

Označevalni jeziki 3

Povezovanje dejstev,
izjave so trojčki,
Resource Description Framework,
svet kot graf.

Povezovanje dejstev

- ❖ Dejstva so sestavina informacijskih objektov.
- ❖ Povezovanje dejstev omogoča povezovanje informacijskih objektov.
- ❖ Zamislimo si iskanje v zbirki:
 - ❖ iskalec dobi bibliografski zapis o filmu F , kjer piše, da ga je režiral režiser R , scenarij pa napisal scenarist S ,
 - ❖ Sistem za poizvedovanje iskalcu sporoči, da
 - ❖ pozna o R informacije tudi iz informacijskih objektov $ioR1$, $ioR2$ in $ioR3$,
 - ❖ pozna informacije o filmih $ioF1$, $ioF2$ in $ioF3$, ki jih je tudi režiral R ,
 - ❖ pozna informacije o filmih $ioF4$, $ioF5$ in $ioF6$, posnetih po scenarijih, ki jih je napisal S ...
 - ❖ *Tak sistem je dober sistem.*

grand scenario: Semantični splet

- ❖ Povezovanje dejstev je jedro Semantičnega spleta.
- ❖ Semantični splet je vizionarski načrt, ki pretvarja Svetovni splet v eno samo razpršeno podatkovno zbirko.
- ❖ Svetovni splet: splet **dokumentov** za človeškega bralca.
- ❖ Semantični splet: splet **dejstev** za programskega bralca.
- ❖ Nekatera orodja in otoki Semantičnega spleta že delujejo.
- ❖ Razvoj Semantičnega spleta vodita Tim Berners-Lee in W3C.

Povezovanje dejstev

- ❖ V vsakem primeru, neglede na nivo, je treba
 - ❖ dejstva prepoznati,
 - ❖ jih izjaviti na standarden način,
 - ❖ prepoznati druge izjave
 - ❖ in jih povezati, če govorijo o istih pojavih.
- ❖ Jedro postopkov povezovanja dejstev je standardizirano izjavljanje o njih.
- ❖ Dejstva zapisujemo v formatu **RDF (Resource Description Framework, Okvir za opisovanje virov)**.
- ❖ **Vir je karkoli**, o čemer se da kaj izjaviti in **ima URI** .

RDF

Enostavna dejstva:

- ❖ Dr. Strangelove **je ime** filma 1.
- ❖ Clockwork Orange **je ime** filma 2.
- ❖ Stanley Kubrick **je režiser** filma 1.
- ❖ Stanley Kubrick **je režiser** filma 2.
- ❖ Režiser **je oseba**.

Izpeljano dejstvo:

- ❖ Filma Dr. Strangelove in Clockwork Orange **je režirala ista oseba**.

Resource Description Framework (RDF)

- ❖ V skladu z RDF izjave o dejstvih zapisujemo v obliki **trojčkov**.
- ❖ Trojček sestavljajo **subjekt**, **predikat** in **objekt**. Predikat imenujemo tudi lastnost.
- ❖ Primer trojčka:
film F (subjekt) je_režiral (predikat) režiser R (objekt).
- ❖ Z enostavnimi trojčki je mogoče izraziti vse podatke, ki jih imamo o pojavih na svetu.
- ❖ RDF le zapisuje, zapisanemu pa ne daje pomena.
- ❖ Pomen dodajamo izjavam s pomočjo **ontologij**.

RDF se običajno zapisuje z XML

```
1. <rdf:RDF
2.   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
3.   xmlns:activity="http://www.binkmovies.org/activities#">
4.     <rdf:Description rdf:about="http://www.binkmovies.org/movies#blade-runner">
5.       <activity:is-directed-by
rdf:resource="http://www.binkmovies.org/people#ridley-scott" />
6.       <!-- alternativna oblika predikata in objekta -->
7.       <activity:is-directed-by>Ridley Scott</activity:is-directed-by>
8.     </rdf:Description>
9. </rdf:RDF>
```

2. URI, kjer je na voljo razlaga ključnih besed RDF, npr. rdf:Description in rdf:resource.

3. URI izmišljenega strežnika, kjer je na voljo razlaga pojmov o dejavnostih v filmski industriji.

4.-7. Opis dejstva je gnezden med `<rdf:Description>` in `</rdf:Description>`.

4. rdf:about uvaja subjekt, opisan je na danem URI.

5. activity je izmišljen imenski prostor, definiran v vrstici 03; activity:is-directed-by je predikat.

Dodajanje semantičnih informacij v RDF (RDFS in OWL)

- ❖ Podatke v RDF dopolnimo s semantičnimi informacijami z uporabo RDFS (RDF Schema) in OWL (Web Ontology Language).
- ❖ Tudi RDFS in OWL sta projekta W3C.
- ❖ Namen ontologije je **klasifikacija stvari in njihovih pomenov**.
 - ❖ "Stvari" so uvrščene v razrede in podrazrede s člani, ki jih imenujemo **individui**, posamezniki.
 - ❖ Individui so uvrščeni v **razrede** glede na njihove skupne lastnosti.
 - ❖ Če je individuum član razreda R to pomeni, da zanj veljajo vse lastnosti razreda R.
- ❖ Tudi ontologija je izražena v RDF.

Zakaj komplicirati, saj imamo splet in zbirke

- ❖ Spletni dokumenti so
 - ❖ slabo strukturirani,
 - ❖ podatki so prepleteni z ostalim besedilom in
 - ❖ avtomatski postopki jih iz njih težko izolirajo.
- ❖ Podatkovne zbirke so dobro strukturirani nosilci informacij, a strukturiranost ni standardna in se razlikuje od zbirke do zbirke.
- ❖ Podatkovne zbirke so zato izolirani otoki informacij.
- ❖ Bolj ko je struktura podatkov splošno sprejeta in dobro definirana, lažje naredimo programe, ki podatke lahko izkoristijo tudi na načine, ki niso vnaprej znani.
- ❖ Povezovanje dejstev in tvorba novega znanja, ki se dogaja sproti, med iskanjem in odvisno od konkretne iskalne zahteve, je tak način izkoriščanja podatkov.

Zakaj komplicirati, saj imamo splet in zbirke

- ❖ Poizvedovanje po podatkih je v zgodovini informatike vedno delovalo na vnaprej pripravljenih podatkovnih strukturah.
- ❖ Česar ni bilo v podatkovni zbirki, tega iskanje ni moglo najti.
- ❖ Želimo narediti sisteme, sposobne povezovanja dejstev in tvorbe novega znanja, ki se dogaja sproti, med iskanjem in odvisno od konkretne iskalne zahteve.
- ❖ Semantični splet obeta samo-organizirajoče se, rastoče zakladnice znanja, sposobne tvorjenja novega znanja in uporabne na načine, ki niso vnaprej predvideni.

Zakaj komplicirati, saj imamo splet in zbirke

- ❖ Verjamemo v prihodnost Semantičnega spleta, ker mislimo,
 - ❖ da je mogoče obstoječe podatke strukturirati tako, da bo njihova struktura splošno sprejeta v globalnih okvirih (to je mogoče doseči le z zelo enostavno standardno strukturo);
 - ❖ da so trojčki RDF taka enostavna, splošno sprejeta in dobro definirana struktura;
 - ❖ da povezovanje podatkov v globalnih okvirih lahko prinese nesluten napredek v informiranju in izrabi informacij.

Zakaj komplicirati, saj imamo splet in zbirke

- ❖ Trojčki RDF, so na spletu shranjeni v t.i. repozitorijih RDF.
- ❖ Aplikacije, ki povezujejo podatke, navadno zgradijo iz trojčkov RDF, ki so na voljo, nekakšno ohlapno podatkovno zbirko.
- ❖ Tako podatkovno zbirko imenujemo *triplestore*.
- ❖ Za povezovanje podatkov ni dovolj le jasna, standardna struktura (npr. RDF) – potrebujemo tudi *standardno razumevanje podatkov*.
- ❖ Standardno razumevanje prinašajo standardne ontologije.

Linking Open Data

- ❖ Končni cilj imenujemo Semantični splet.
- ❖ Pot do njega pa gibanje Linked Data, utemeljeno v projektu **Linking Open Data (LOD)**, ki ga vodi W3C.
- ❖ Projekt se je začel leta 2007.
- ❖ Cilji:
 - ❖ identifikacija obstoječih podatkovnih setov, ki so v odprtem dostopu,
 - ❖ pretvorba v trojčke RDF in njihovo objavljanje na spletu v repozitorijih RDF,
 - ❖ povezovanje repozitorijev RDF v globalno mrežo.

Kako napreduje LOD

❖ Click to edit Master text styles

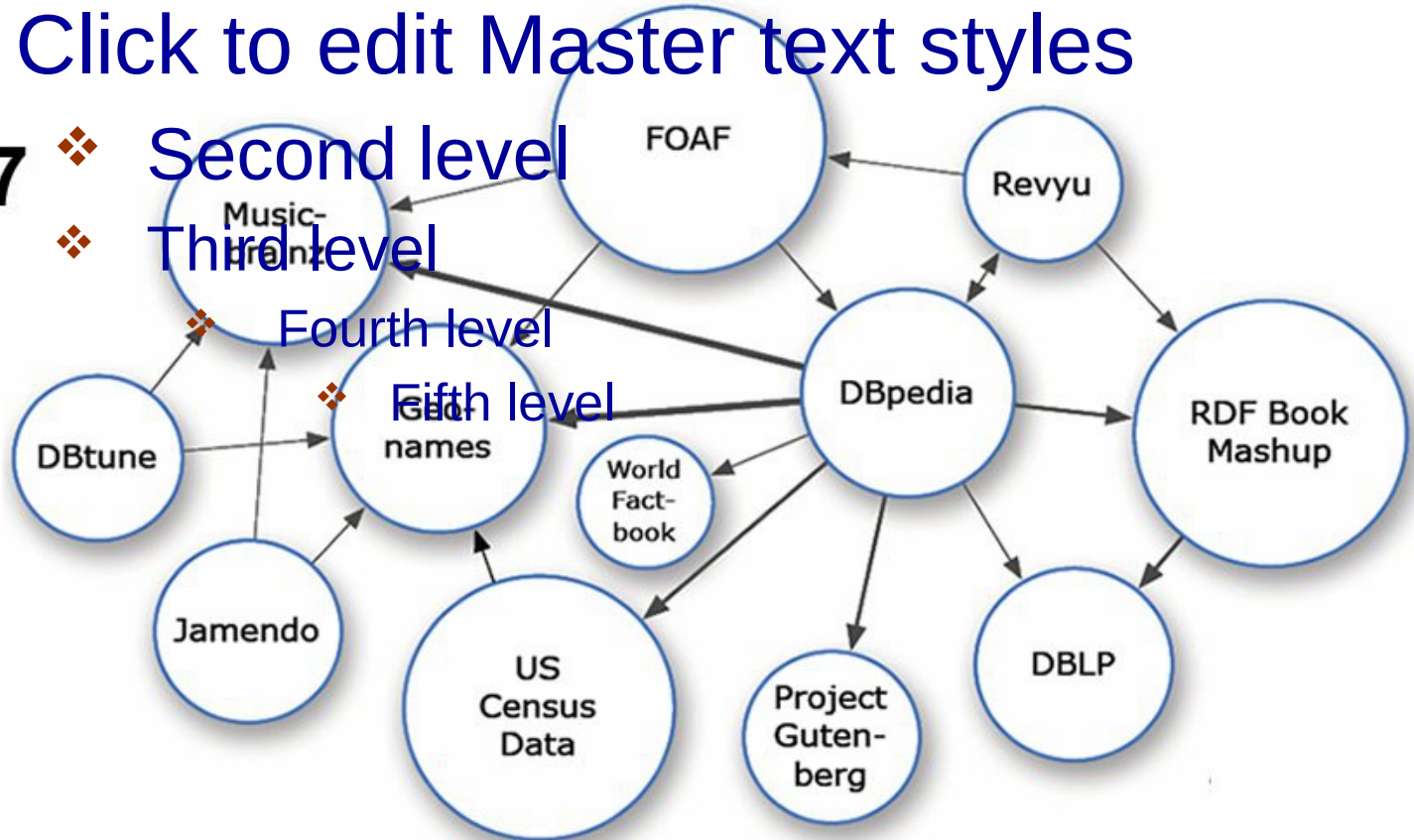
May 2007

❖ Second level

❖ Third level

❖ Fourth level

❖ Fifth level



Kako napreduje LOD

September 2008

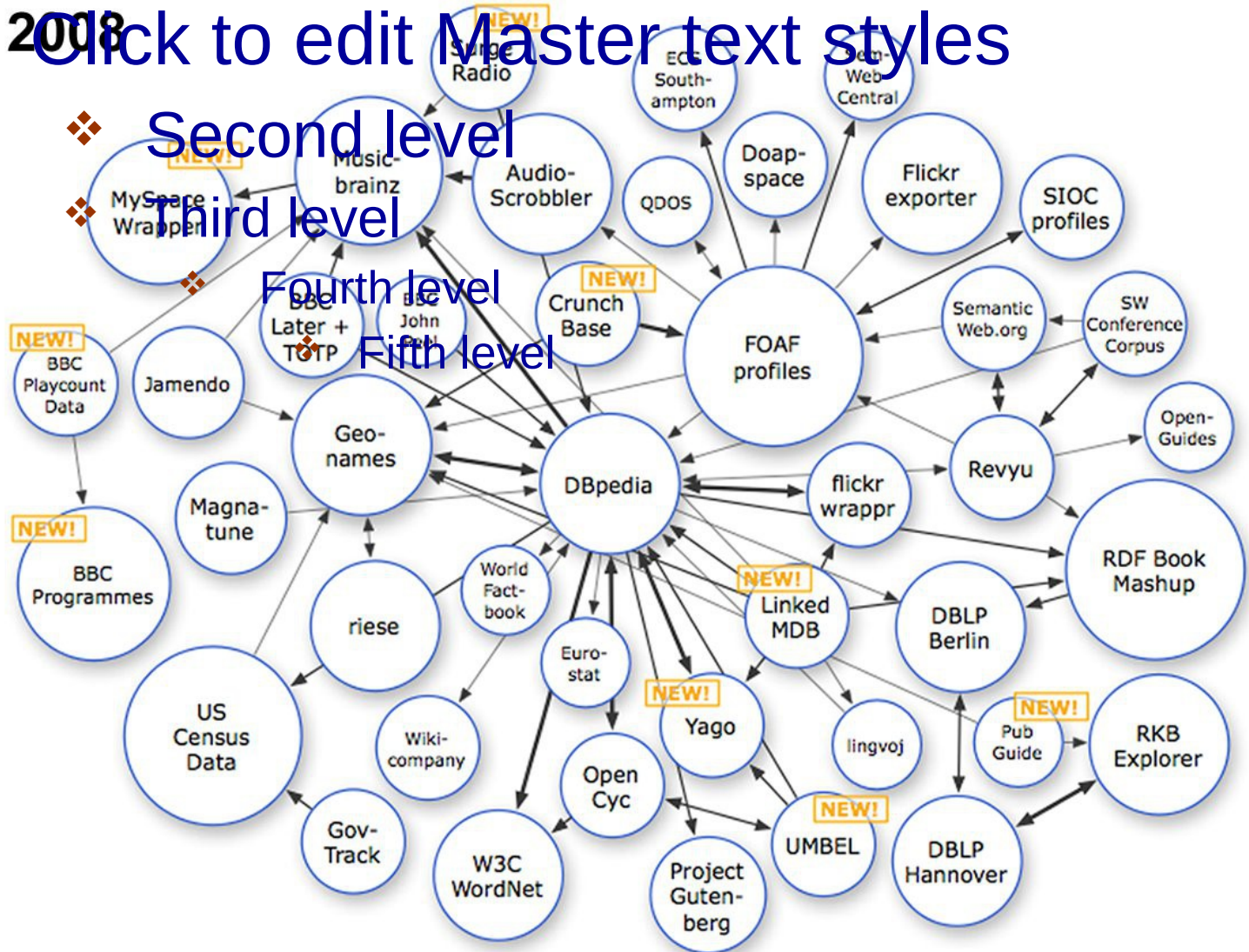
Click to edit Master text styles

❖ Second level

❖ Third level

Fourth level

❖ Fifth level



Kako napreduje LOD

September 2010

Click to edit Master text styles

Second level

Third level

Fourth level

Fifth level



Kako napreduje LOD

September 2011

Click to edit Master text styles

Second level

Third level

Fourth level

Fifth level



DBpedia

❖ Click to edit Master

❖ Trojčki o stvarih, ki so strukturirane in avtomatsko identificirane na straneh Wikipedie.

❖ Podatki v trojčkih RDF pretežno prihajajo iz t.i. "infoboxes", podnapisov slik, zemljepisnih koordinat in kazalcev na druge spletne strani.

❖ Nastala 2007 v Free University of Berlin in University of Leipzig, od tedaj se pospešeno razvija.



The screenshot shows the DBpedia infobox for the project 'DBpedia'. It features the DBpedia logo at the top, which consists of a network of nodes and edges. Below the logo, the text 'DBpedia' is written in a large, stylized font. The infobox contains the following information:

- Initial release:** 23 January 2007
- Stable release:** DBpedia 3.8 / 6 Aug 2012^[1]
- Written in:** Scala, Java, VSP
- Operating system:** Virtuoso Universal S
- Type:** Semantic Web, Link Data
- License:** GNU General Public License
- Website:** dbpedia.org

primera infoboxes



The screenshot shows the Wikipedia infobox for the project 'Wikipedia'. It features the Wikipedia logo at the top, which is a globe made of puzzle pieces with various characters and symbols. Below the logo, the text 'WIKIPEDIA' is written in a large, bold font. The infobox contains the following information:

- URL:** wikipedia.org
- Slogan:** The Free Encyclopedia
- Commercial?** No
- Type of site:** Internet encyclopedia
- Registration:** Optional, but is required for certain tasks such as editing protected pages, creating pages in English Wikipedia and uploading files
- Available language(s):** 275 active editions (285 in total)
- Users:** 35,000,000 (total registered in all editions)^[1]
- Content license:** Creative Commons Attribution/Share-Alike 3.0 (most text also dual-licensed under GFDL) Media licensing varies
- Owner:** Wikimedia Foundation (non-profit)
- Created by:** Jimmy Wales, Larry Sanger^[2]
- Launched:** January 15, 2001 (11 years ago)
- Alexa rank:** 6 (January 2013)^[3]
- Current status:** Active

DBpedia v številkah

- ❖ Več kot 1 milijarda trojčkov o 3,64 milijonih stvareh.
- ❖ Med drugim
 - ❖ 416.000 oseb,
 - ❖ 526.000 krajev,
 - ❖ 106.000 glasbenih albumov,
 - ❖ 60.000 filmov,
 - ❖ 17.500 video iger,
 - ❖ 169.000 organizacij,
 - ❖ 183.000 bioloških vrst in
 - ❖ 5.400 bolezni).

DBpedia v številkah

- ❖ Stvari poimenovane v 97 jezikih.
- ❖ Kazalci na
 - ❖ 2.724.000 slik izven Wikipedie,
 - ❖ 6.300.000 spletnih strani izven Wikipedie,
 - ❖ 6.200.000 zunanjih repozitorijev RDF.
- ❖ Primer poizvedovanja po LOD z jezikom SPARQL

DBpedia: rojeni v Ljubljani pred 1900-01-01

SPARQL results:

name	birth	death	person
"Ferdinand Augustin Haller von Hallerstein, or Liu Songling"@en	"1703-09-27"^^xsd:date	"1774-10-29"^^xsd:date	:Ferdinand_Augustin_Hallerstein 
"Ferdinand Augustin Hallerstein, or Liu Songling"@en	"1703-09-27"^^xsd:date	"1774-10-29"^^xsd:date	:Ferdinand_Augustin_Hallerstein 
"Albin Prepeluh"@en	"1881-02-22"^^xsd:date	"1937-11-20"^^xsd:date	:Albin_Prepeluh 
"Albin Prepeluh"@en	"1881-02-22"^^xsd:date	"1937-11-20"^^xsd:date	:Albin_Prepeluh 
"Albin Prepeluh"@en	"1881-02-22"^^xsd:date	"1937-11-20"^^xsd:date	:Albin_Prepeluh 
"Albin Prepeluh"@en	"1881-02-22"^^xsd:date	"1937-11-20"^^xsd:date	:Albin_Prepeluh 
"Anton Alexander von Auersperg"@en	"1806-04-11"^^xsd:date	"1876-09-22"^^xsd:date	:Count_Anton_Alexander_von_Auersperg 
"Anton Alexander von Auersperg"@en	"1806-04-11"^^xsd:date	"1876-09-22"^^xsd:date	:Count_Anton_Alexander_von_Auersperg 
"Anton Alexander von Auersperg"@en	"1806-04-11"^^xsd:date	"1876-09-22"^^xsd:date	:Count_Anton_Alexander_von_Auersperg 
"Anton Alexander von Auersperg"@en	"1806-04-11"^^xsd:date	"1876-09-22"^^xsd:date	:Count_Anton_Alexander_von_Auersperg 
"Anton Funtek"@en	"1862-10-30"^^xsd:date	"1932-10-21"^^xsd:date	:Anton_Funtek 
"Ferdinand Augustin Hallerstein"@en	"1703-09-27"^^xsd:date	"1774-10-29"^^xsd:date	:Ferdinand_Augustin_Hallerstein 
"Ferdinand Augustin Hallerstein"@en	"1703-08-27"^^xsd:date	"1774-10-29"^^xsd:date	:Ferdinand_Augustin_Hallerstein 
"Ferdo Kozak"@en	"1894-10-18"^^xsd:date	"1957-12-08"^^xsd:date	:Ferdo_Kozak 
"Fran Ramovš"@en	"1890-09-14"^^xsd:date	"1952-09-16"^^xsd:date	:Fran_Ramov%C5%A1 
"Fran Ramovš"@en	"1890-09-14"^^xsd:date	"1952-09-16"^^xsd:date	:Fran_Ramov%C5%A1 
"Fran Ramovš"@en	"1890-09-14"^^xsd:date	"1952-09-16"^^xsd:date	:Fran_Ramov%C5%A1 
"Fran Ramovš"@en	"1890-09-14"^^xsd:date	"1952-09-16"^^xsd:date	:Fran_Ramov%C5%A1 
"Fritz Pregl"@en	"1869-09-03"^^xsd:date	"1930-12-13"^^xsd:date	:Fritz_Pregl 
"Fritz Pregl"@en	"1869-09-03"^^xsd:date	"1930-12-13"^^xsd:date	:Fritz_Pregl 
"Fritz Pregl"@en	"1869-09-03"^^xsd:date	"1930-12-13"^^xsd:date	:Fritz_Pregl 
"Fritz Pregl"@en	"1869-09-03"^^xsd:date	"1930-12-13"^^xsd:date	:Fritz_Pregl 
"Gvido Birolla"@en	"1881-06-12"^^xsd:date	"1963-05-29"^^xsd:date	:Gvido_Birolla 
"Gvido Birolla"@en	"1881-06-12"^^xsd:date	"1963-05-29"^^xsd:date	:Gvido_Birolla 

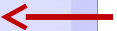
Click to edit Master text styles

Second level

Third level

Fourth level

Fifth level



DBpedia: stvar, imenovana Ferdinand Hallerstein

rdf:type	owl:Thing
rdf:type	foaf:Person
rdf:type	dbpedia:ontology/Person
rdf:type	dbpedia:ontology/Agent
rdf:type	dbpedia:ontology/Scientist
rdf:type	<http://schema.org/Person>
owl:sameAs	<http://de.dbpedia.org/resource/August_von_Hallerstein>
owl:sameAs	<http://rdf.freebase.com/ns/m.0cmbkz>
rdfs:label	"August von Hallerstein"@de
rdfs:label	"Ferdinand Augustin Hallerstein"@en
rdfs:label	"Augustin Haller von Hallerstein"@no
rdfs:comment	"August(in) Ferdinand Haller von Hallerstein (chinesisch 劉松齡 / 刘松齡 Liú Sōnglíng; * 27. August 1703 in Laibach; † 29. Oktober 1774 in Peking war ein deutsch-österreichischer Jesuit und Missionar."@de
rdfs:comment	"Ferdinand Augustin Haller von Hallerstein (Slovene: Ferdinand Avguštin Haller von Hallerstein, also known as August Allerstein, simplified Chinese: 刘松龄; traditional Chinese: 劉松齡; pinyin: LiuSongling, Jesuit missionary in China and Chinese astronomer from Carniola, who spent 35 years at the Emperor Qianlong court as the Head of the Imperial Astronomical Bureau and Board of Mathematics and created an armillary sphere with rotating rings at the Beijing observatory; the Chinese first demographer who precisely calculated the exact number of Chinese population of the time . "@en
rdfs:comment	"Augustin Haller von Hallerstein, født 18. august 1703 i Ljubljana i det som i dag er Slovenia med da var Laibach i Østerrike, død 29. oktober 1774 i Beijing i Kina) var en tyskættet jesuitt som ble sendt som misjonær til Kina og ble mandarin og president for det astronomisk-matematiske byrå ved det keiserlige hoff."@no
dbpedia:ontology/deathDate	"1774-10-29"^^xsd:date
dbpedia:ontology/birthPlace	:Ljubljana
dbpedia:ontology/birthPlace	:Slovenia
dbpedia:ontology/birthPlace	:Carniola

DBpedia: stvar, imenovana Ferdinand Hallerstein

dbpedia:ontology/birthDate	"1703-08-27"^^xsd:date
dbpedia:ontology/birthDate	"1703-09-27"^^xsd:date
foaf:givenName	"Ferdinand Augustin"@en
foaf:name	"Ferdinand Augustin Hallerstein"@en
foaf:name	Ferdinand Augustin Haller von Hallerstein, or Liu Songling"@en
dbpedia2:birthPlace	"Ljubljana, Carniola"@en
<http://purl.org/dc/terms/subject>	:Category:Slovenian_Jesuits
<http://purl.org/dc/terms/subject>	:Category:1774_deaths
<http://purl.org/dc/terms/subject>	:Category:1703_births
<http://purl.org/dc/terms/subject>	:Category:Slovenian_mathematicians
<http://purl.org/dc/terms/subject>	:Category:Slovenian_astronomers
<http://purl.org/dc/terms/subject>	:Category:Slovenian_Christian_missionaries

❖ Click to edit Master text styles

❖ Second level

❖ Third level

❖ Fourth level

❖ Fifth level