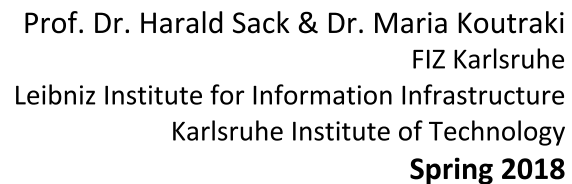


6.4 Knowledge Mining and Information Extraction III



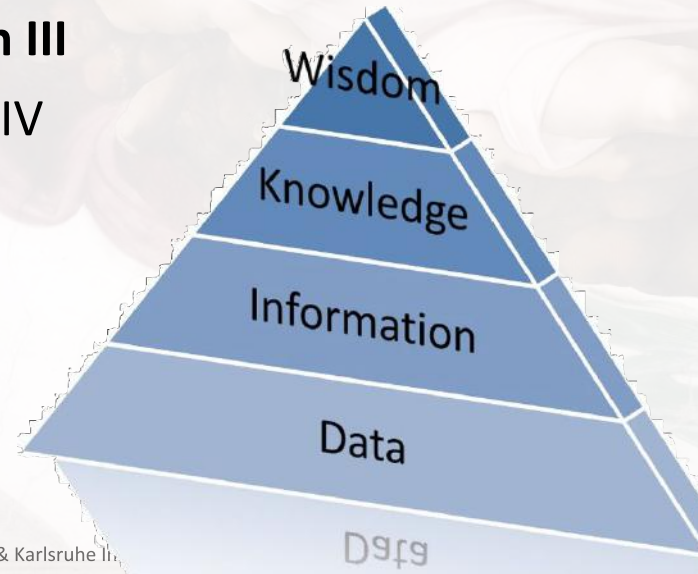
HPI Hasso Plattner Institut



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



Hands-On Data Analysis Example Task

Information Service Engineering - Lecture 6 Putting it All Together - 6.4 Knowledge Mining and Information Extraction III



Hands-On Data Analysis Example Task

Information Service Engineering - Lecture 6 Putting it All Together - 6.4 Knowledge Mining and Information Extraction III

3. CleanUp Data

This might require
Several rounds...

6.3 - Soccer Stats Example

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

	A	B	C	D	E	F	G	H	I
1	total_goals	height	weight	year	pos	bday			
2	1859	180	74	1943	forward	1943-06-11			
3	1457	178	89	1982	fly-half	1982-03-05			
4	1358	172	75	1940	midfielder	1940-10-23			
5	1318	190	84	1970	midfielder	1970-08-05			
6	1142	177	78	1977	forward	1977-06-08			
7	946	187	83.5	1985	forward	1985-02-05			
8	902	170	77	1977	winger	1977-02-26			
9	849	164	57	1975	forward	1975-02-08			
10	831	184	80	1955	goalkeeper	1955-05-16			
11	691	181	84	1976	goalkeeper	1976-10-21			
12	646	180	68	1947	midfielder	1947-04-25			
13	628	192	80	1954	midfielder	1954-02-19			
14	612	165	68	1960	midfielder	1960-10-30			
15	611	172	72	1907	forward	1907-09-09			
16	594	186	73	1972	midfielder	1972-04-19			
17	592	172	72	1927	forward	1927-04-02			
18	586	170	72	1933	midfielder	1933-10-28			
19	581	171	69	1973	forward	1973-08-11			
20	574	180	82	1976	midfielder	1976-09-27			
21	562	188	85	1982	goalkeeper	1982-02-13			
22	560	182	82	1963	defender	1963-03-21			
23	557	176	68	1945	forward	1945-11-03			
24	556	183	72	1976	centre-back	1976-09-29			
25	552	185	85	1972	goalkeeper	1972-12-14			
26	548	183	78	1969	midfielder	1969-09-29			
27	525	181	83	1985	midfielder	1985-06-04			
28	522	181	76	1977	forward	1977-10-25			
29	497	173	75	1913	forward	1913-09-06			

+ Wikidata CleanedUp Wikidata original DBpedia

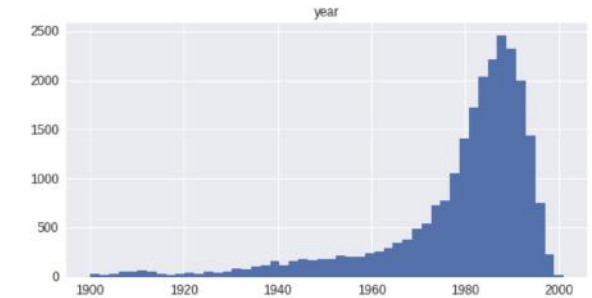
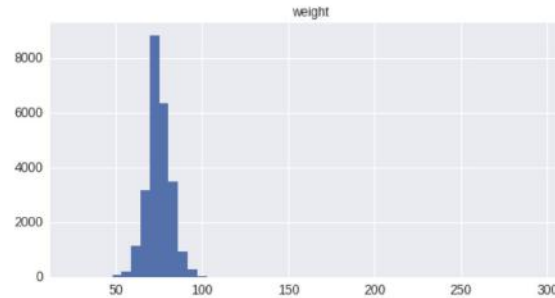
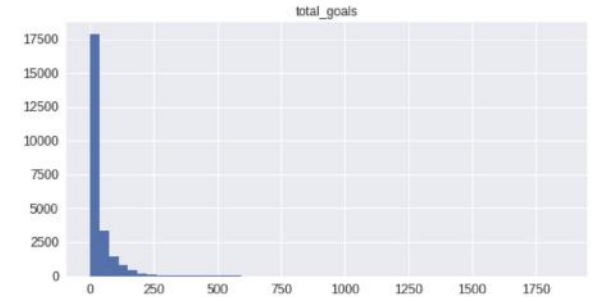
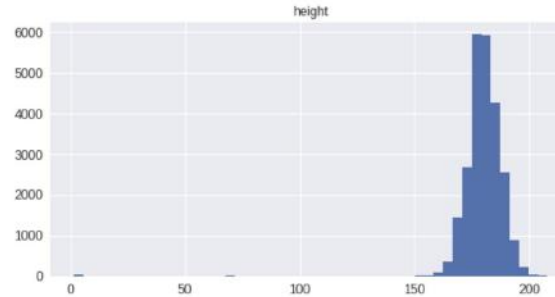
[Data in Google Doc Spreadsheet](#)

Hands-On Data Analysis Example Task

Information Service Engineering - Lecture 6 Putting it All Together - 6.4 Knowledge Mining and Information Extraction III

4. Analyse the Data

E.g. via python



[Data in Google Collab Notebook](#)

Tabula terræ Nouæ Zemblæ,
 in qua fretum sinusq; WAIGATS,
 item ora littoralis TARTARIÆ atq;
 RUSSIÆ, ad urbem usq; KILDEINAM,
 præscriptum adhuc cursus quem inde
 naves in reditu tenent scilicet septentrionalis
 litus et TRAIECTUS
 prope fretum WAIGATS ad Rusiæ
 terram et ad promontorium CANDENOS
 atq; fauces usq; MARIS ALBI.
 AUTHORE GERHARDO DE VEER.

Knowledge Mining and Information Extraction IV

Picture References:

- P.6: Tabula terrae Nouae Zemblae in qua fretum sinusq Waigats item ord littoralis Tartariae atq Russiae, ad urgem usq Kildeinam, Autore Gerhardo De Veer. 1601,
https://commons.wikimedia.org/wiki/File:1601_De_Bry_and_de_Veer_Map_of_Nova_Zembla_and_the_Northeast_Passage_-_Geographicus_-_Nova_Zembla-debry-1601.jpg