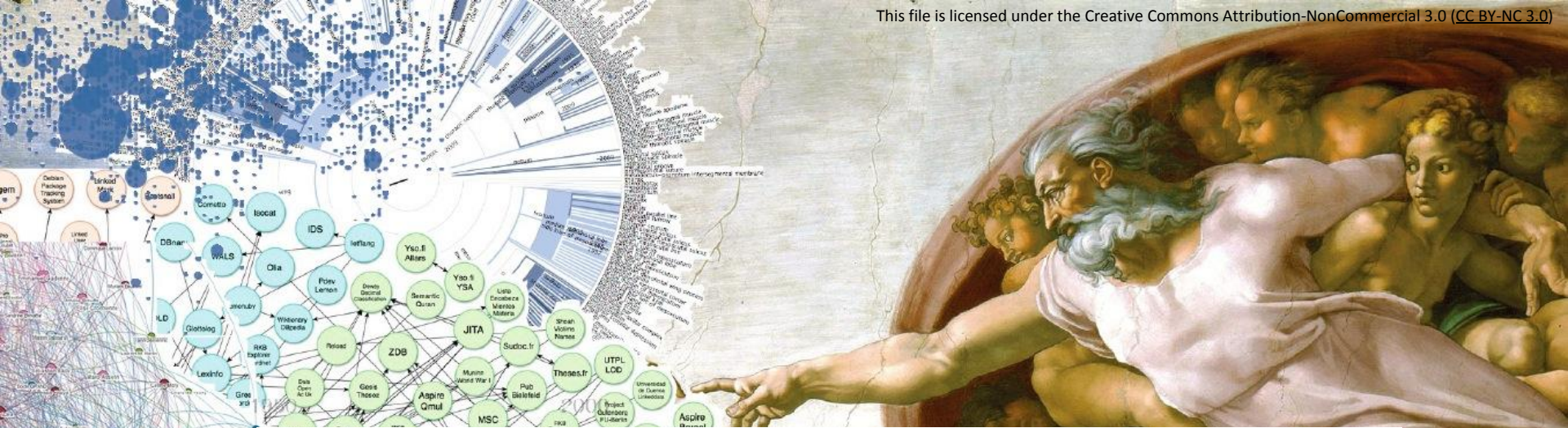




Information Service Engineering

Lecture 6 - Putting it All Together

6.2 Knowledge Mining and Information Extraction



FIZ Karlsruhe

Leibniz Institute for Information Infrastructure



Hasso
Plattner
Institut



Karlsruher Institut für Technologie

Prof. Dr. Harald Sack & Dr. Maria Koutraki
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Spring 2018

Information Service Engineering

Lecture 6: Putting it All Together

6.1 Linked Data Programming

6.2 Knowledge Mining and Information Extraction

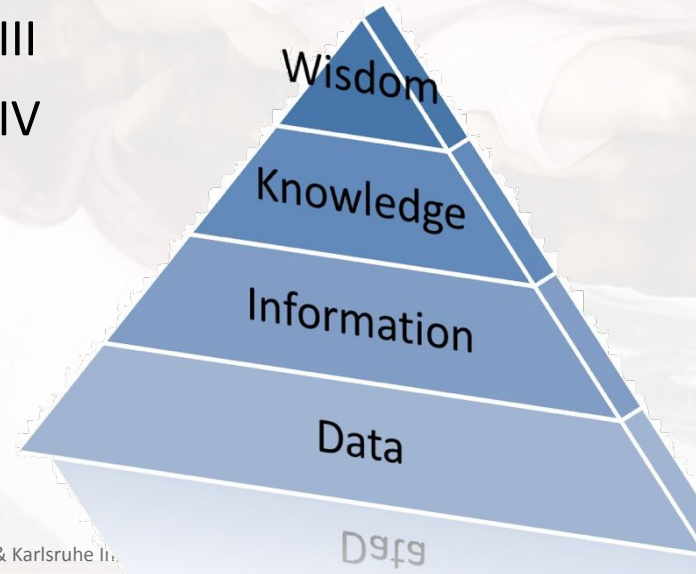
6.3 Knowledge Mining and Information Extraction II

6.4 Knowledge Mining and Information Extraction III

6.5 Knowledge Mining and Information Extraction IV

6.6 Semantic Search

6.7 Exploratory Search



Knowledge Mining

understanding principles
(why?, what is best?, doing things right)

collective application of knowledge in context

understanding patterns
(principles: how to?)

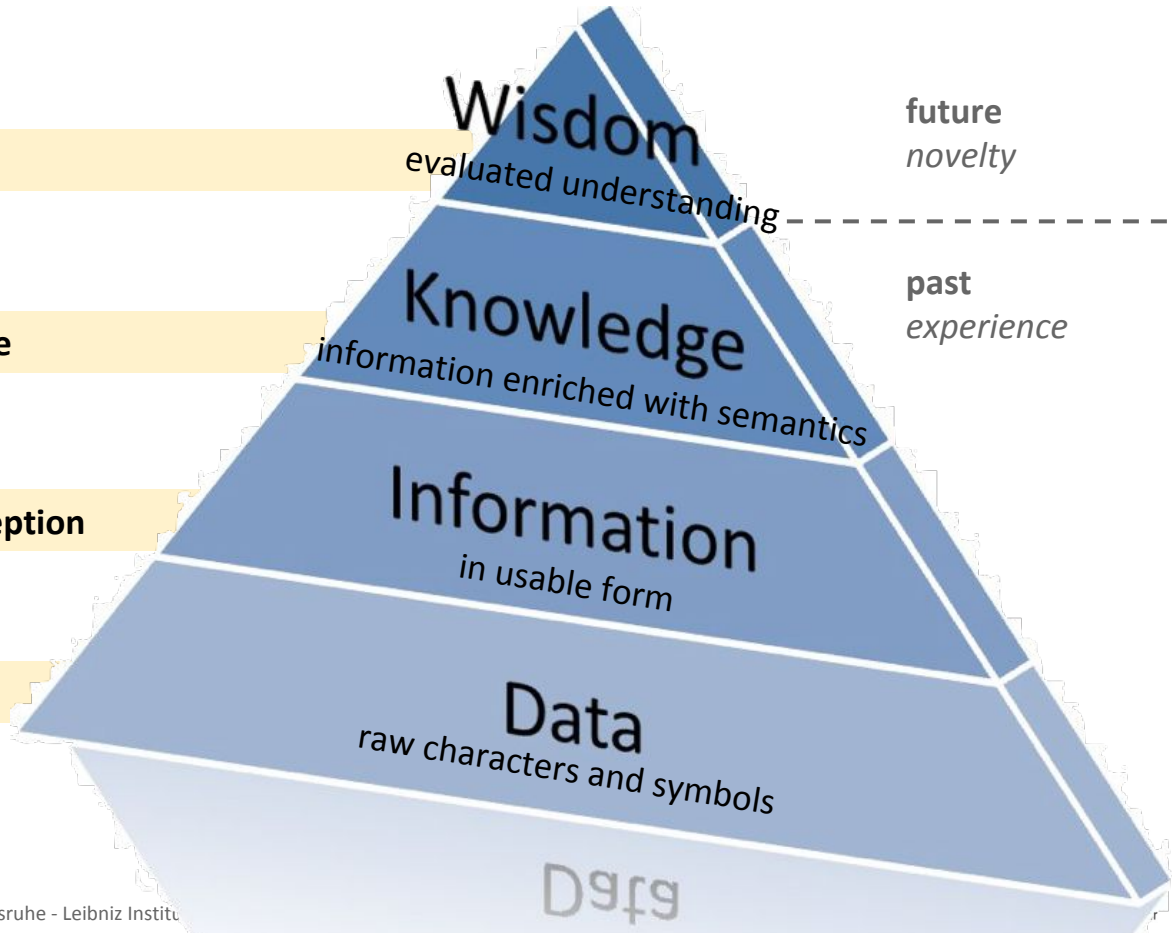
experience, context, value applied to a message

understanding relations
(description: what?)

a message meant to change the receivers perception

discrete objective facts about event

DIKW Pyramid, Ackoff 1989 [1]



Data

Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **Data** is raw.
- It simply exists and has **no significance** beyond its existence (in and of itself).
- It can exist **in any form**, usable or not.

Information

Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **Information** is data that has been given **meaning** by way of **relational connection**.
- This "meaning" can be **useful**, but does not have to be.
- **Information** is **contained in descriptions**, answers to questions that begin with such words as *who*, *what*, *when*, *where*, and *how many*

Knowledge

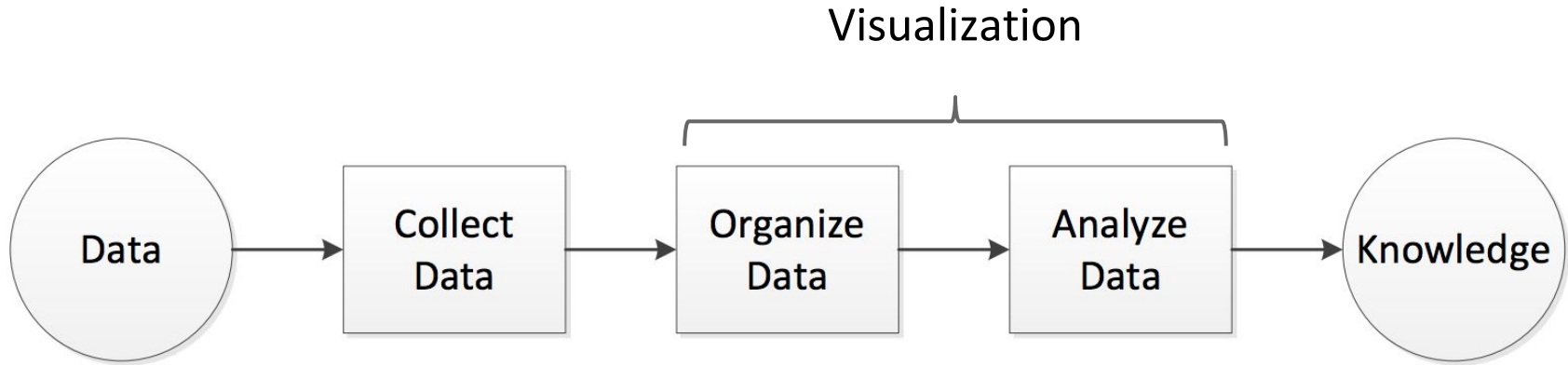
Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **Knowledge** is the appropriate collection of information, such that it's intent is to be useful.
- **Understanding** is a continuum that leads from **data**, through **information** and **knowledge**, and ultimately to **wisdom**
- **Data** transforms to **information** by convention, **information** to **knowledge** by cognition, and **knowledge** to **wisdom** by contemplation

Knowledge Mining

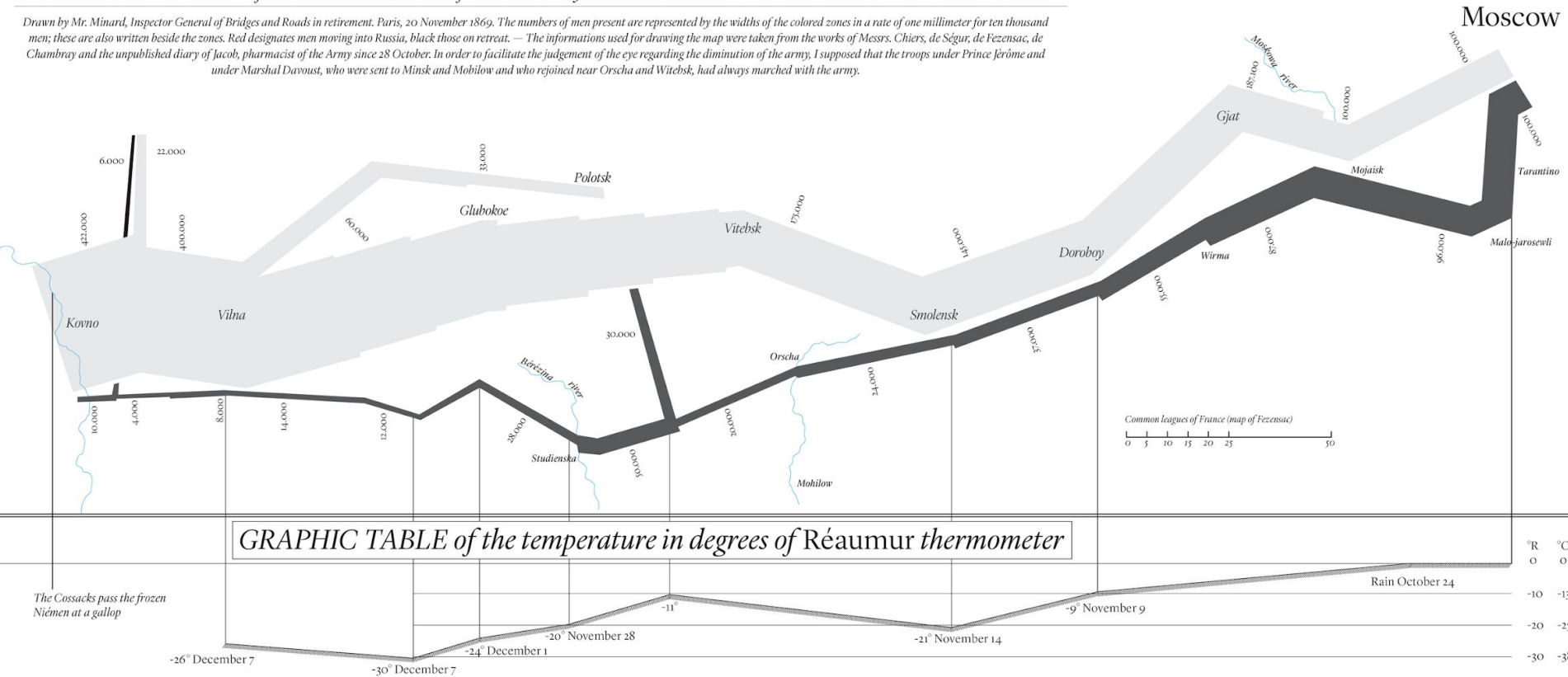
Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- How do we transform data into knowledge?
 1. Collect Data
 2. Organize Data
 3. Analyze Data



FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement, Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Chiers, de Ségur, de Fenezac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davoust, who were sent to Minsk and Mohilow and who rejoined near Orscha and Witebsk, had always marched with the army.

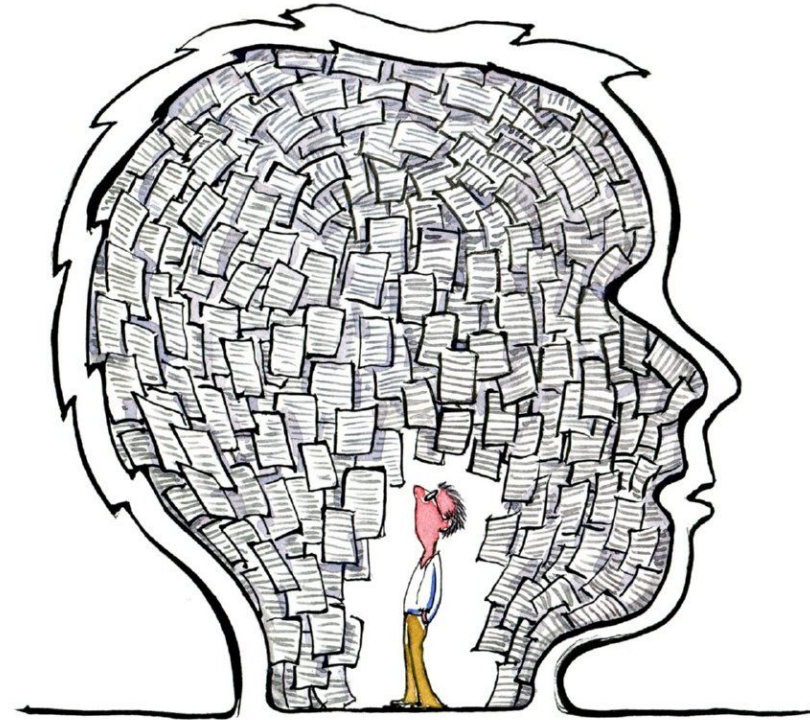


Famous Infographics by Charles Minard (1781-1870)

A Picture is worth a 1000 words

Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- Pictures have been used to convey information long before the development of writing
- A single picture can be processed (“understood”) much faster than a (linear) text page
- Human perception is processing in **parallel**, text analysis is limited by the **sequential** process of reading



https://commons.wikimedia.org/wiki/File%3AA_picture_is_worth_a_thousand_words.jpg

Information Visualization

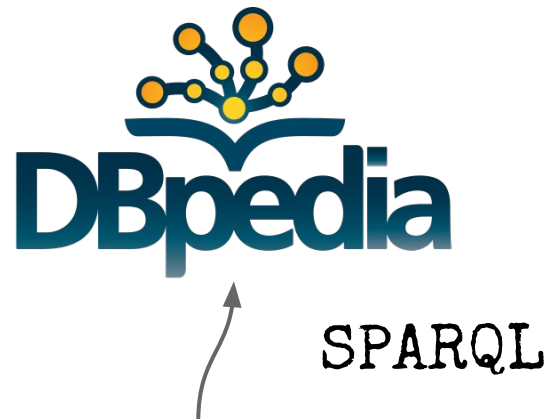
Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **Information Visualization** is the study of (interactive) visual representations of abstract data to reinforce human cognition
- **Information graphics** or **infographics** are graphic visual representations of information, data or knowledge intended to present information quickly and clearly
 - a static form of information visualization
 - aims to emphasize specific findings gained from the visualized data
 - **Mandatory precondition:** Data Analysis

Quick Data Visualization Example

Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **Dataset:** DBpedia



- **Task:**
Draw a **map chart** which indicates the **number of soccer players per country**

Geo Chart



e.g. via GoogleDoc

Quick Data Visualization Example

Information Service Engineering - Lecture 6 Putting it All Together - 6.2 Knowledge Mining and Information Extraction

- **SPARQL Query:**
 - **Country and Number of soccerplayers per country**
 - `?s rdf:type dbo:SoccerPlayer .`
 - `?s dbo:birthPlace ?birthplace .`
 - `?birthplace dbo:country ?country .`
 - `?country rdf:type dbo:Country .`
 - `?country rdfs:label ?countryLabel`
 - `FILTER (lang(?countryLabel)="en")`
 - `GROUP BY ?countryLabel`
 - `COUNT(DISTINCT ?s)`

Default Data Set Name (Graph IRI)

Query Text

```
select ?clabel (COUNT (DISTINCT ?s) as ?count) WHERE {
?s rdf:type dbo:SoccerPlayer ;
  dbo:birthPlace ?birthplace .
?birthplace dbo:country ?country .
?country rdf:type dbo:Country FILTER NOT EXISTS {?country dbp:lifeSpan ?d .}.
?country rdfs:label ?countryLabel FILTER (lang(?countryLabel)="en").
BIND (STR(?countryLabel) as ?clabel)
}
GROUP BY ?clabel
ORDER BY DESC(?count)
```

[SPARQL query](#)

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

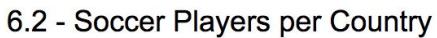
Execution timeout:

 milliseconds (values less than 1000 are ignored)

Options:

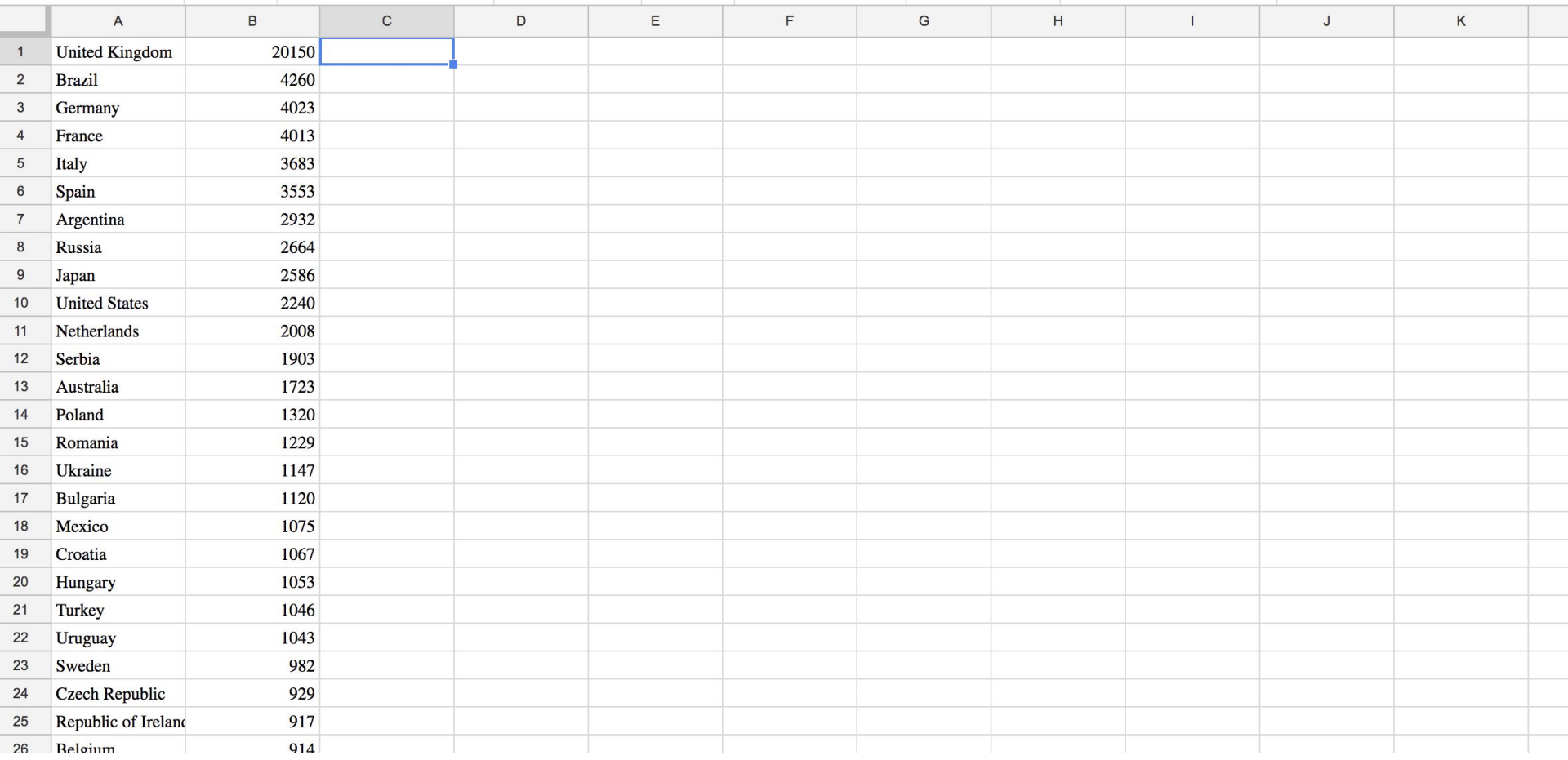
- ☒ Strict checking of void variables
- ☐ Log debug info at the end of output (has no effect on some queries and output formats)
- ☐ Generate SPARQL compilation report (instead of executing the query)

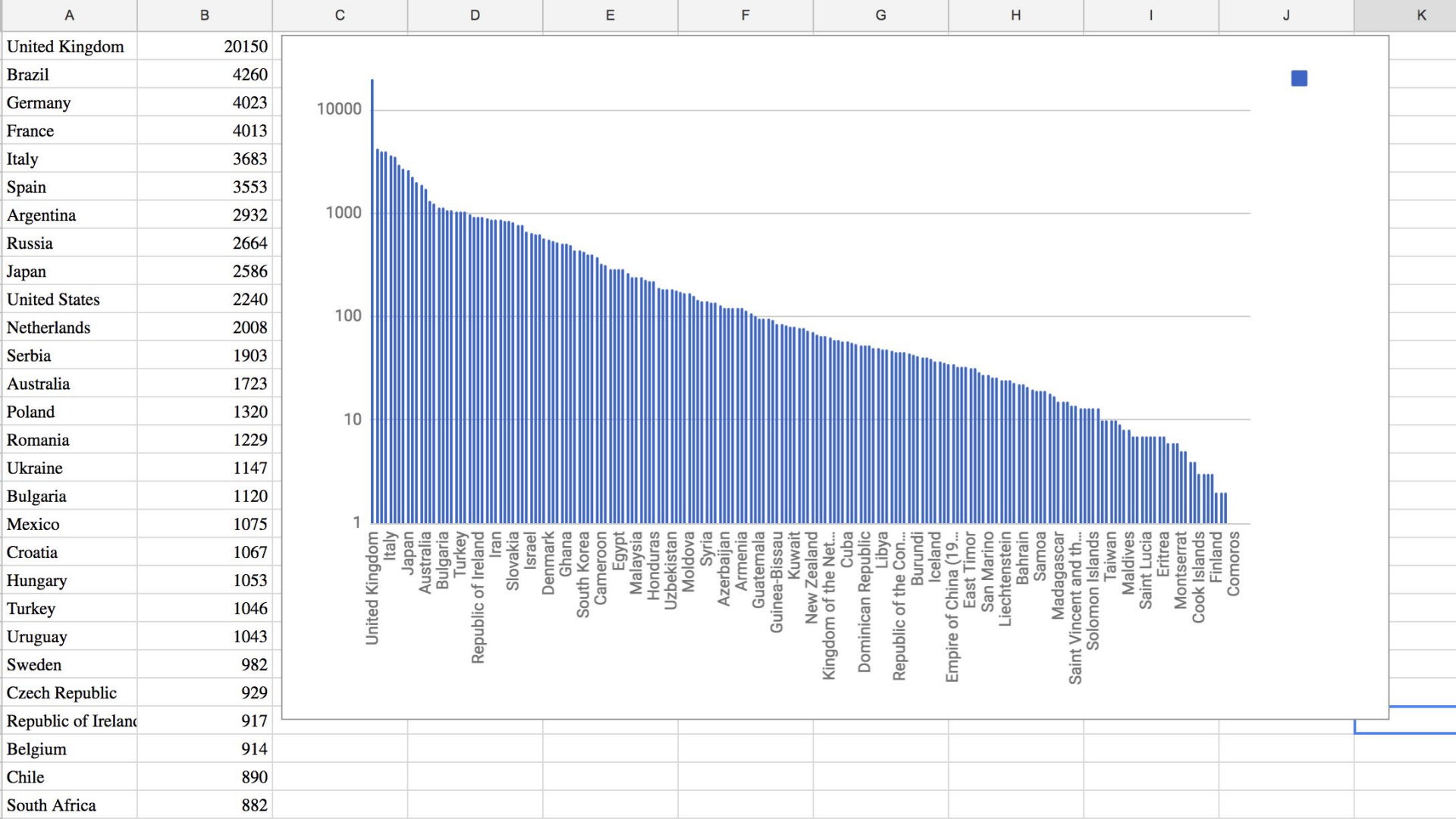
(The result can only be sent back to browser, not saved on the server, see [details](#))

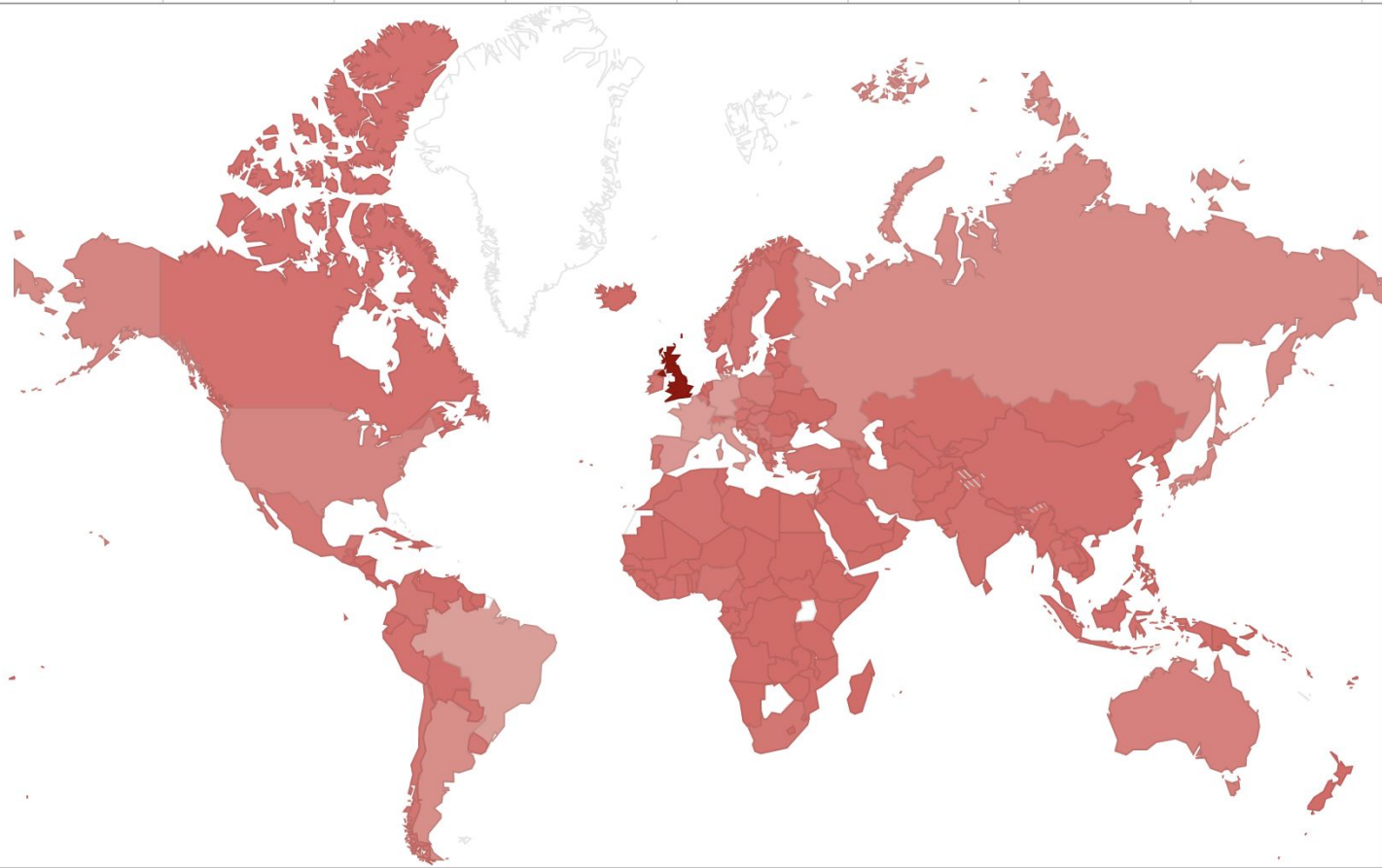


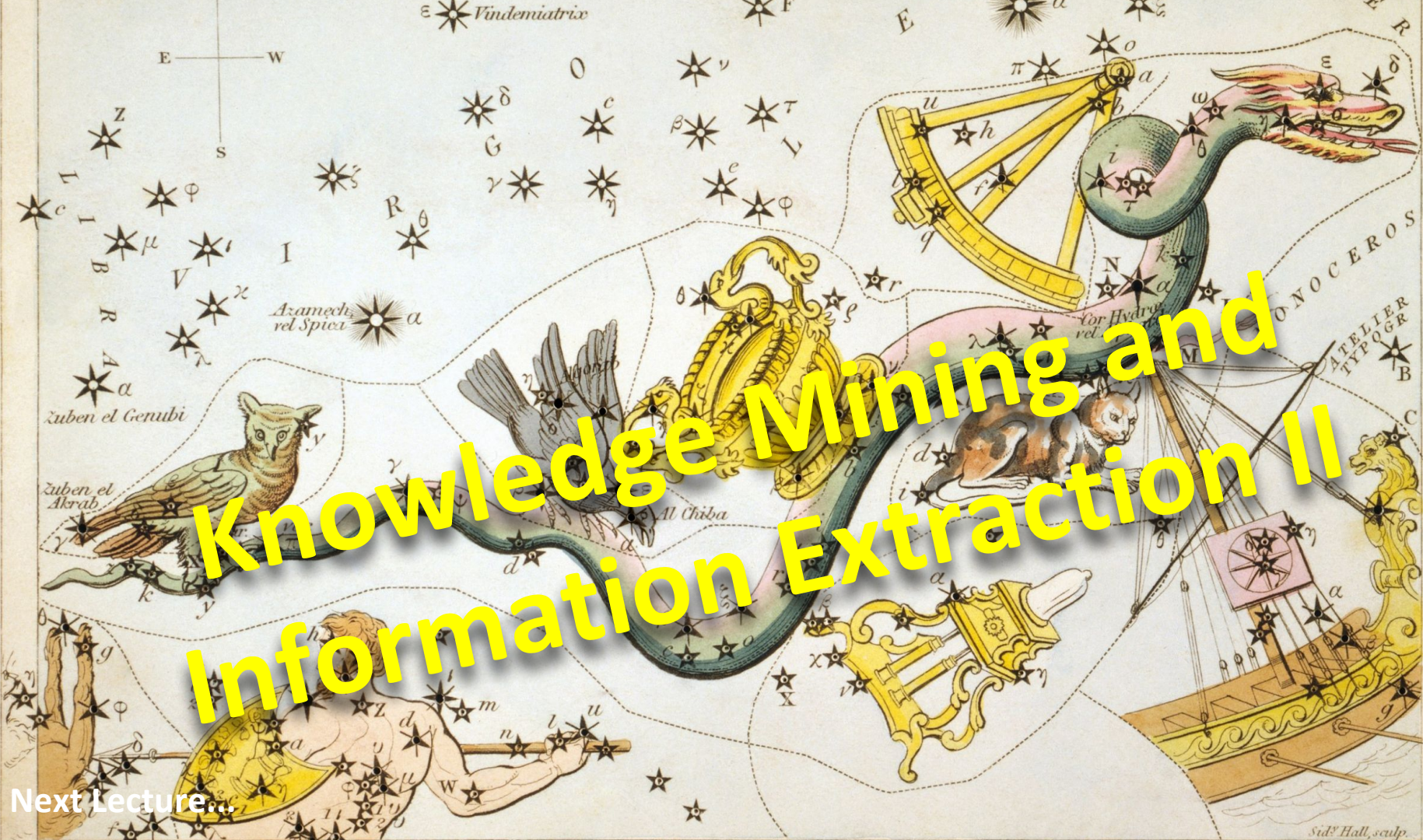
File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

[GoogleDoc Sheet with Data](#)





A	B	C	D	E	F	G	H	I	J	K
United Kingdom	20150									
Brazil	4260									
Germany	4023									
France	4013									
Italy	3683									
Spain	3553									
Argentina	2932									
Russia	2664									
Japan	2586									
United States	2240									
Netherlands	2008									
Serbia	1903									
Australia	1723									
Poland	1320									
Romania	1229									
Ukraine	1147									
Bulgaria	1120									
Mexico	1075									
Croatia	1067									
Hungary	1053									
Turkey	1046									
Uruguay	1043									
Sweden	982									
Czech Republic	929									
Republic of Ireland	917									
Belgium	914									
Chile	890									
South Africa	882									



Next Lecture...

Picture References:

- P.8: Charles Minard's map of Napoleon's disastrous Russian campaign of 1812.
https://en.wikipedia.org/wiki/Charles_Joseph_Minard#/media/File:Minard_map_of_napoleon.png
- P.9: HiKingArtist, A picture is worth a thousand words, https://commons.wikimedia.org/wiki/File%3AA_picture_is_worth_a_thousand_words.jpg
- P.17: Sidney Hall, "Noctua, Corvus, Crater, Sextans Uraniaë, Hydra, Felis, Lupus, Centaurus, Antlia Pneumatica, Argo Navis, and Pyxis Nautica", plate 32 in Urania's Mirror, a set of celestial cards accompanied by A familiar treatise on astronomy ... (1825)
https://commons.wikimedia.org/wiki/File:Sidney_Hall_-_Urania%27s_Mirror_-_Noctua,_Corvus,_Crater,_Sextans_Urani%C3%A6,_Hydra,_Felis,_Lupu_s,_Centaurus,_Antlia_Pneumatica,_Argo_Navis,_and_Pyxis_Nautica.jpg