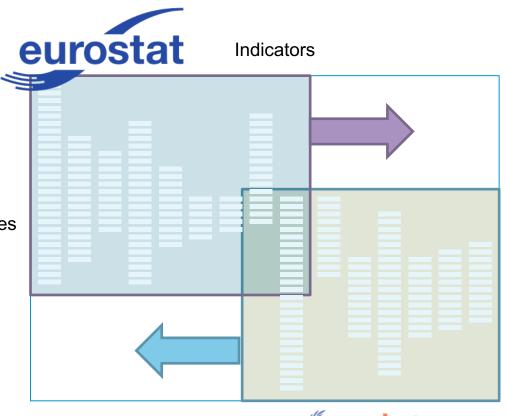


## City Data Pipeline: could be viewed as a cognitive computing use case



Goal: understanding, Cities combining, enriching, different open datasets

using both firstwave and second wave AI methods







### **City Data Pipeline**

### citydata.wu.ac.at

- Search for indicators & cities
- obtain results incl. sources
- Integrated data served as Linked Open Data
- Predicted values AND estimated error rates for missing data...



### **SIEMENS**

### Vienna 🍕

#### Municipal waste (1000 t)

- > 2004: 778.905392176222 1000 t (from http://citydata.wu.ac.at /ns#Prediction, predicted by with an estimated error of %RMSE)
- > 2005: 813.77643147163 1000 t (from http://citydata.wu.ac.at /ns#Prediction, predicted by with an estimated error of %RMSE)
- 2006: 813.889824195497 1000 t (from <a href="http://citydata.wu.ac.at/ns#Prediction">http://citydata.wu.ac.at/ns#Prediction</a>, predicted by with an estimated error of %RMSE)
- > 2007: 811.538914636665 1000 t (from http://citydata.wu.ac.at /ns#Prediction, predicted by with an estimated error of %RMSE)
- 2008: 811.010344391444 1000 t (from <a href="http://citydata.wu.ac.at/ns#Prediction">http://citydata.wu.ac.at/ns#Prediction</a>, predicted by with an estimated error of %RMSE)





Sustainable Cities Results



## **SIEMENS**

#### **Berlin**

Population male 2012

1717645.0 persons

(Source: http://epp.eurostat.ec.europa.eu/)

Population male 2011

1695438.0 persons (Source: http://data.un.org/)

Population male 2011

1695438.0 persons

(Source: http://epp.eurostat.ec.europa.eu/)

Population male 2010

1686256.0 persons

(Source: http://epp.eurostat.ec.europa.eu/)

Population male 2009

1686256.0 persons

#### Vienna

Population male 2011

821605.0 persons (Source: http://data.un.o

Population male 2010

812867.0 persons (Source: http://data.un.o

Population male 2009

807088.0 persons (Source: http://data.un.o

Population male 2009

807088.0 persons

(Source: http://epp.eurostat.ec.europa.eu/)

(Source, http://e

Population male 2008 801776.0 persons (Source: http://data.un.o

Population male 2008

800361.0 persons

...it's not finished, but: assumption: Predictions get better, the more Open data we integrate...



## Our research: data.wu.ac.at



- What is the status of Open Data and what are the challenges using Open Data?
  - OpenData PortalWatch a project at WU
- How can Open Data be used by enterprises?
  - Open City Data Pipeline a joint project with Siemens on using Open Data in an Enterprise context!
- What's next?
  - Improving Open Data Quality and Access: ADEQUATE.at
  - Making Open Data Searchable
  - Building an Open Data Knowledge Graph!



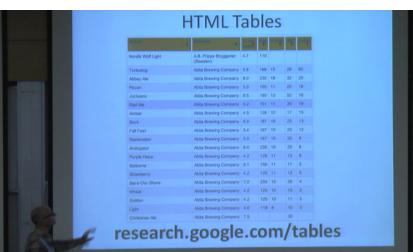


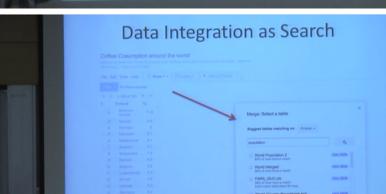
## Why is Search in Open Data a problem?



https://www.youtube.com/watch?v=kCAymmbYIvc

Structured Data in Web Search by Alon Halevy







NUTS2	NUTS3	DISTRICT_CODE	SUB_DISTRICT_CODE	POP_TOTAL	POP_MEN	POP_WOMEN	REF_DATE
AT13	AT130	90101	0	16131	7726	8405	01.01.2014
AT13	AT130	90201	0	99597	48650	50947	01.01.2014
AT13	AT130	90301	0	86454	41085	45369	01.01.2014
AT13	AT130	90401	0	31452	14903	16549	01.01.2014
AT13	AT130	90501	0	53610	26299	27311	01.01.2014
AT13	AT130	90601	0	30613	14833	15780	01.01.2014
AT13	AT130	90701	0	30792	14703	16089	01.01.2014
AT13	AT130	90801	0	24279	11855	12424	01.01.2014
AT13	AT130	90901	0	40528	19286	21242	01.01.2014
AT13	AT130	91001	0	186450	91638	94812	01.01.2014
AT13	AT130	91101	0	93440	45541	47899	01.01.2014
AT13	AT130	91201	0	90874	43752	47122	01.01.2014

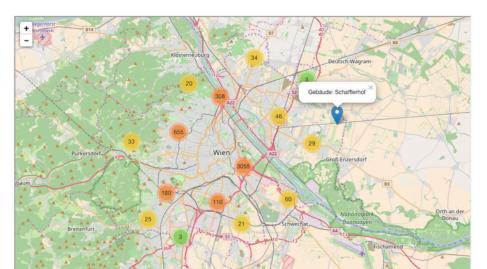
### Open Data Search is hard...

- a) No natural language "cues" like in Web tables...
- b) Existing knowledge graphs don't cover the domain of "Open Data"
- c) Open Data is not properly geo-referenced





- First baby steps on building an Open Data Knowledge Graph:
- Ongoing work to make
- Open Data geo-searchable e.g. in our project <u>communidata.at</u>:



#### International Semantic Web conference 2016:

#### Multi-level semantic labelling of numerical values

Sebastian Neumaier<sup>1</sup>, Jürgen Umbrich<sup>1</sup>, Josiane Xavier Parreira<sup>2</sup>, and Axel Polleres<sup>1</sup>

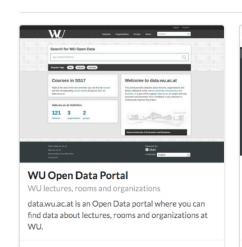
Vienna University of Economics and Business, Vienna, Austria
Siemens AG Österreich, Vienna, Austria

Abstract. With the success of Open Data a huge amount of tabular data sources became available that could potentially be mapped and linked into the Web of (Linked) Data. Most existing approaches to "semantically label" such tabular data rely on mappings of textual information to classes, properties, or instances in RDF knowledge bases in order to link - and eventually transform - tabular data into RDF. However, as we will illustrate, Open Data tables typically contain a large portion of numerical columns and/or non-textual headers; therefore solutions that solely focus on textual "cues" are only partially applicable for mapping such data sources. We propose an approach to find and rank candidates of semantic labels and context descriptions for a given bag of numerical values. To this end, we apply a hierarchical clustering over information taken from DBpedia to build a background knowledge graph of possible "semantic contexts" for bags of numerical values, over which we perform a nearest neighbour search to rank the most likely candidates. Our evaluation shows that our approach can assign fine-grained semantic labels, when there is enough supporting evidence in the background knowledge graph. In other cases, our approach can nevertheless assign high level contexts to the data, which could potentially be used in combination with other approaches to narrow down the search space of possible labels.



## Ongoing Projects (data.wu.ac.at)







#### DBpedia Wayback Machine

Extract past DBpedia versions

0

The DBpedia Wayback Machine aims at providing the wayback functionality for DBpedia based on the revisions of their Wikipedia article.

#### **Projects**



#### Open Data Portal Watch

Monitoring & exposing portals' metadata

Open Data Portal Watch assesses the evolution of the (meta) data quality of about 260 Open Data portals over since September 2014.

III 259 portals





#### Jupyter Notebook Server

Programming & Documentation

Notebook documents are documents which contain both computer code (e.g. python) and human-readable rich text elements.

Only available within local WU Vienna network



#### Open Data AT Assistant

Search chatbot for Austrian datasets

The assistant will help you to explore the content of the austrian open data portals: data.gv.at and opendataportal.at.



## Our research: data.wu.ac.at



- What is the status of Open Data and what are the challenges using Open Data?
  - OpenData PortalWatch a project at WU
- How can Open Data be used by enterprises?
  - Open City Data Pipeline a joint project with Siemens on using Open Data in an Enterprise context!
- What's next?
  - Improving Open Data Quality and Access: ADEQUATE.at
  - Making Open Data Searchable
  - Building an Open Data Knowledge Graph!
- A striving Data Economy needs no silos... re-democratise the Web by Congitive Intelligence based on Open Data?

### Quora

Q Search for questions, people, and topics

Quora uses cookies to improve your experience. Read more

### This is that IB

Google Knowledge Graph Facebook Graph Search

### Out of Facebook Graph Search and Google Knowledge Graph, which is more revolutionary, creative and useful?

Say after both these graphs grow to their full extent Compare and contrast Facebook's Introducing Graph Search ☑ and Google's http://www.google.co.in/insidese... I in terms of

- Revolution to the internet
- 2.Creativity in their design
- 3.Usefulness to the users

#### 3 Answers

"Both our knowldege graph and google's habe roots in Wikipedia and freebase"

- but none of Google and FB make their knowledge graphs freely and openly available again as Open Data!



Justin Moore, Engineering Manager at Facebook Written Mar 19, 2013

You're comparing apples to oranges. Facebook has graph search \*and\* a knowledge graph (although we didn't give it a name externally that I know of). Both our knowledge graph and google's have roots in Wikipedia and freebase and both are semantic knowledge stores. Search for baseball (sport) on Facebook and scroll through the page to see our knowledge graph about players, teams, etc.

Linking

Graph search is different. Its structured semantic search on top of structured data like MCCr. knowledge graph but also all of your connections to people, photos, places. You can't really judge the two any more than comparing the Internet and google. http:/

ohn P. Эk.



## This is a fundamental threat to the Web itself:



https://www.theguardian.com/technology/2017/mar/11/tim-berners-lee-web-inventor-save-internet

Internet

Tim Berners-Lee: I invented the web. Here are three things we need to change to save it

It has taken all of us to build the web we have, and now it is up to all of us to build the web we want - for everyone

- 1) We've lost control of our personal data
- 2) It's too easy for misinformation to spread on the web
- 3) Political advertising online needs transparency and understanding





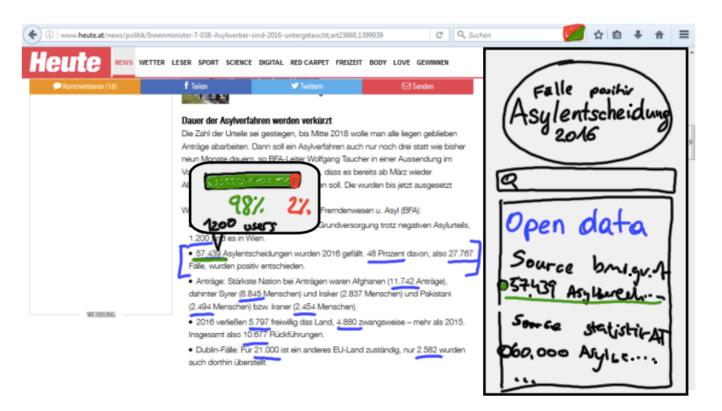
1 Sir Tim Berners-Lee, inventor of the worldwide web. Photograph: Sarah Lee for the Guardian



## Open-Data-Fueled Cognitive Computing to the rescue!



 http://www.communidata.at/austrian-opendata-day-students-co-create-open-data-ideas/





## Open-Data-Fueled Cognitive Computing to the rescue!

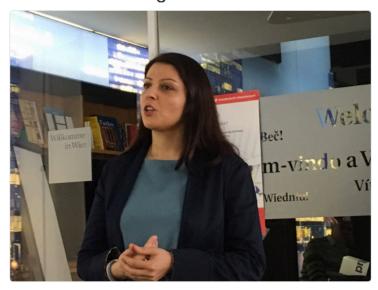




Roland Ledinger @Roland\_Ledinger · 3. März

**£3** 10

Business Treff Open Data Day Statement von Staatssekretärin Duzdar offene Daten sind für Alle wichtig





Axel Polleres @AxelPolleres · Mar 3

Staatssekretärin Muna #Duzdar at #opendataday: #fakenews? #opendata to the rescue!



# Bottomline: Let's build Open Knowledge Graphs as a basis for Cognitive Computing for all businesses!

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

23.3.2017 16:30 bis 23.03.2017 18:30

## KICK-OFF: DIGITAL INFORMATION MANAGEMENT COMMUNITY AUSTRIA

Datum 23.3.2017 16:30 - 23.03.2017 18:30

Ort Impact Hub Vienna

künstliche Intelligenz virtuelle Assistenten werden im privaten und geschäftlichen Alltag immer präsenter. Digitale Lösungen nutzen Cognitive Computing afbereitete Informationen und

Prozessautomatisierung. Disruptive digitale Businessmodelle revolutionieren ganze Branchen.

Mit welchen Themen sollten wir uns befassen?

Mit Weieren

Welche Lösungen sind im Einsatz?

spannende Diskussionen sowie einen projektorientierten Erfahrungsaustausch zwischen Experten aus dem Public Sector und der Privatwirtschaft. #OpenData





Wenn Sie mehr zu den Themen **Semantic Web, Cognitive Computing, Linked Open Data** erfahren wollen:

https://iswc2017.semanticweb.org/



Wir veranstalten von 21.-25.10. an der WU die

16. International Semantic Web Conference (ISWC2017)

## Things I did NOT have time to talk about:



- Open Data and licences → DALICC
- Open Data adoption barriers → see our recent paper to be presented at <u>CEDEM2017</u>
- Privacy and data on the Web → <a href="http://privacylab.at">http://privacylab.at</a>