

Semantic-Web und Wikidata

Konzepte formaler Beschreibungs- und Abfragesprachen

Immanuel Normann

immanuel.normann@uni-muenster.de

Arbeitskreis Digital Humanities
Münster – 1.3.2019

POP



ART

**by ANDY
WARHOL**

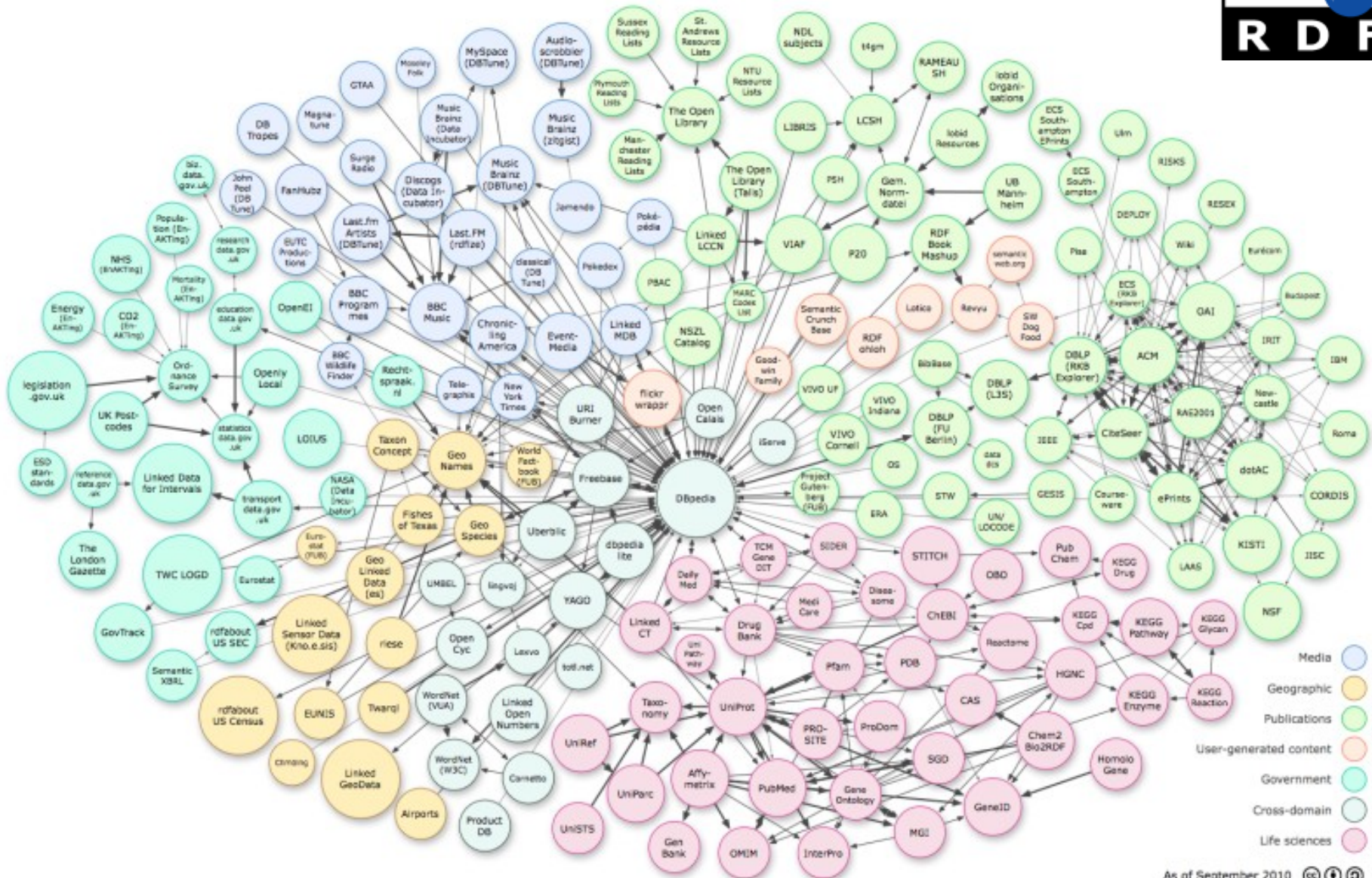
POP

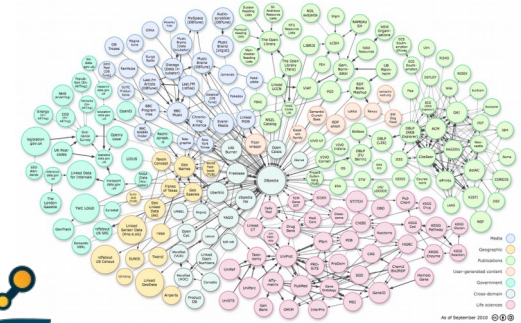


ART



Semantic
Web





WIKIPEDIA

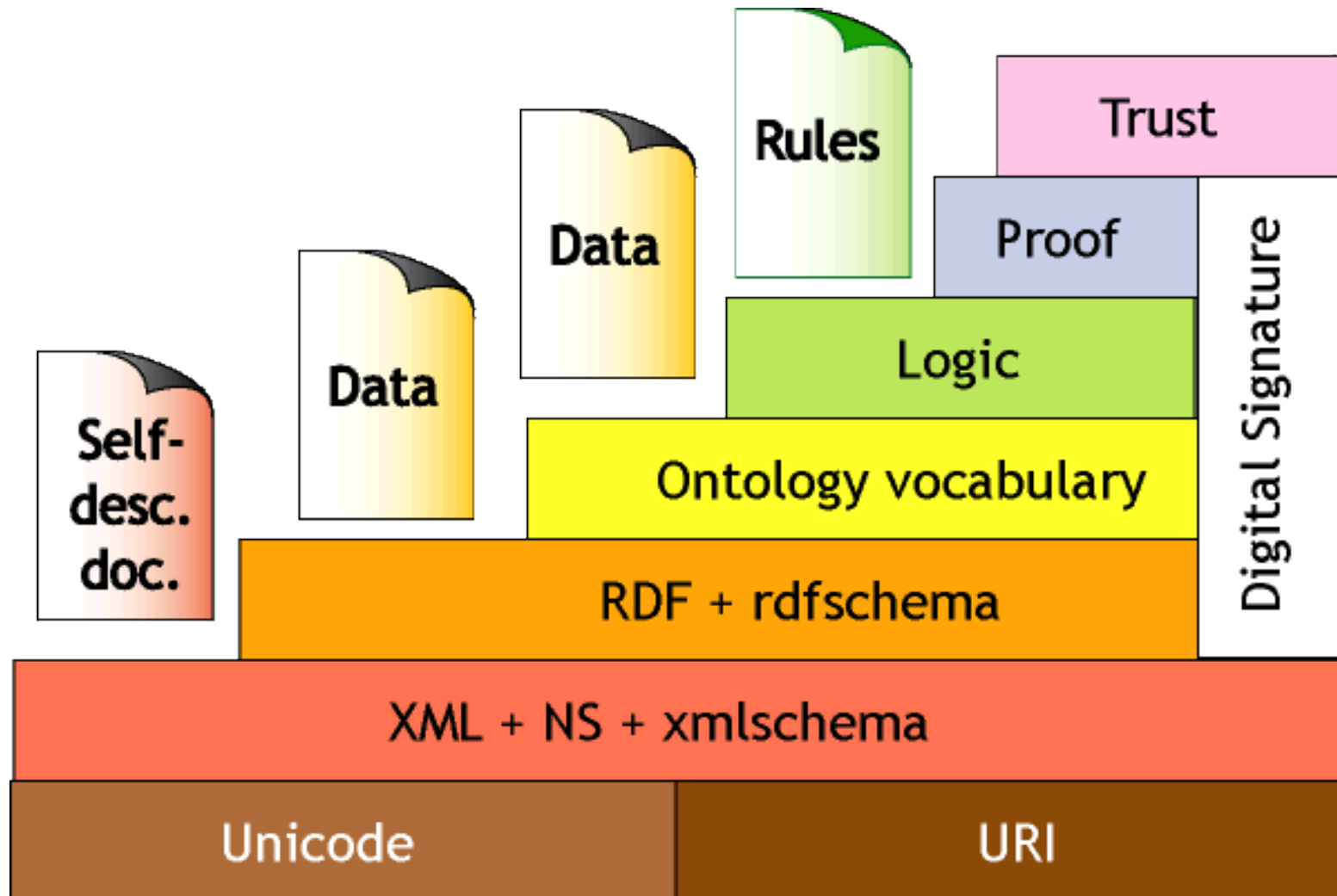


<HTML>

Web 1.0

Web 2.0

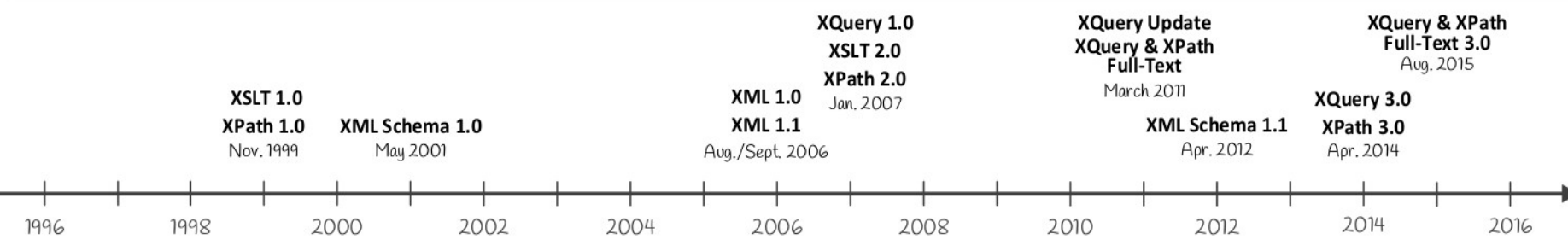
Web 3.0



XML World

Introduced

Standardized



Semantic Web

Introduced

Standardized



Beschreibungssprache (RDF/XML) für ein semantisches Netz

```
<!DOCTYPE rdf:RDF [  
  <ENTITY xsd "http://www.w3.org/2001/XMLSchema#">  
  <ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#">  
  <ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#">  
  <ENTITY owl "http://www.w3.org/2002/07/owl#">  
  <ENTITY wikibase "http://wikiba.se/ontology#">  
>  
  
<rdf:RDF xmlns:xsd="&xsd;" xmlns:rdf="&rdf;" xmlns:rdfs="&rdfs;" xmlns:owl="&owl;">  
  
  <owl:Ontology rdf:about="&wikibase;">  
    <rdfs:label>Wikibase system ontology</rdfs:label>  
    <rdfs:comment>The system ontology of Wikibase</rdfs:comment>  
    <rdfs:seeAlso rdf:resource="https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format" />  
    <rdfs:seeAlso rdf:resource="https://www.mediawiki.org/wiki/Extension:WikibaseLexeme/RDF_mapping" />  
  </owl:Ontology>  
  
  <owl:Class rdf:about="&wikibase;Dump">  
    <rdfs:label>Dump</rdfs:label>  
    <rdfs:comment>A dump of wikidata content.</rdfs:comment>  
  </owl:Class>  
  
  <owl:Class rdf:about="&wikibase;Entity">  
    <rdfs:label>Entity</rdfs:label>  
    <rdfs:comment>Wikibase entity.</rdfs:comment>  
  </owl:Class>  
  
  <owl:Class rdf:about="&wikibase;Item">  
    <rdfs:label>Item</rdfs:label>  
    <rdfs:comment>Wikibase item.</rdfs:comment>  
    <rdfs:subClassOf rdf:resource="&wikibase;Entity"/>  
  </owl:Class>  
  
  <owl:Class rdf:about="&wikibase;Property">  
    <rdfs:label>Property</rdfs:label>  
    <rdfs:comment>Wikibase property.</rdfs:comment>  
    <rdfs:subClassOf rdf:resource="&wikibase;Entity"/>  
  </owl:Class>  
  
  <owl:Class rdf:about="&wikibase;Statement">  
    <rdfs:label>Statement</rdfs:label>  
    <rdfs:comment>A reified statement.</rdfs:comment>  
  </owl:Class>
```


Beschreibungssprache (TTL) für ein semantisches Netz

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
```

```
<http://wikiba.se/ontology#>  
  a owl:Ontology ;  
  rdfs:label "Wikibase system ontology" ;  
  rdfs:comment "The system ontology of Wikibase" ;  
  rdfs:seeAlso <https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format>, <https://www.mediawiki.org/wiki/Extension:WikibaseLexeme/  
RDF_mapping> .
```

```
<http://wikiba.se/ontology#Dump>  
  a owl:Class ;  
  rdfs:label "Dump" ;  
  rdfs:comment "A dump of wikidata content." .
```

```
<http://wikiba.se/ontology#Entity>  
  a owl:Class ;  
  rdfs:label "Entity" ;  
  rdfs:comment "Wikibase entity." .
```

```
<http://wikiba.se/ontology#Item>  
  a owl:Class ;  
  rdfs:label "Item" ;  
  rdfs:comment "Wikibase item." ;  
  rdfs:subClassOf <http://wikiba.se/ontology#Entity> .
```

```
<http://wikiba.se/ontology#Property>  
  a owl:Class ;  
  rdfs:label "Property" ;  
  rdfs:comment "Wikibase property." ;  
  rdfs:subClassOf <http://wikiba.se/ontology#Entity> .
```

```
<http://wikiba.se/ontology#Statement>  
  a owl:Class ;  
  rdfs:label "Statement" ;  
  rdfs:comment "A reified statement." .
```

Anfragesprache (SPARQL) für ein semantisches Netz

Battles per year per continent and country last 80 years (animated)

```
SELECT (SAMPLE(?_continentLabel) AS ?continent) (COUNT(?_country) AS ?battles) (SAMPLE(?_countryLabel)
AS ?_countryLabel) (SAMPLE(?year) AS ?year) WHERE {
  ?subj wdt:P31 wd:Q178561.      # instance of battle
  OPTIONAL { ?subj wdt:P580 ?d1. } # start date
  OPTIONAL { ?subj wdt:P585 ?d2. } # point in time
  OPTIONAL { ?subj wdt:P582 ?d3. } # end date
  BIND(IF(!BOUND(?d1), IF(!BOUND(?d2), ?d3, ?d2), ?d1) AS ?date)
  BIND(str(YEAR(?date)) AS ?year)
  FILTER(BOUND(?year))
  ?subj wdt:P276/wdt:P17 ?_country.
  ?_country wdt:P30 ?_continent.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
    ?_country rdfs:label ?_countryLabel.
    ?_continent rdfs:label ?_continentLabel.}
  BIND((NOW()) - ?date AS ?distance)
  FILTER((0 <= ?distance) && (?distance < 31 * 12 * 80))
}
GROUP BY ?year ?_country
```

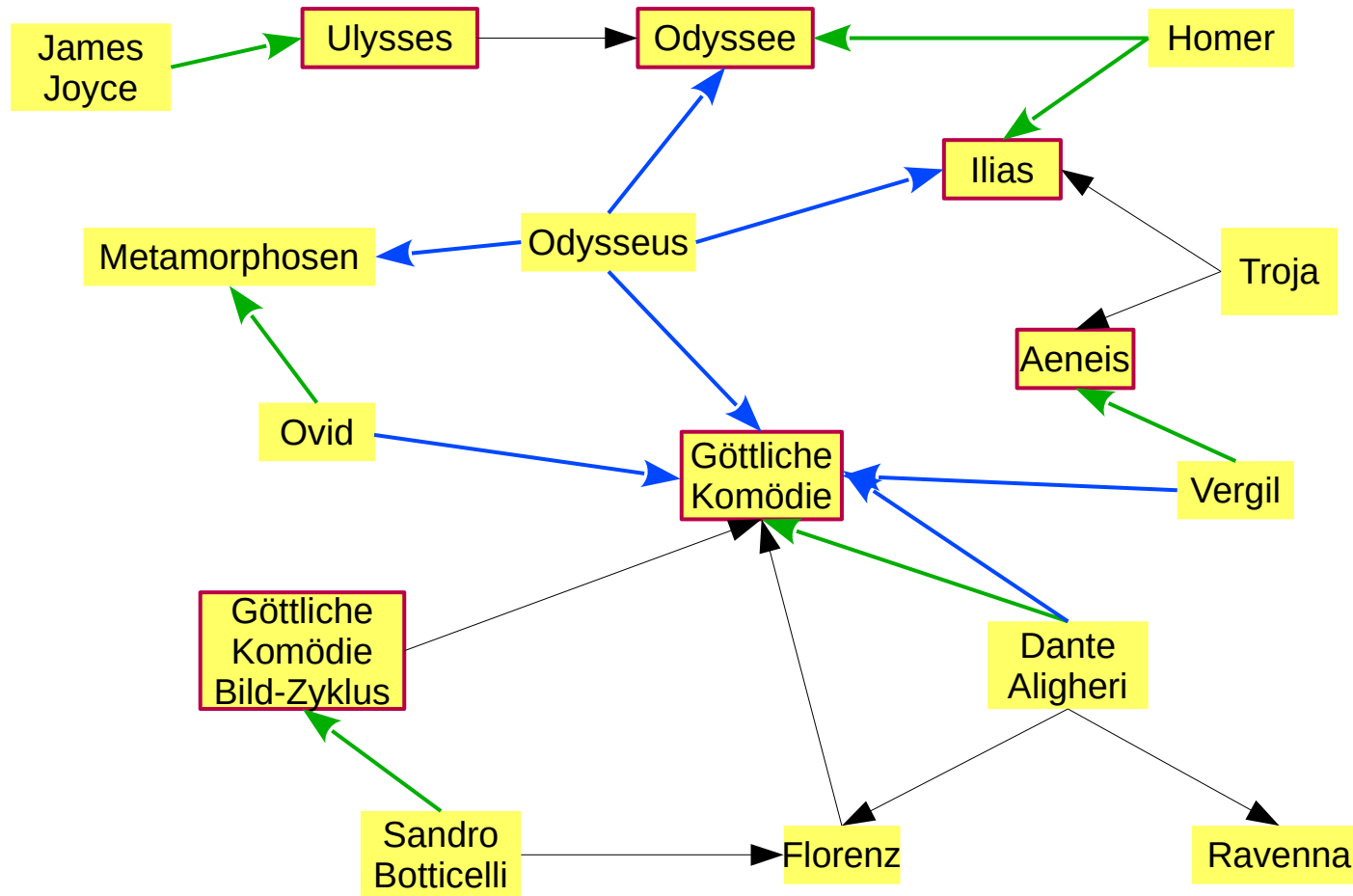
Anfragesprache (SPARQL)


aus der Sicht des Computers

```

00000000: 01010011 01000101 01001100 01000101 01000011 01010100 SELECT
00000006: 00100000 00101000 01010011 01000001 01001101 01010000 (SAMP
0000000c: 01001100 01000101 00101000 00111111 01011111 01100011 LE(?_c
00000012: 01101111 01101110 01110100 01101001 01101110 01100101 ontine
00000018: 01101110 01110100 01001100 01100001 01100010 01100101 ntLabe
0000001e: 01101100 00101001 00100000 01000001 01010011 00100000 l) AS
00000024: 00111111 01100011 01101111 01101110 01110100 01101001 ?conti
0000002a: 01101110 01100101 01101110 01110100 00101001 00100000 nent)
00000030: 00101000 01000011 01001111 01010101 01001110 01010100 (COUNT
00000036: 00101000 00111111 01011111 01100011 01101111 01110101 (?_cou
0000003c: 01101110 01110100 01110010 01111001 00101001 00100000 ntry)
00000042: 01000001 01010011 00100000 00111111 01100010 01100001 AS ?ba
00000048: 01110100 01110100 01101100 01100101 01110011 00101001 ttles)
0000004e: 00100000 00101000 01010011 01000001 01001101 01010000 (SAMP
00000054: 01001100 01000101 00101000 00111111 01011111 01100011 LE(?_c
0000005a: 01101111 01110101 01101110 01110100 01110010 01111001 ountry
00000060: 01001100 01100001 01100010 01100101 01101100 00101001 Label)
00000066: 00100000 01000001 01010011 00100000 00111111 01011111 AS ?_
0000006c: 01100011 01101111 01110101 01101110 01110100 01110010 countr
00000072: 01111001 01001100 01100001 01100010 01100101 01101100 yLabel
00000078: 00101001 00100000 00101000 01010011 01000001 01001101 ) (SAM
0000007e: 01010000 01001100 01000101 00101000 00111111 01111001 PLE(?y
00000084: 01100101 01100001 01110010 00101001 00100000 01000001 ear) A
0000008a: 01010011 00100000 00111111 01111001 01100101 01100001 S ?yea
00000090: 01110010 00101001 00100000 00100000 01010111 01001000 r) WH
00000096: 01000101 01010010 01000101 00100000 01111011 00001010 ERE {.
0000009c: 00100000 00100000 00111111 01110011 01110101 01100010 ?sub
  
```

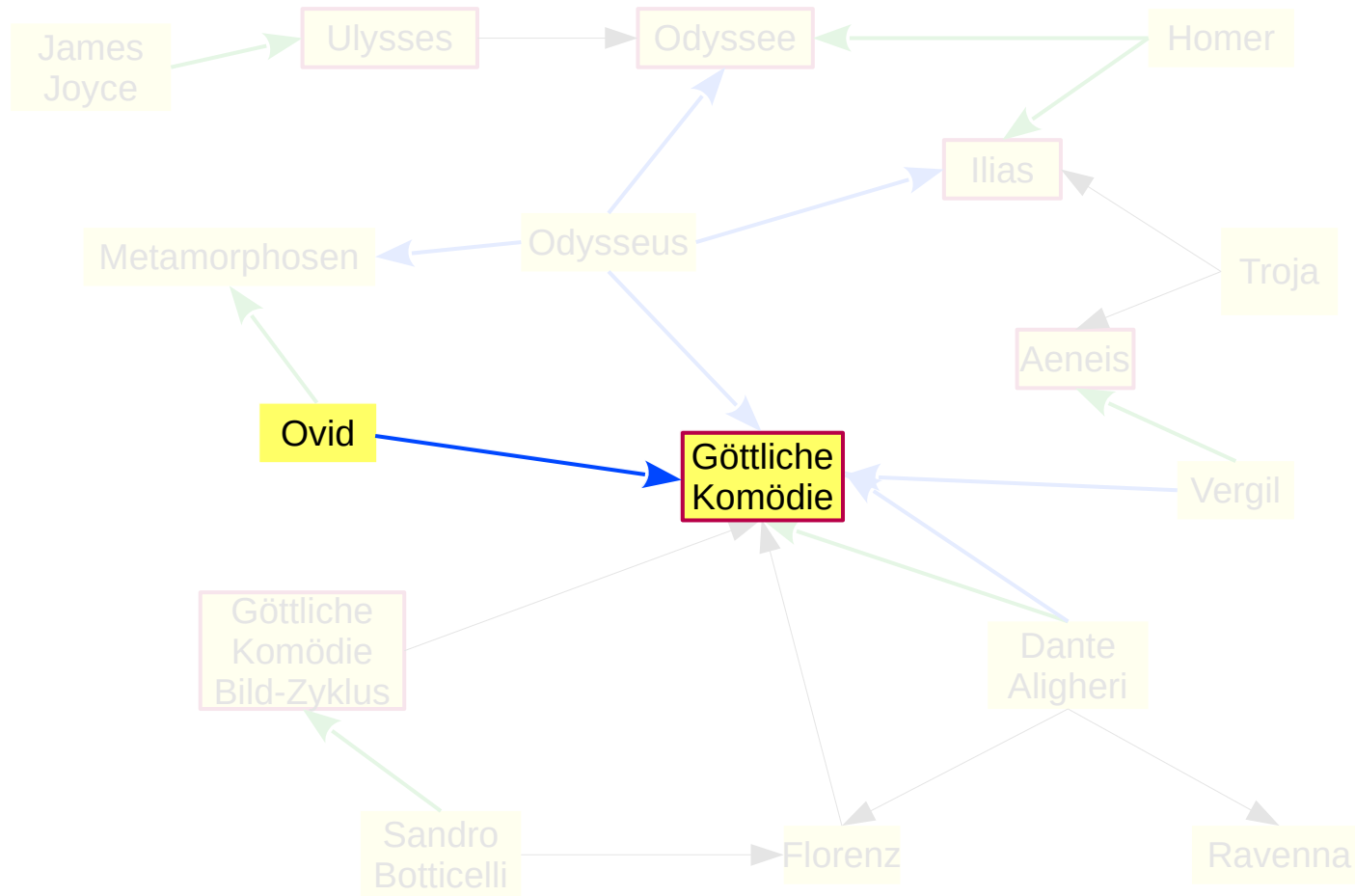
Semantisches Netz



Figur in   *Werk*

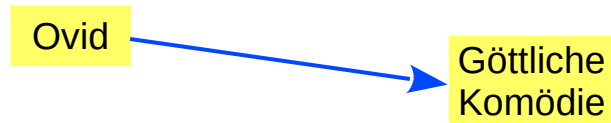
Autor von   *Werk*

Semantisches Netz

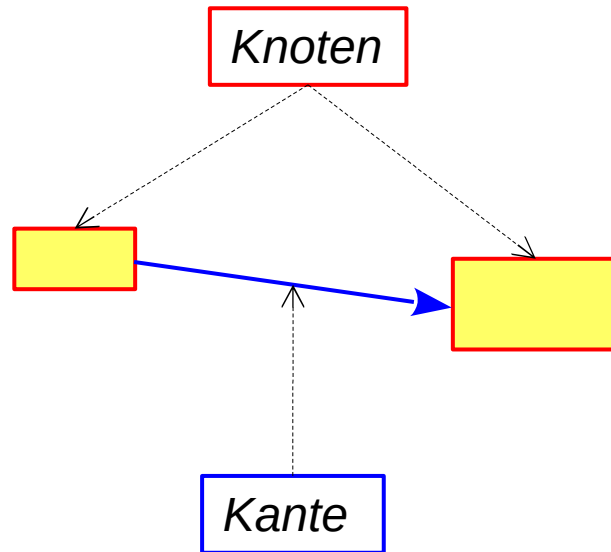


Figur in   *Werk*
 Autor von   *Werk*

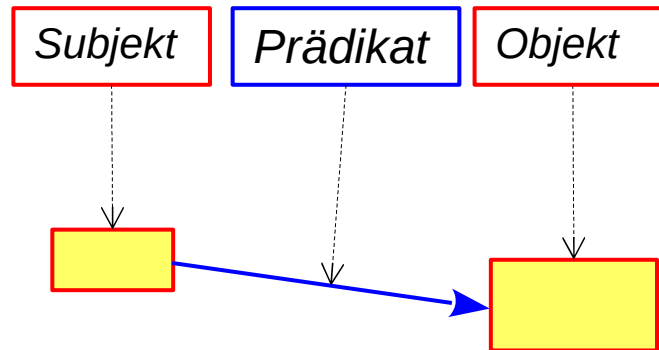
Semantisches Netz



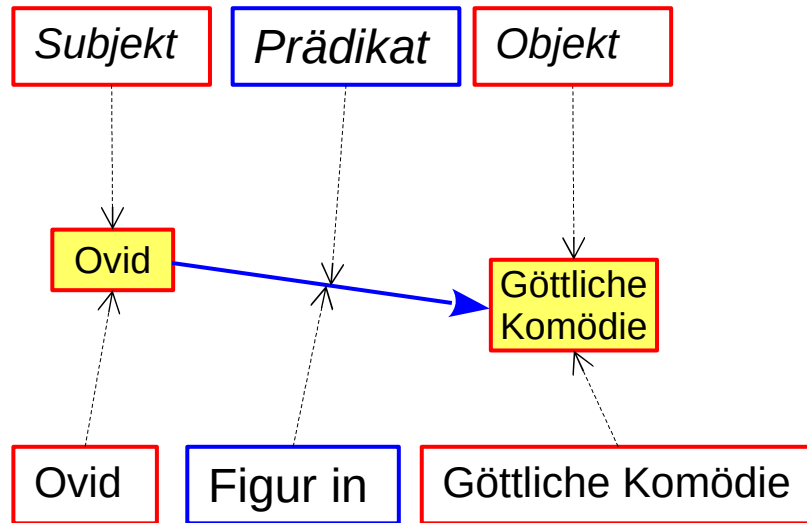
Semantisches Netz



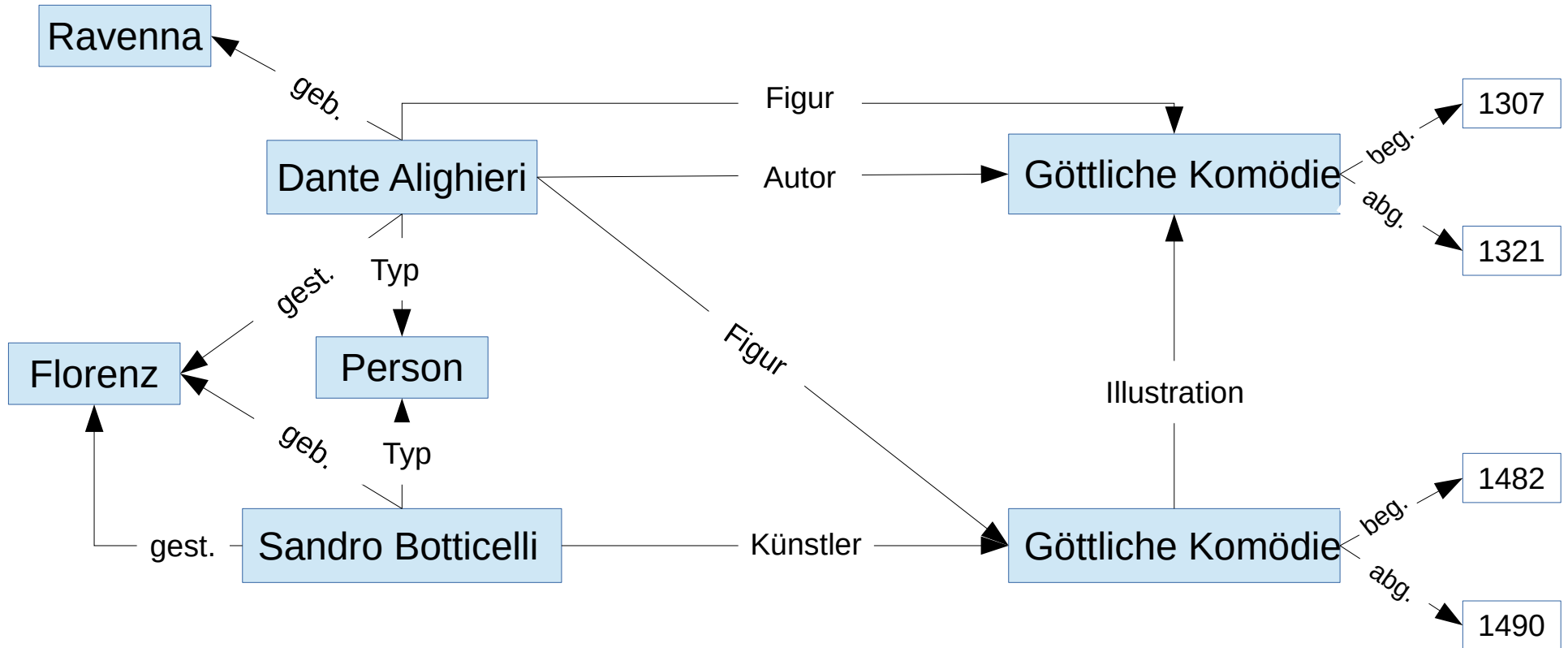
Semantisches Netz



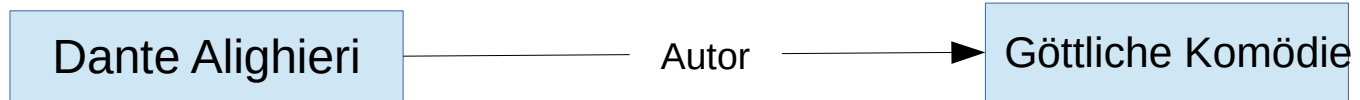
Semantisches Netz



RDF-Graph



RDF-Graph



RDF-Graph

Göttliche Komödie

RDF-Graph

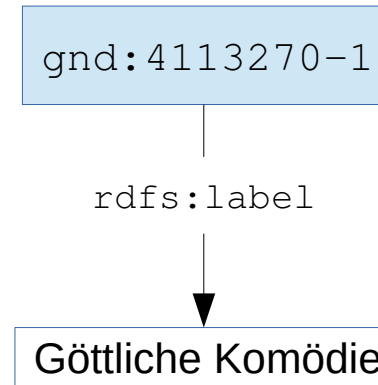
`http://d-nb.info/gnd/4113270-1`

RDF-Graph

gnd:4113270-1

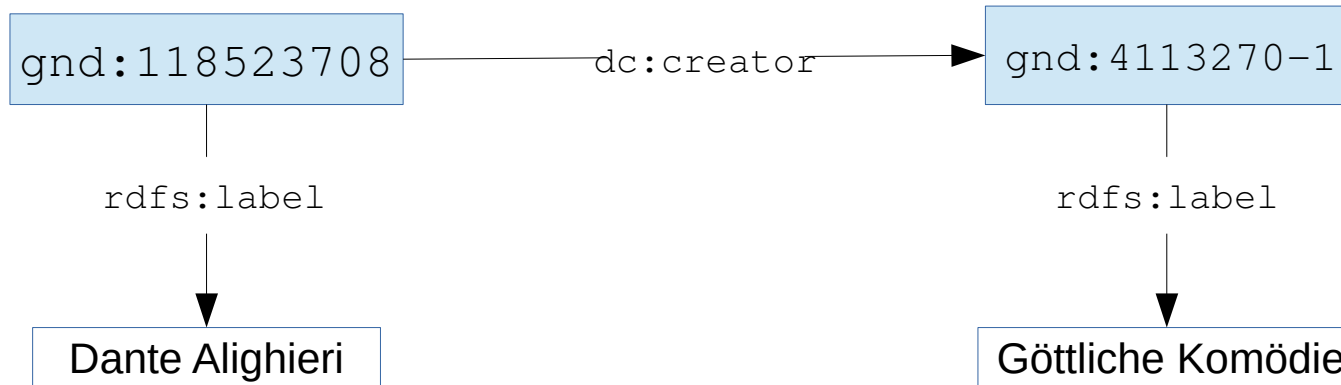
gnd: <http://d-nb.info/gnd/>

RDF-Graph



gnd: <http://d-nb.info/gnd/>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

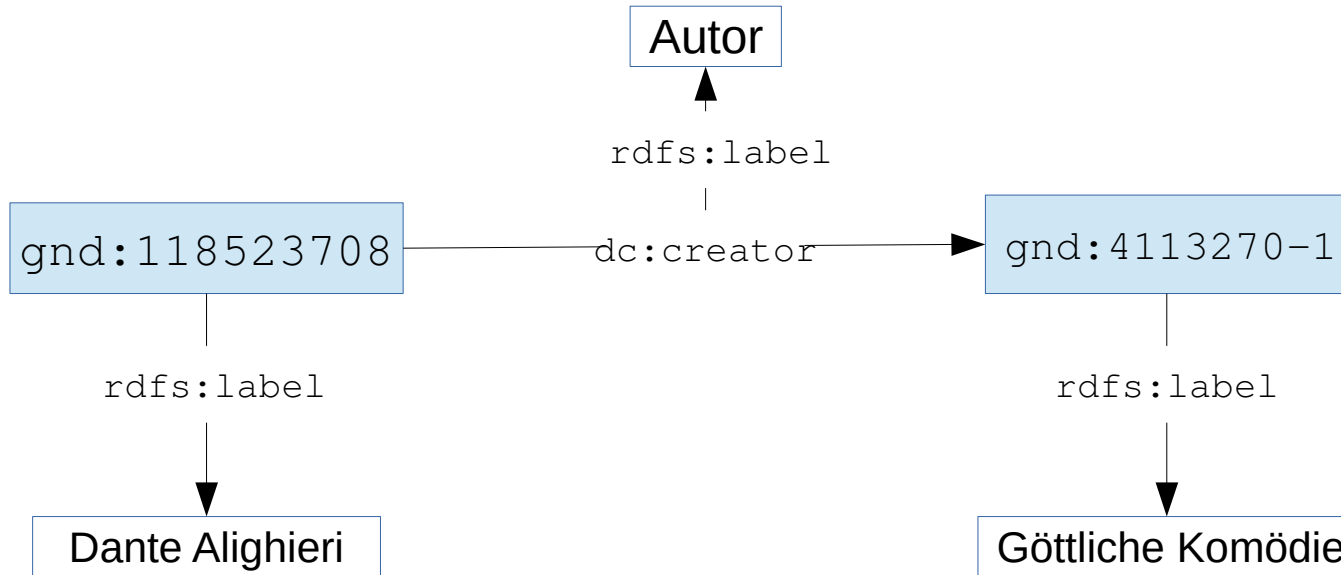


dc: <http://purl.org/dc/terms/>

gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

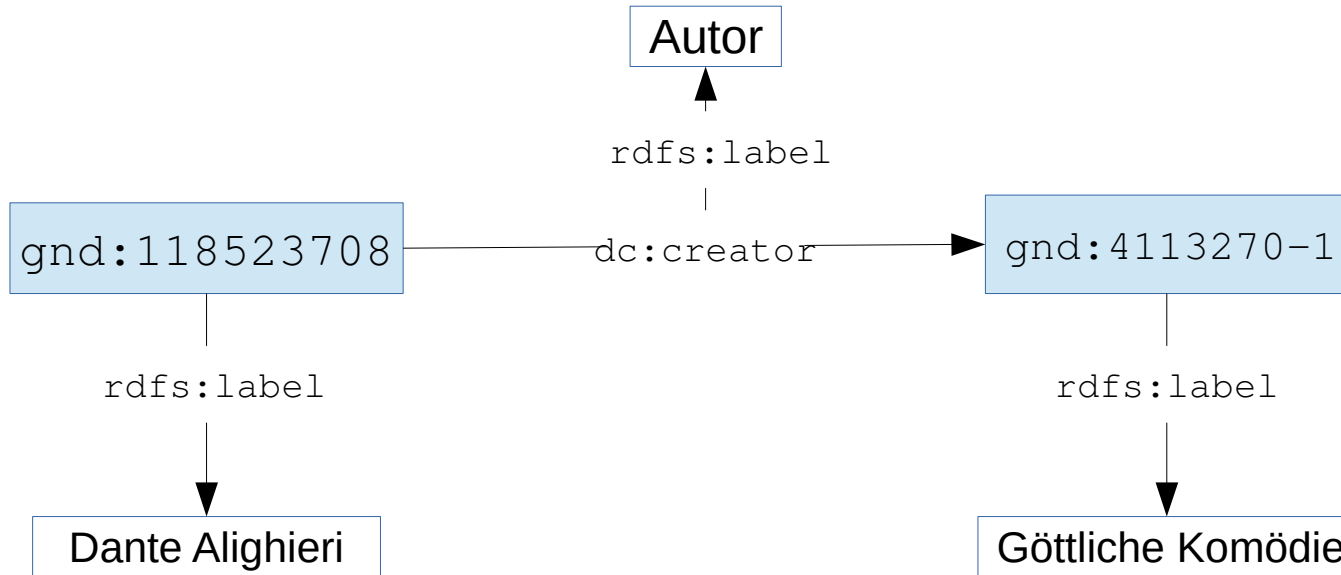


dc: <http://purl.org/dc/terms/>

gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

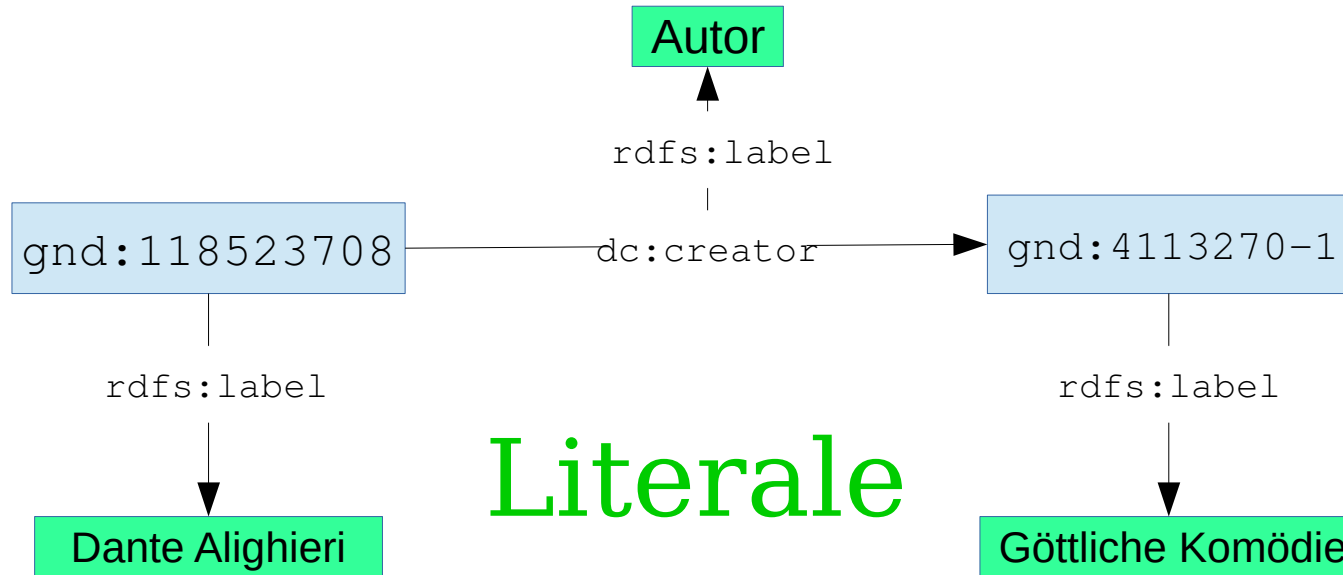


dc: <http://purl.org/dc/terms/>

gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

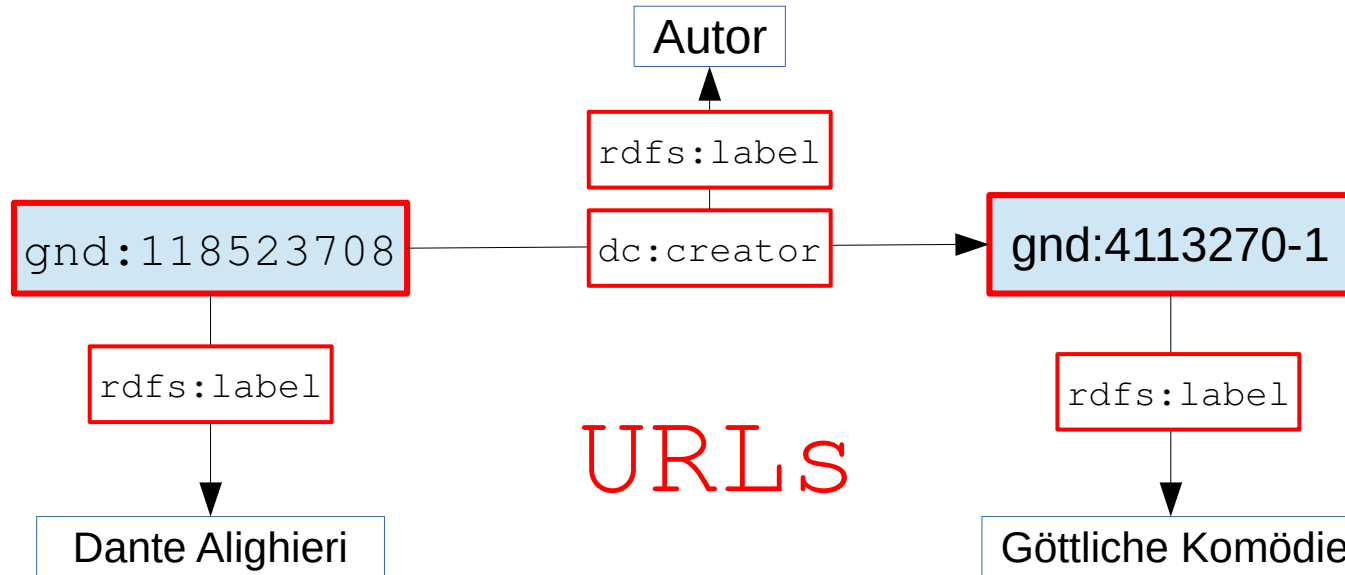


dc: <http://purl.org/dc/terms/>

gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

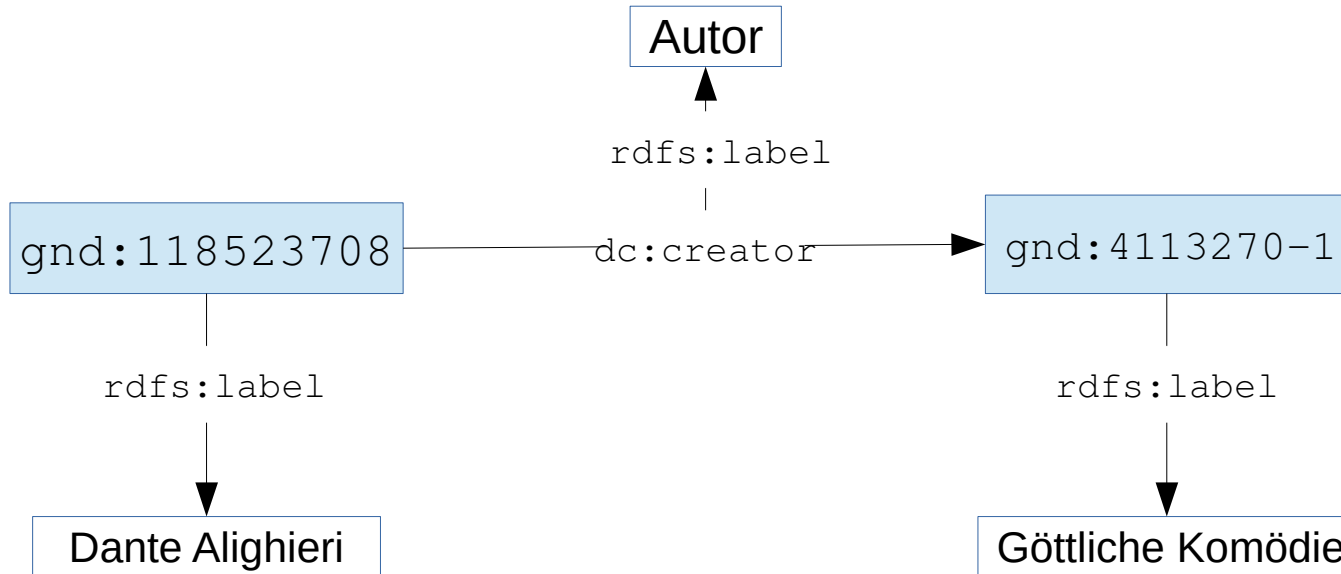


dc: <http://purl.org/dc/terms/>

gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

RDF-Graph

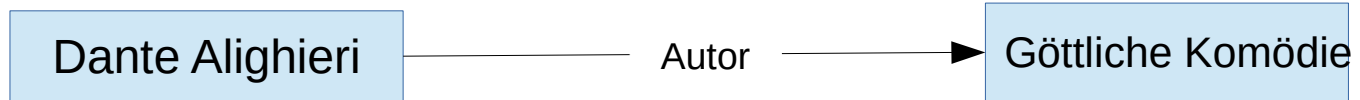


dc: <http://purl.org/dc/terms/>

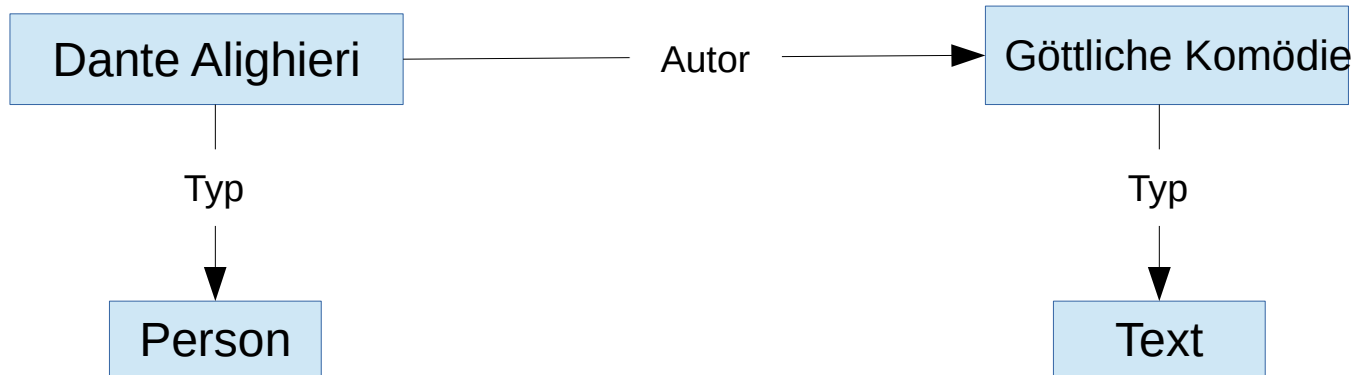
gnd: <http://d-nb.info/gnd/>

rdfs: <http://www.w3.org/2000/01/rdf-schema#>

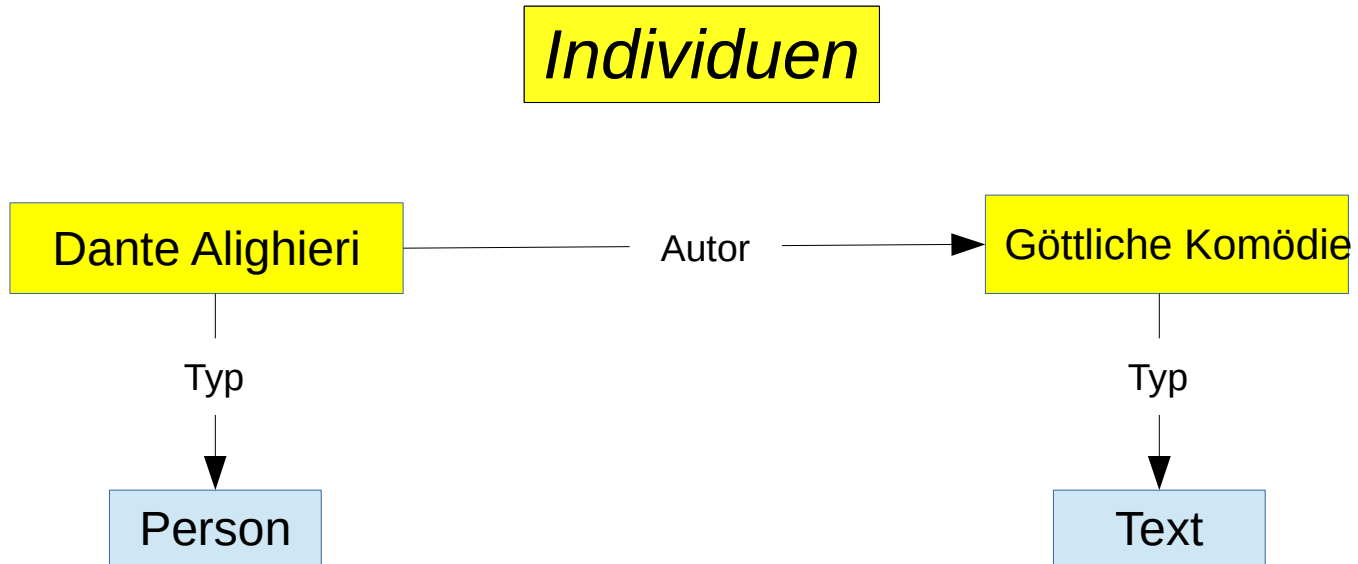
RDF-Graph



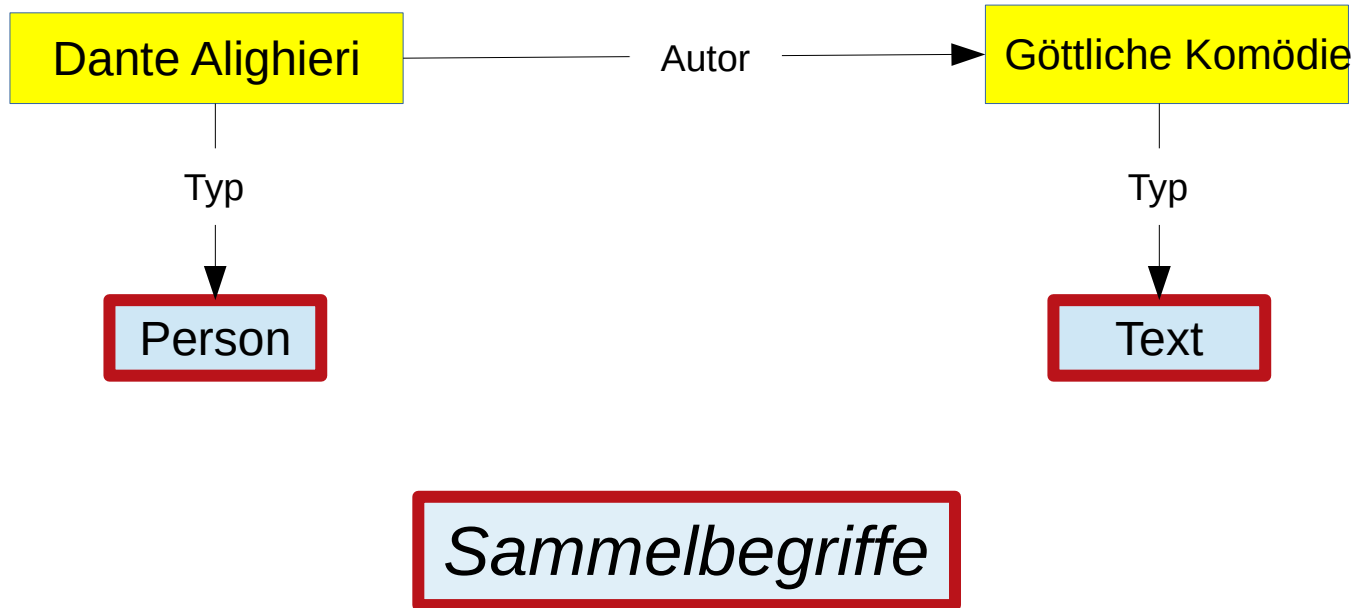
RDF-Schema



RDF-Schema

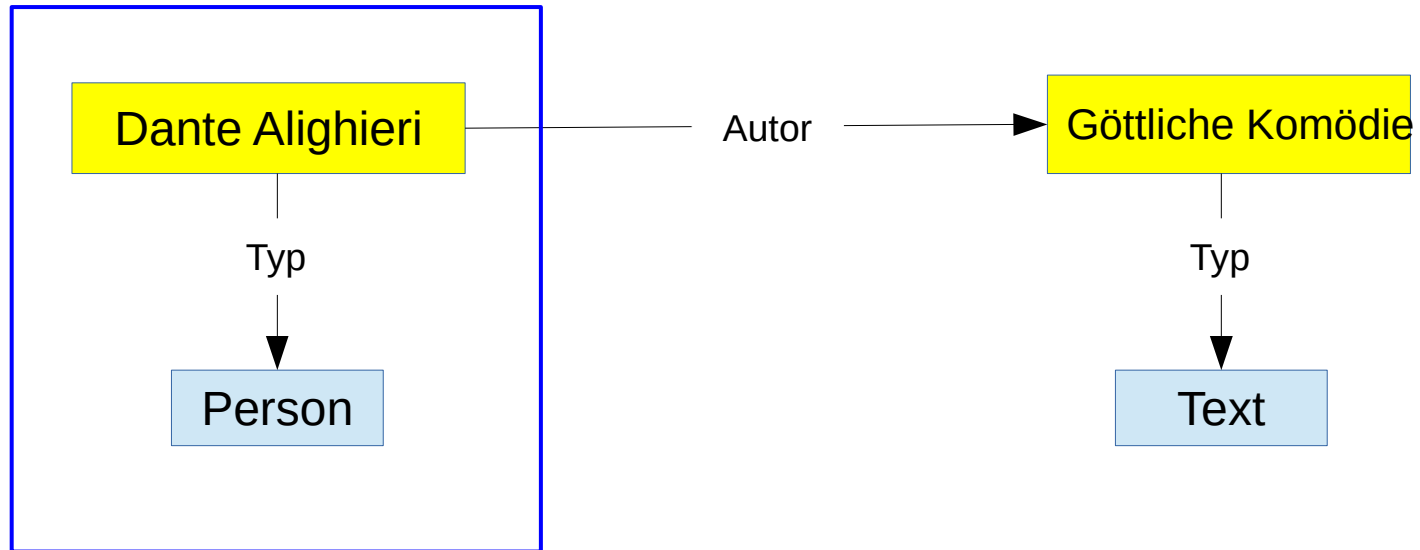


RDF-Schema

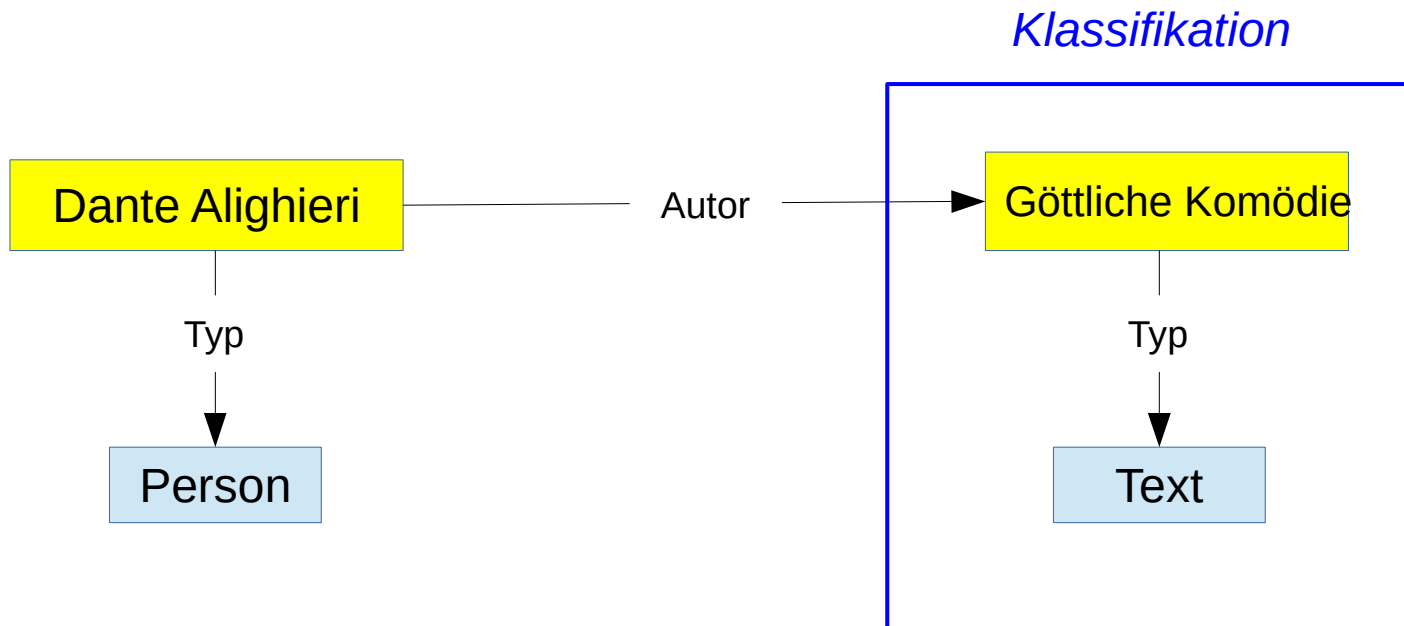


RDF-Schema

Klassifikation



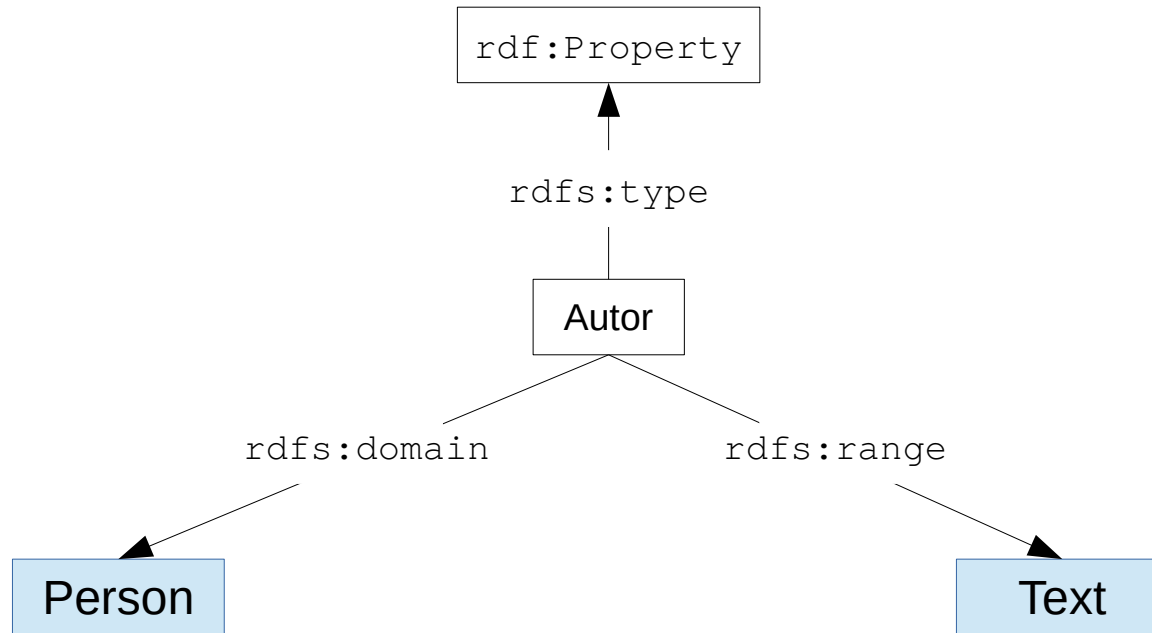
RDF-Schema



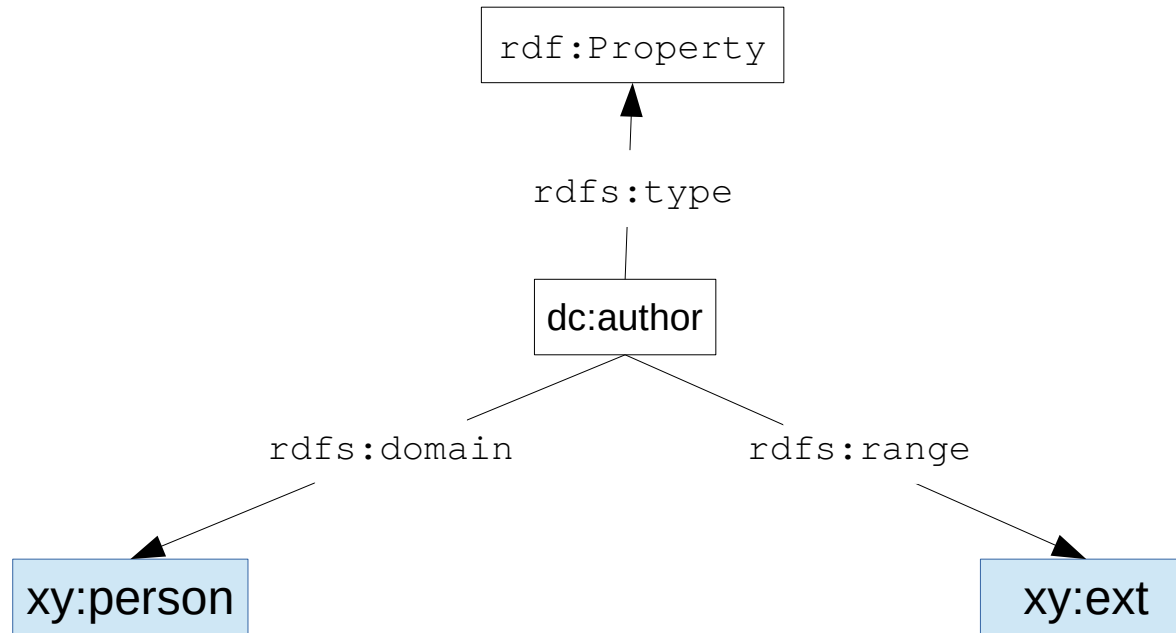
RDF-Schema



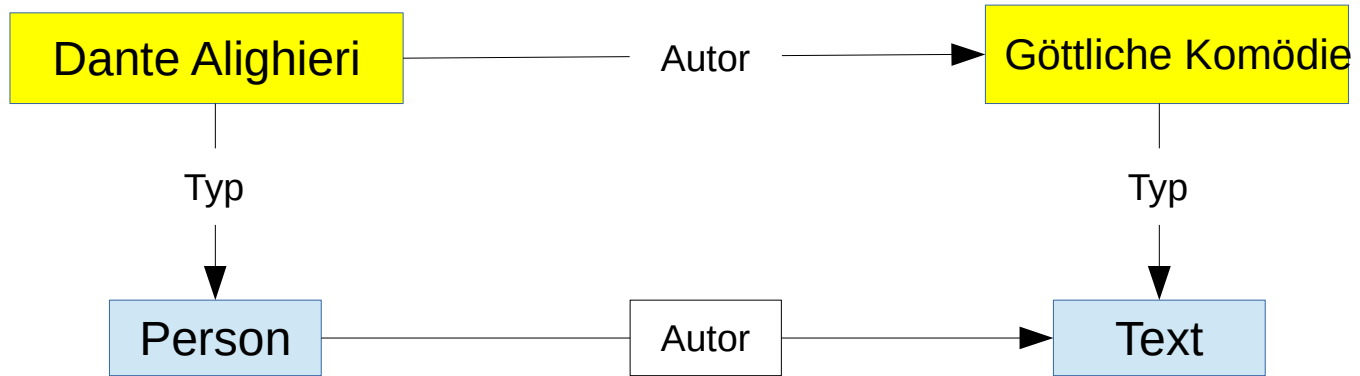
RDF-Schema



RDF-Schema

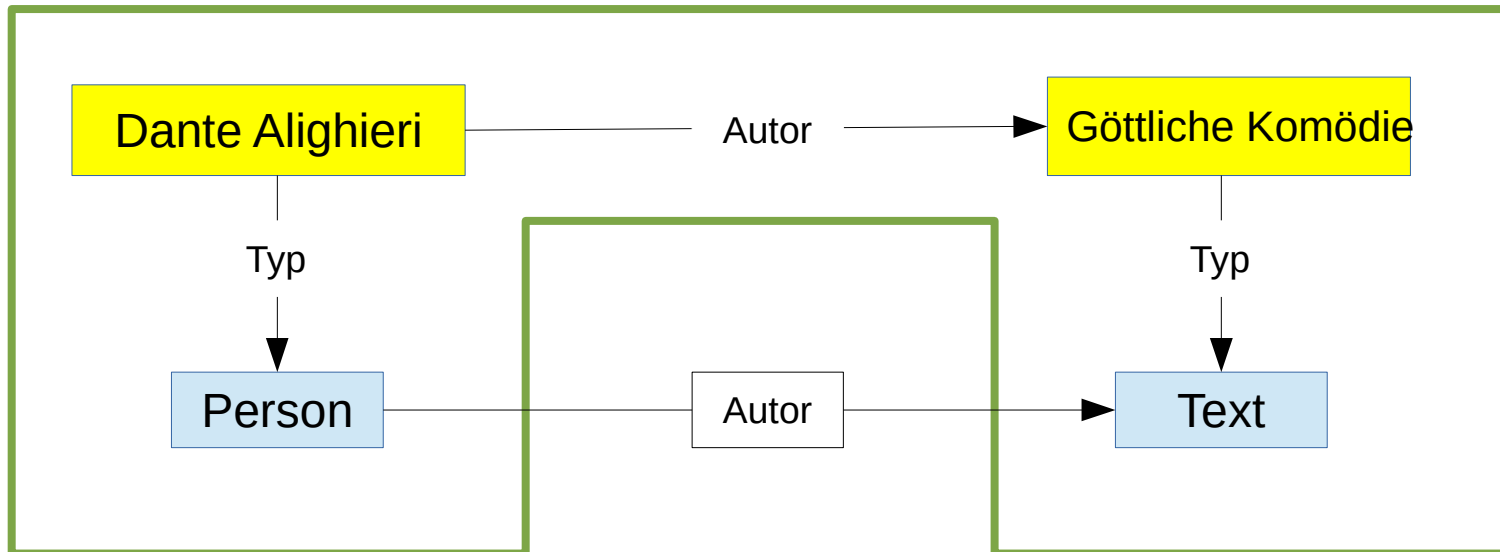


RDF-Schema

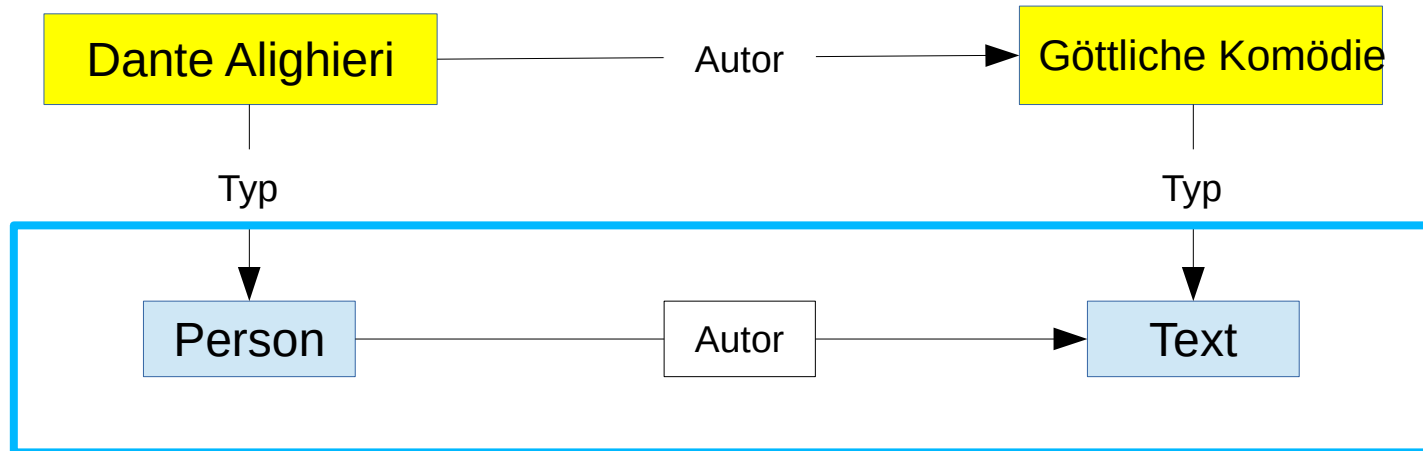


RDF-Schema

Fakten



RDF-Schema



Schema

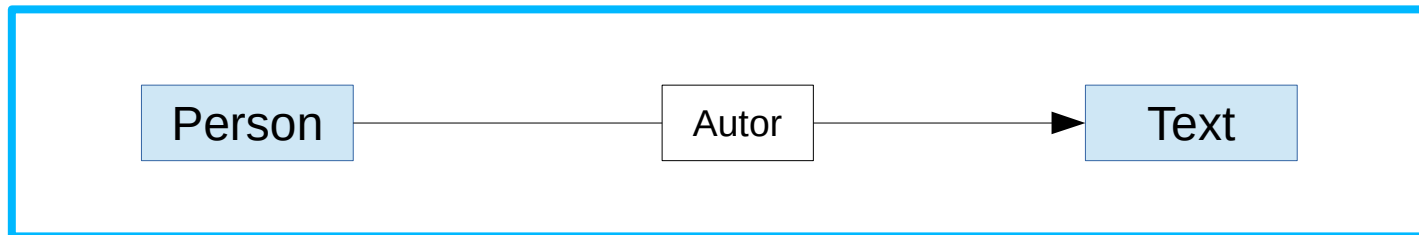
Wissensmodellierung

Modellierung = Schema-Entwurf = Ontologie-Entwurf

Rollen:

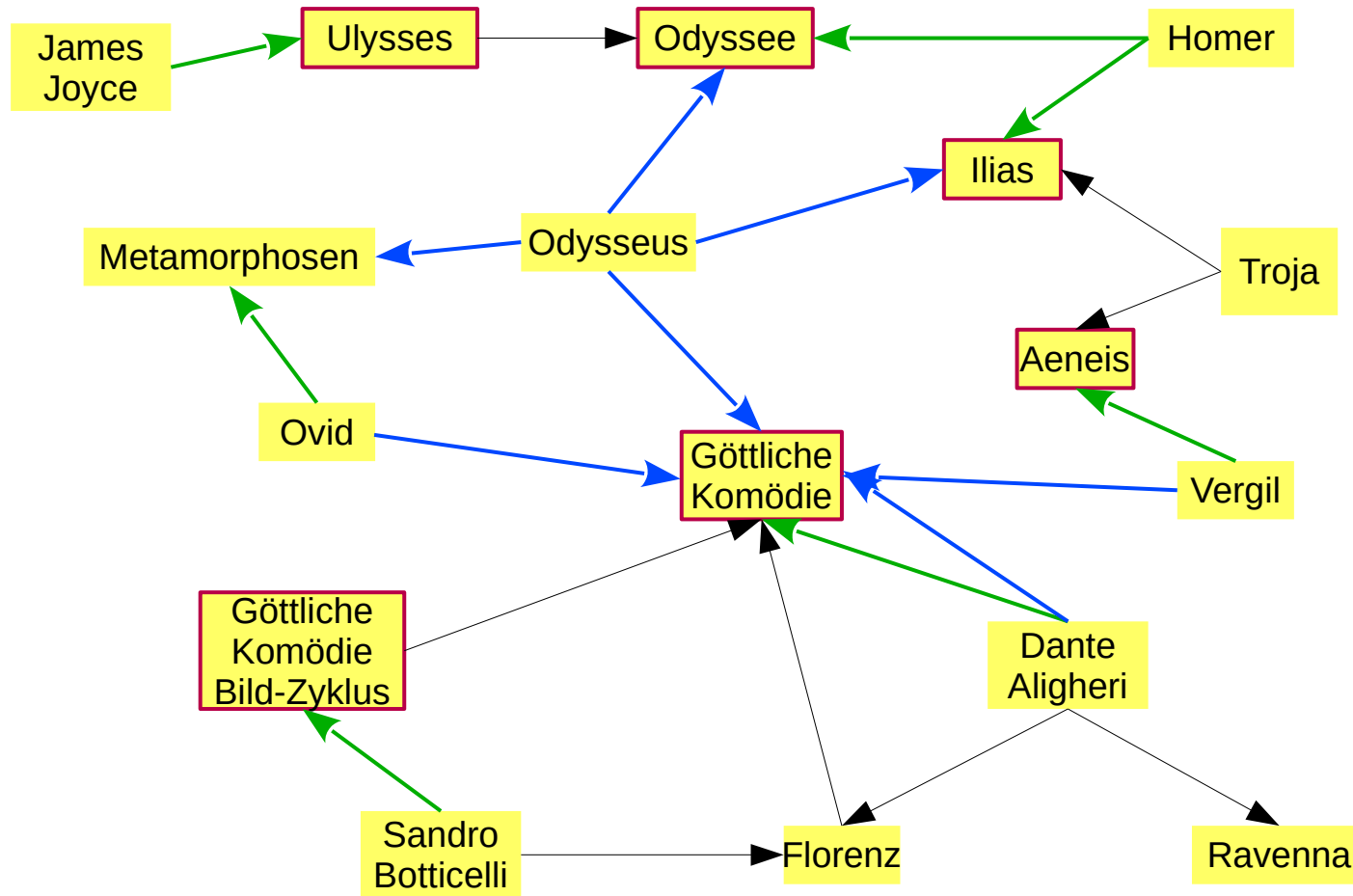
domain expert: Fachliche Expertise

knowledge engineer: Formalisierungsexpertise



Schema

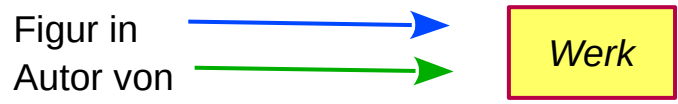
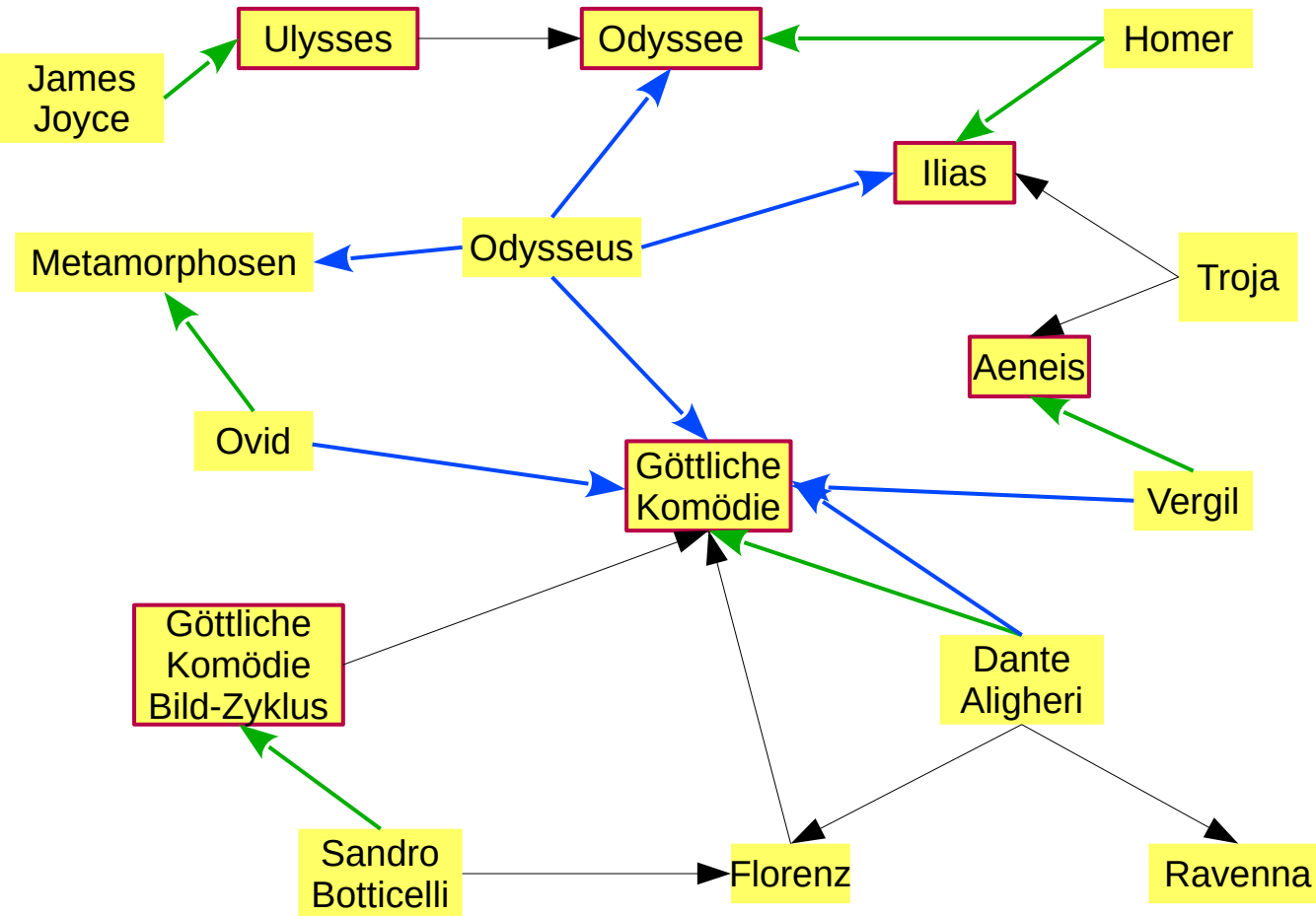
Faktensammlung



Figur in   *Werk*

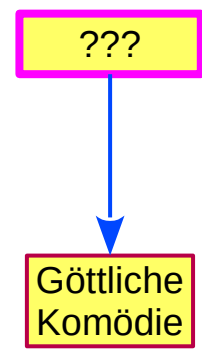
Autor von   *Werk*

Suche



Beispiel 1

Welche Figuren kommen in der Göttlichen Komödie vor ?

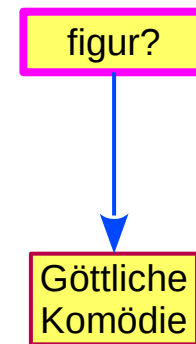


```
SELECT ?figur
```

```
WHERE {  
  ?figur figur_in gnd:4113270-1 .  
}
```

Beispiel 1

Welche Figuren kommen in
der Göttlichen Komödie vor ?



```
SELECT ?figur
```

```
WHERE {
```

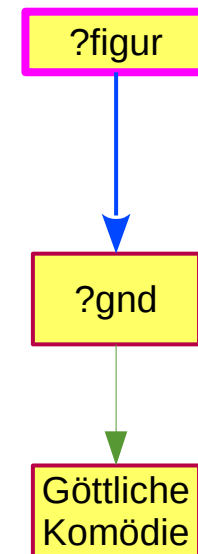
```
  ?figur figur_in ?gnd .
```

```
  ?gnd rdfs:label „Göttliche Komödie“ .
```

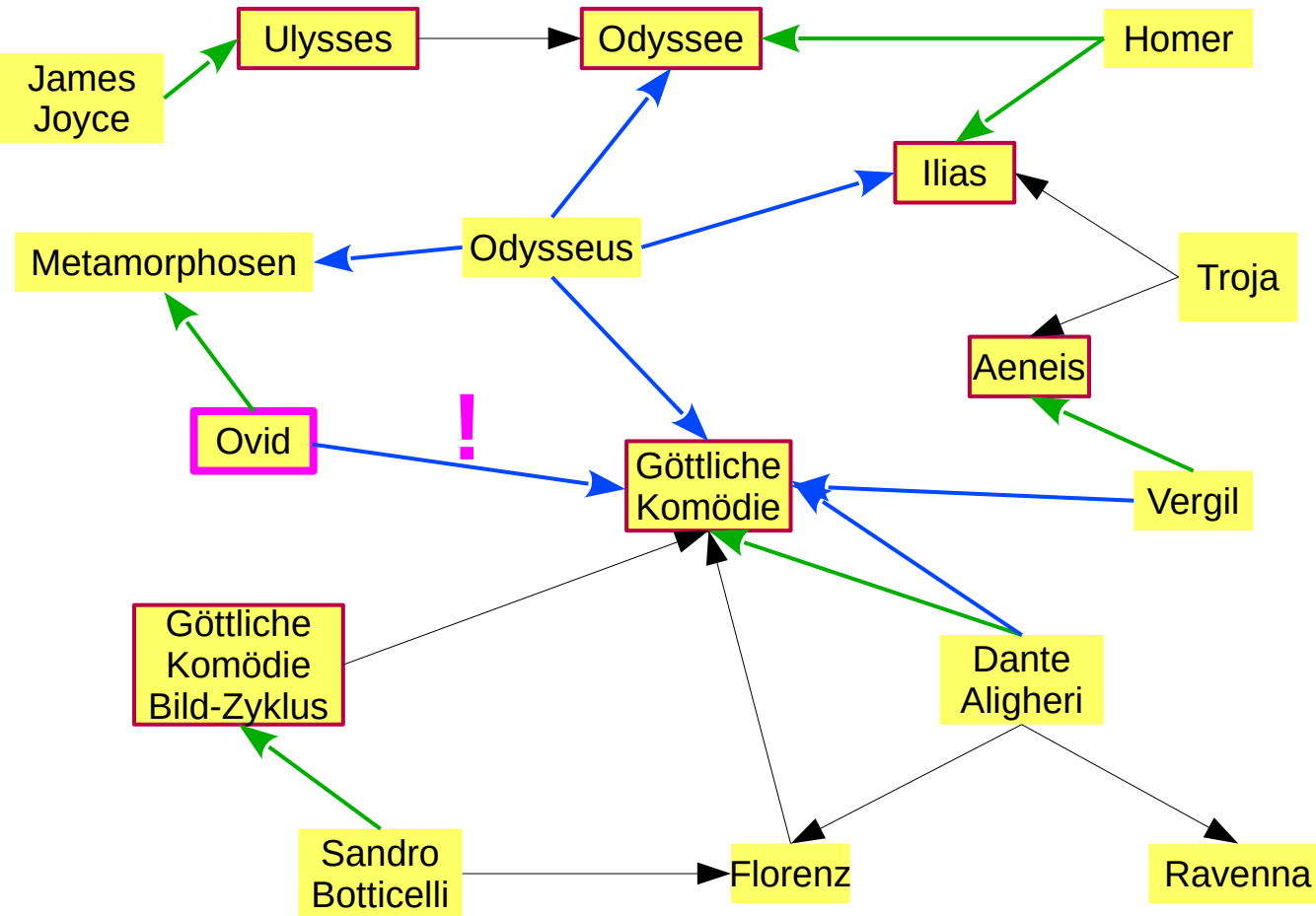
```
}
```

Beispiel 1

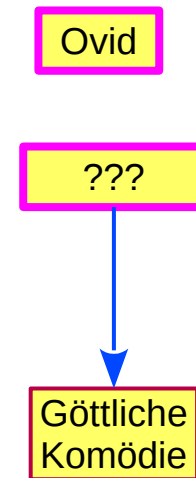
Welche Figuren kommen in der Göttlichen Komödie vor ?





Suche

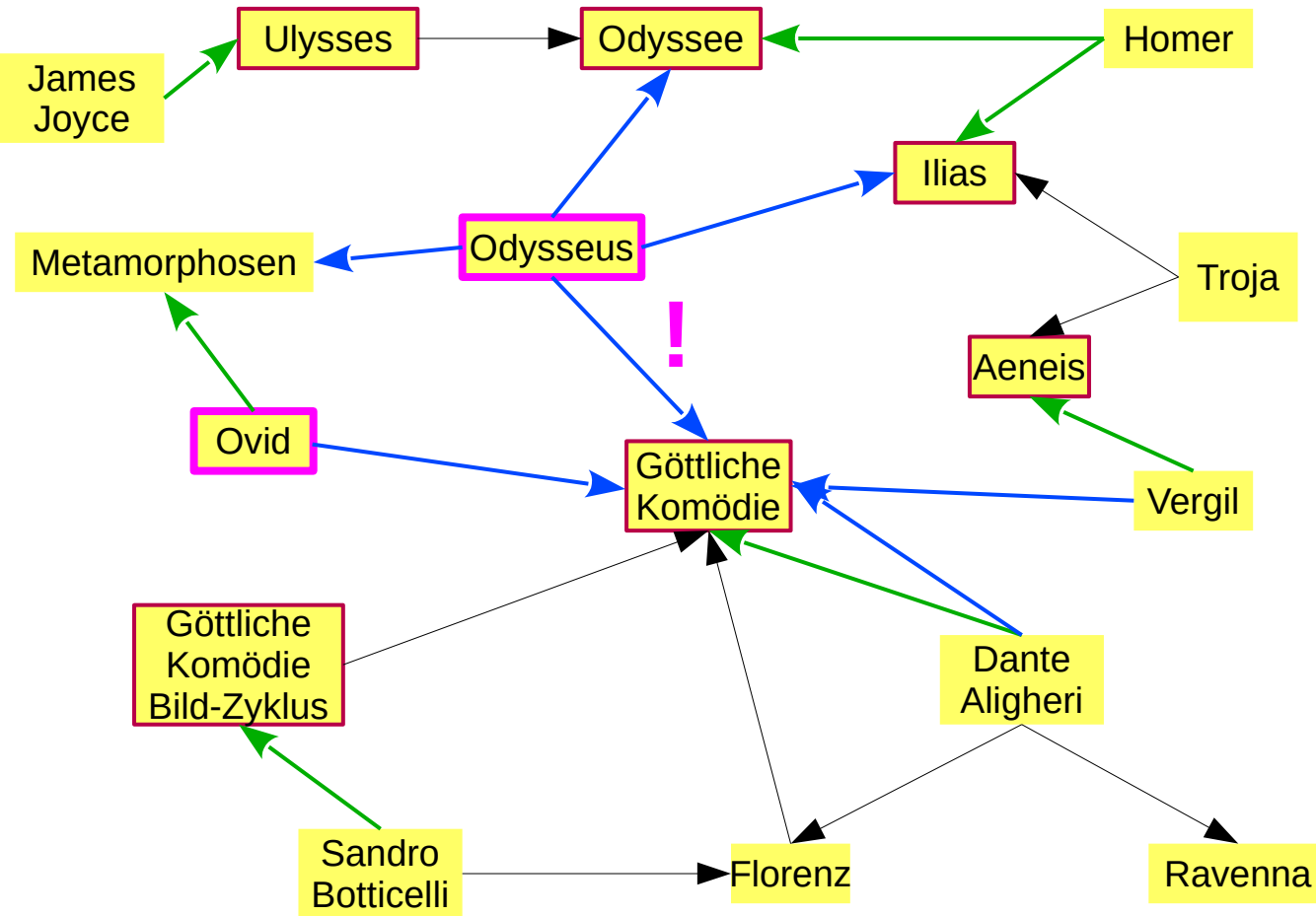


Welche Figuren kommen in der Göttlichen Komödie vor ?

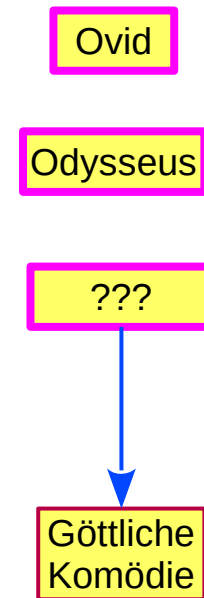





Figur in   *Werk*
 Autor von   *Werk*

Suche

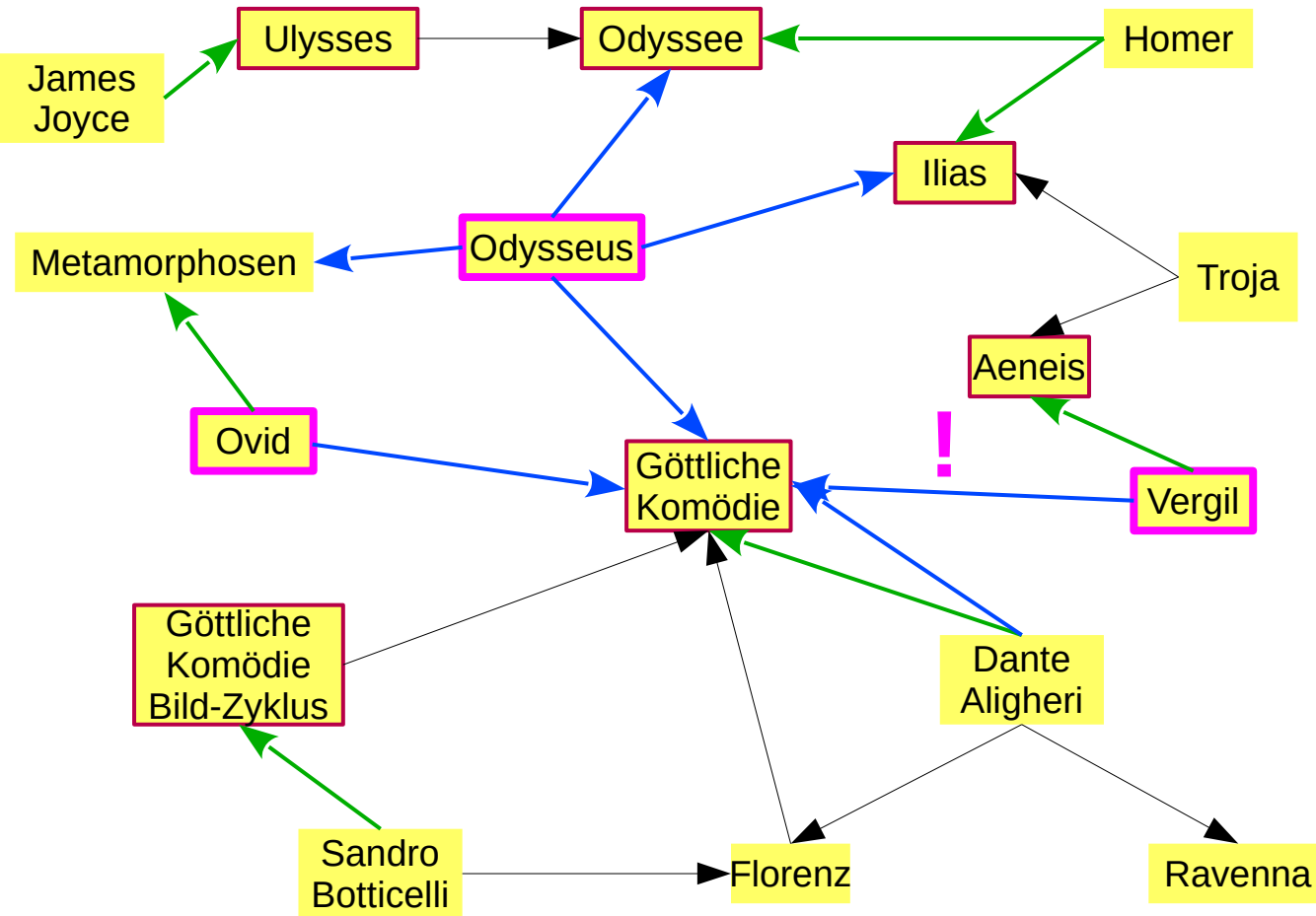


Welche Figuren kommen in der Göttlichen Komödie vor ?







Figur in   *Werk*
 Autor von   *Werk*

Suche

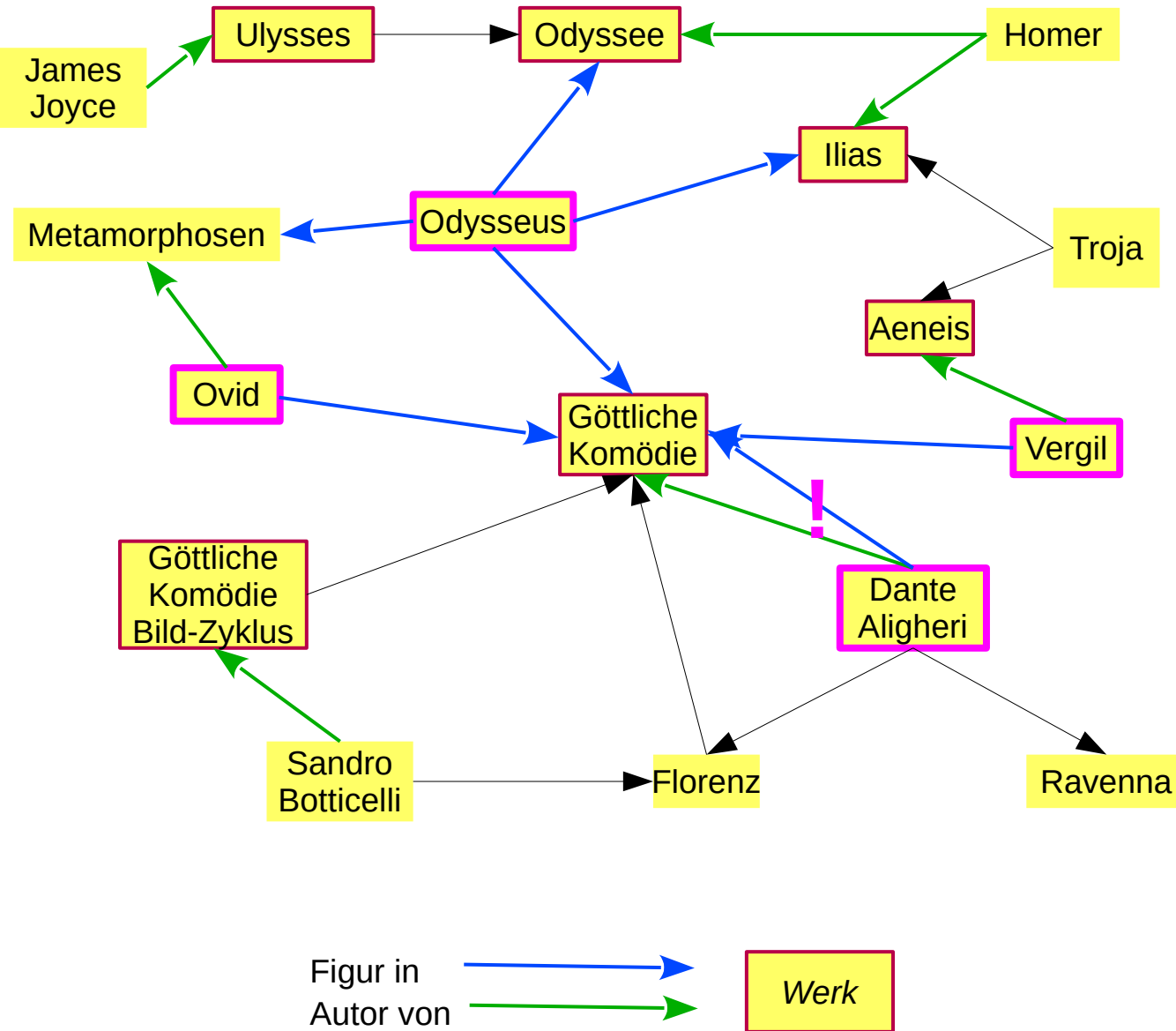


Welche Figuren kommen in der Göttlichen Komödie vor ?



Figur in   *Werk*
 Autor von   *Werk*

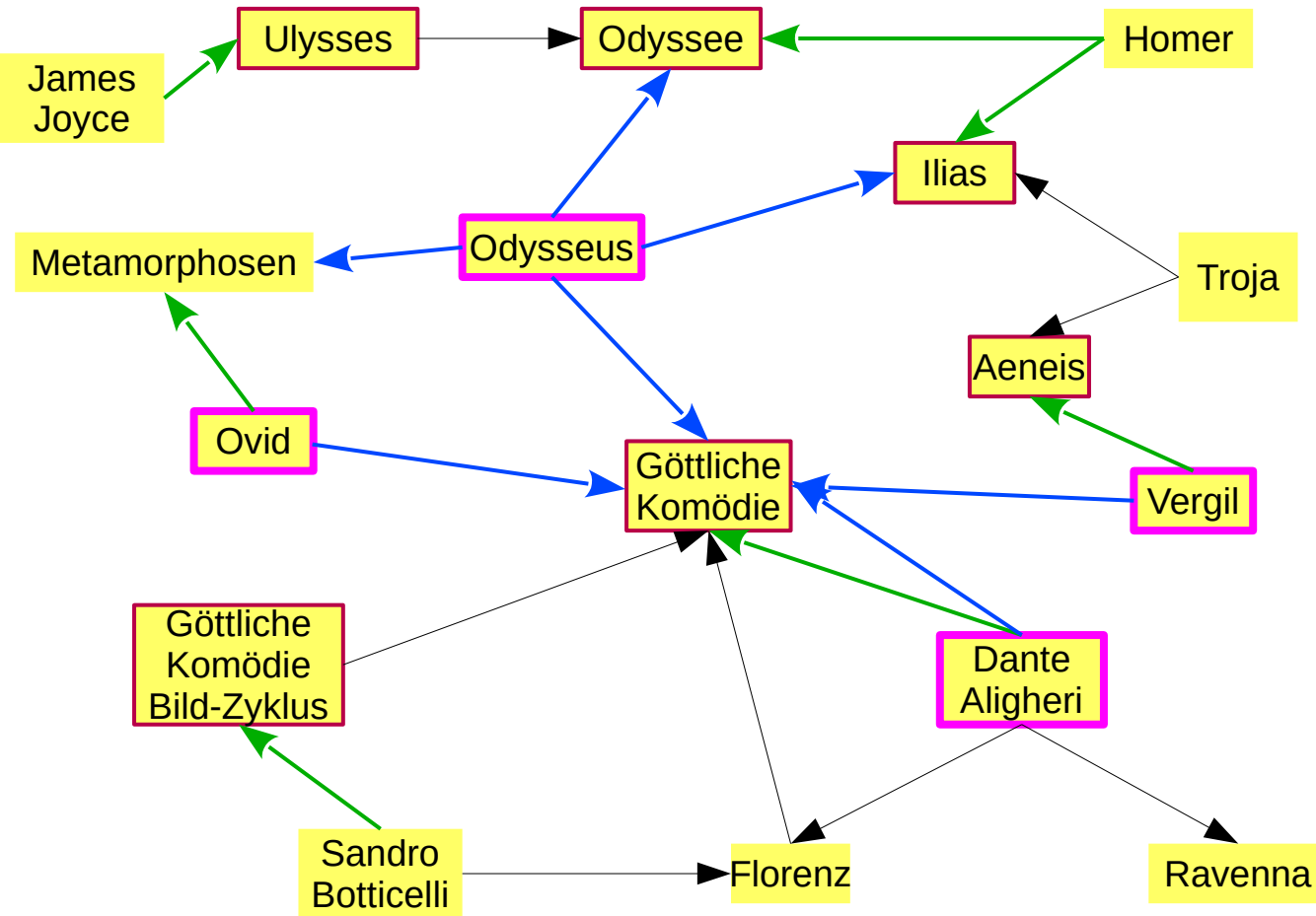
Suche



Welche Figuren kommen in der Göttlichen Komödie vor ?

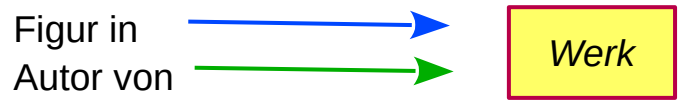


Suche

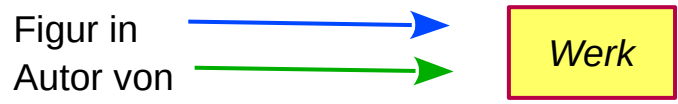
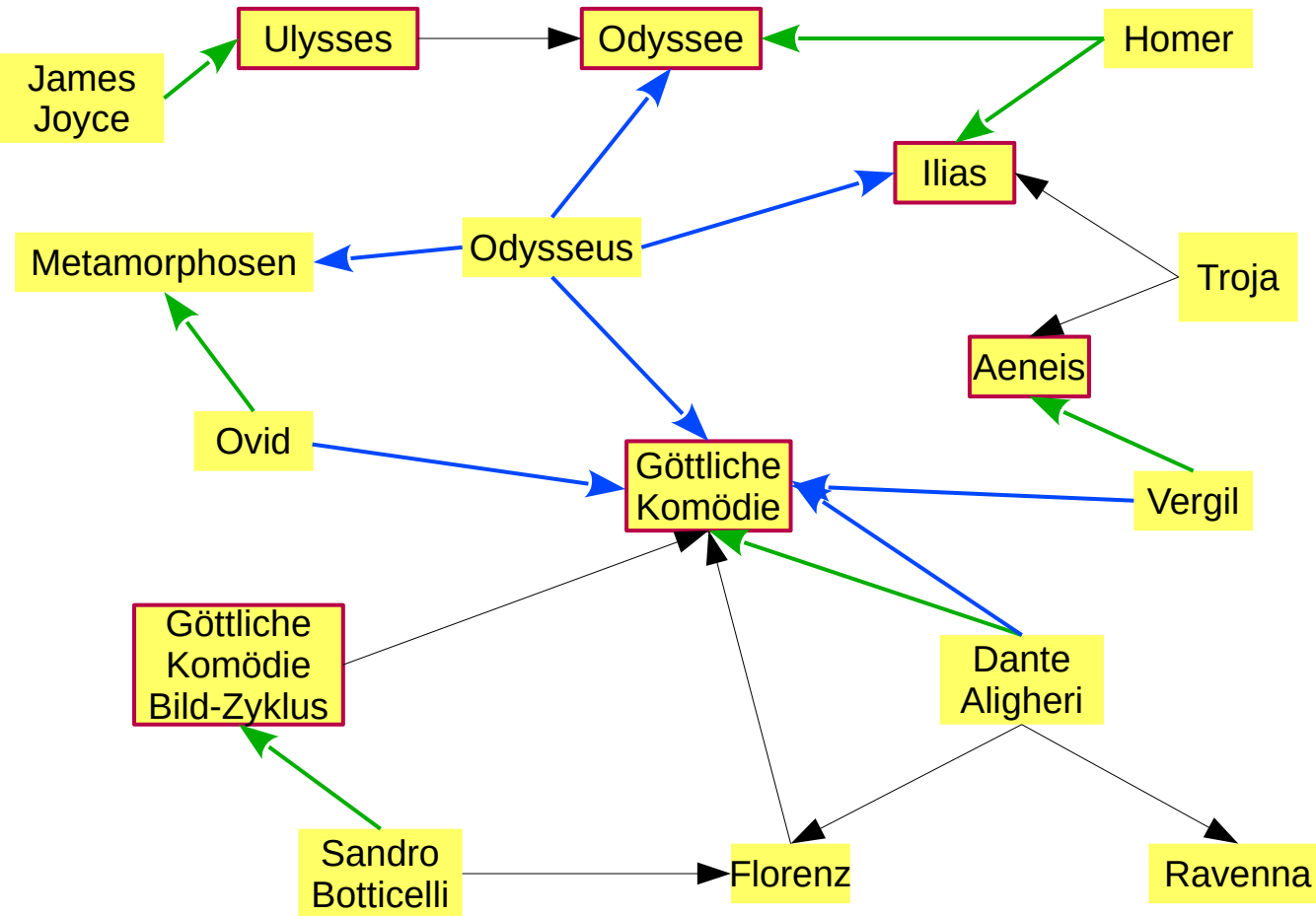


Welche Figuren kommen in der Göttlichen Komödie vor ?

- Ovid
- Odysseus
- Vergil
- Dante Aligheri

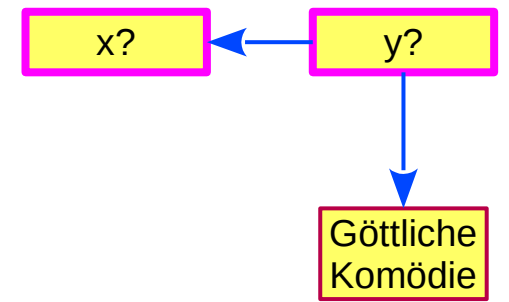


Suche

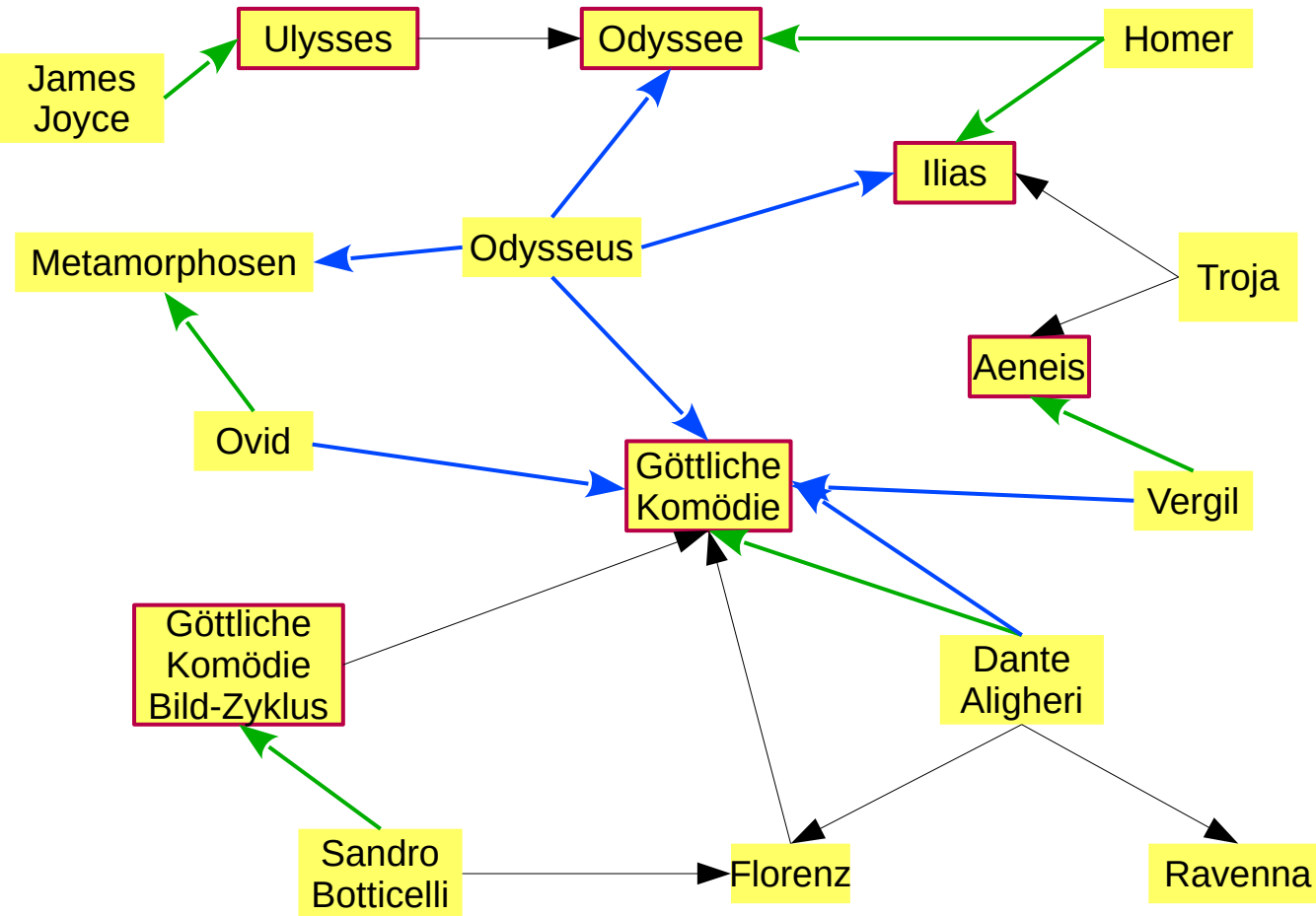


Beispiel 2

Welche Werke teilen eine Figur mit der Göttlichen Komödie?

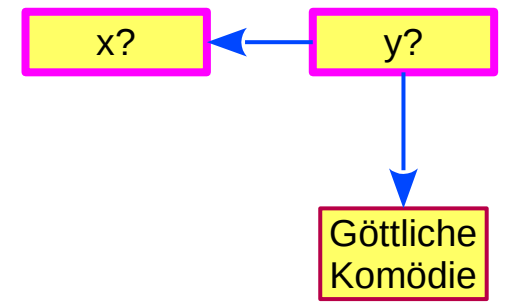


Suche

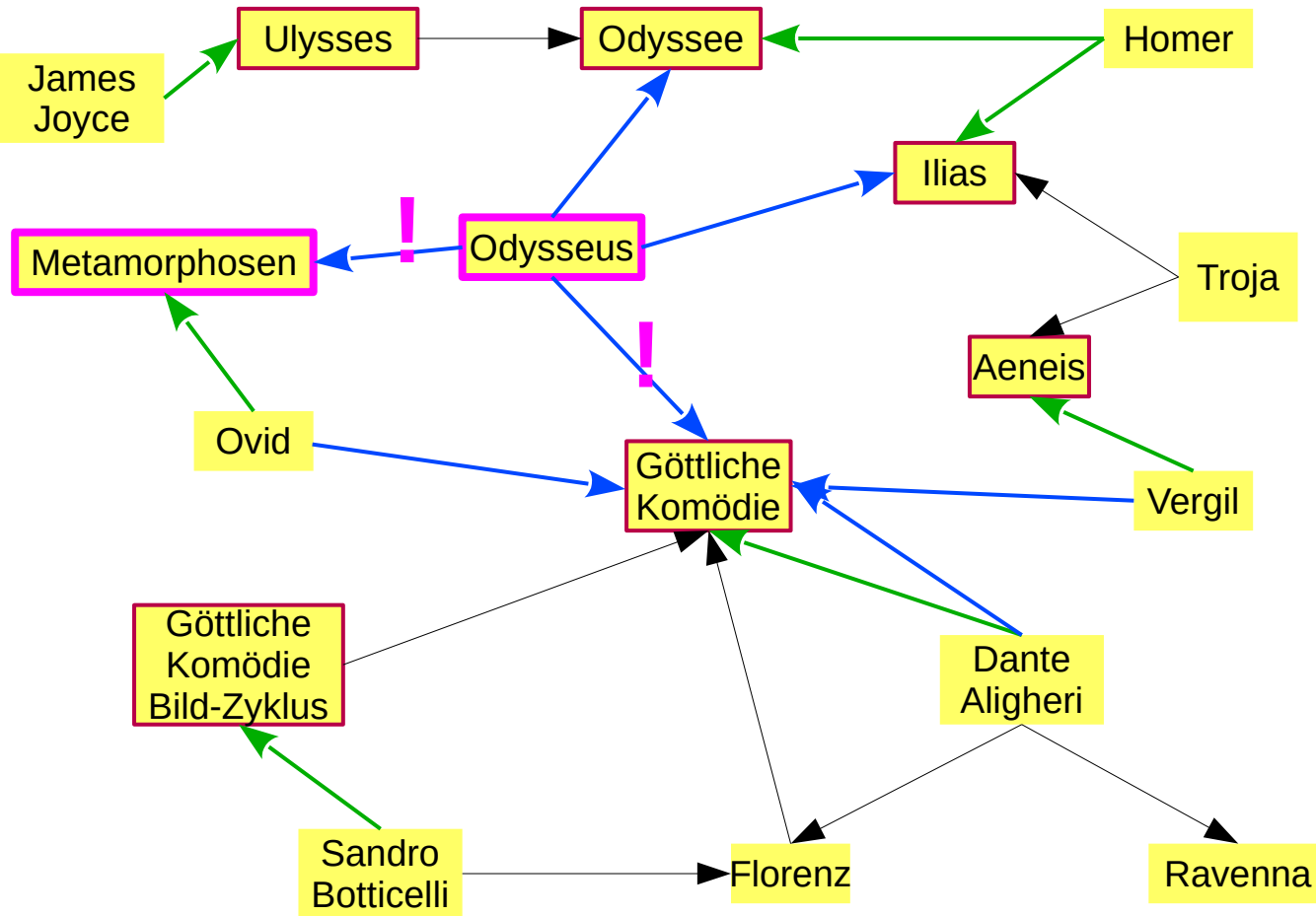


Beispiel 2

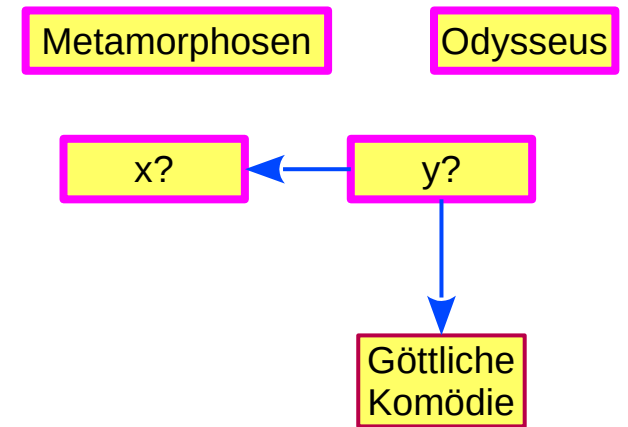
Welche Werke teilen eine Figur mit der Göttlichen Komödie?







Suche

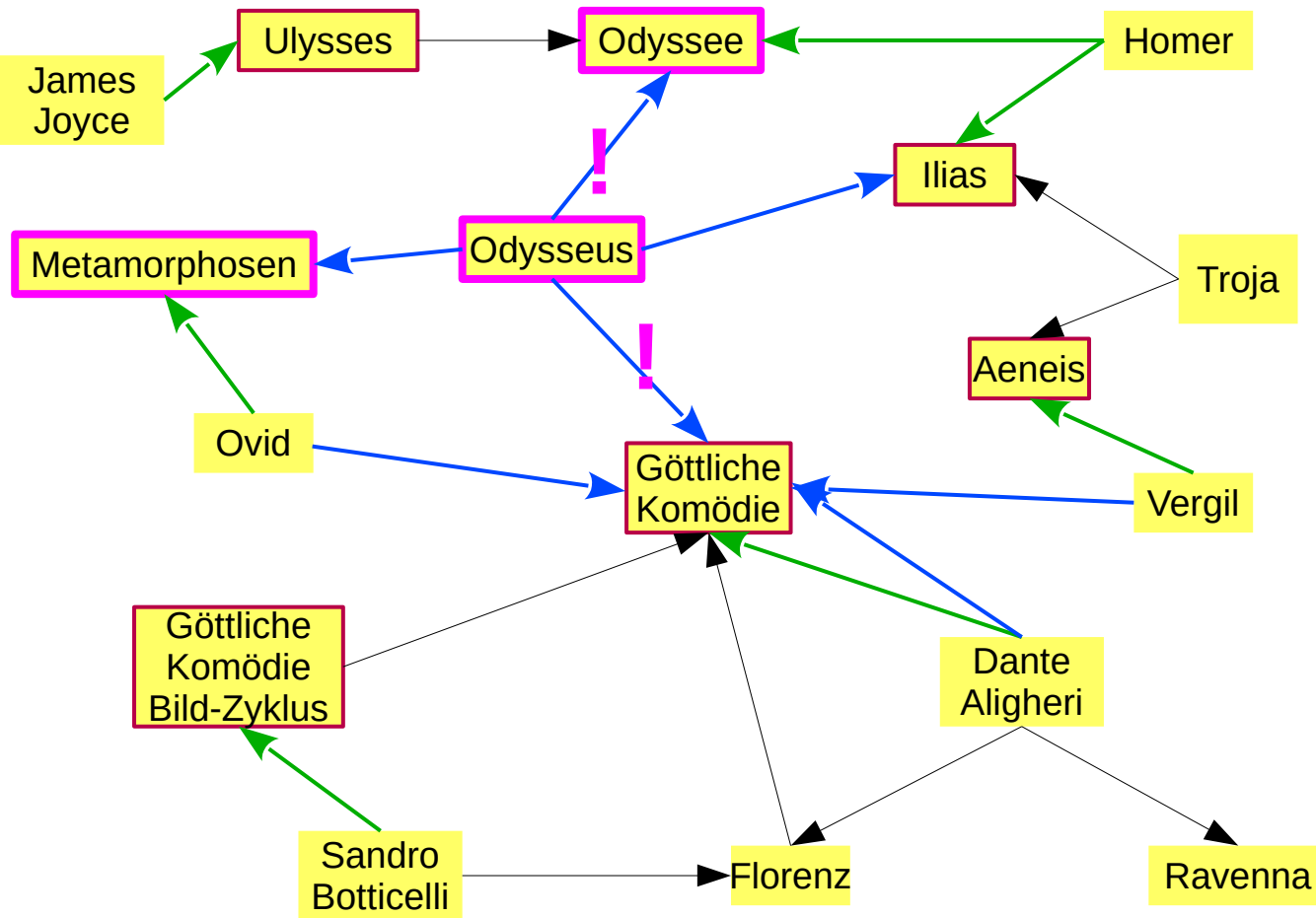


Welche Werke teilen eine Figur mit der Göttlichen Komödie?

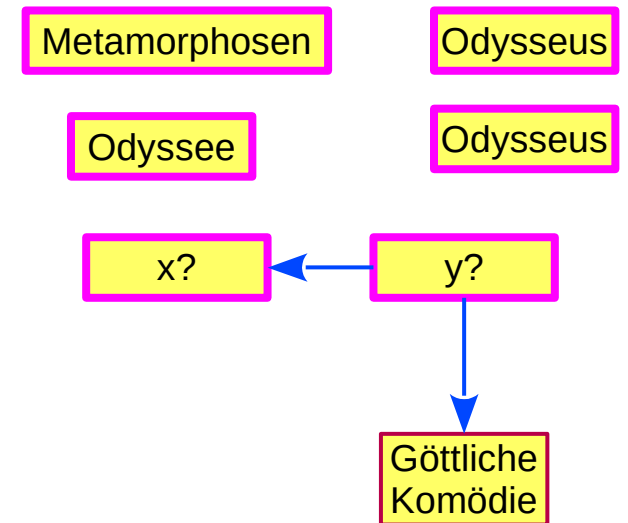






Figur in   *Werk*
 Autor von   *Werk*

Suche

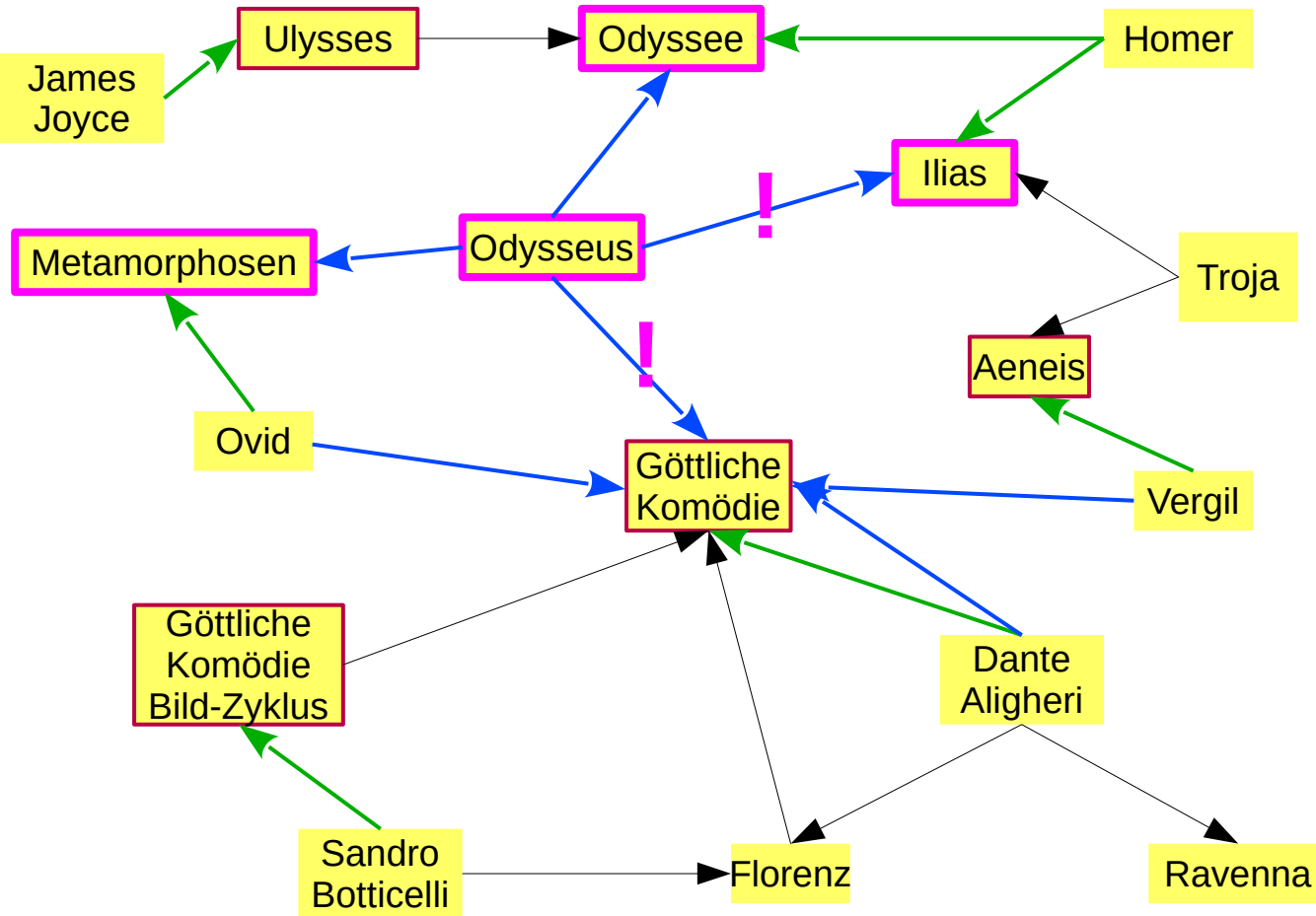


Welche Werke teilen eine Figur mit der Göttlichen Komödie?

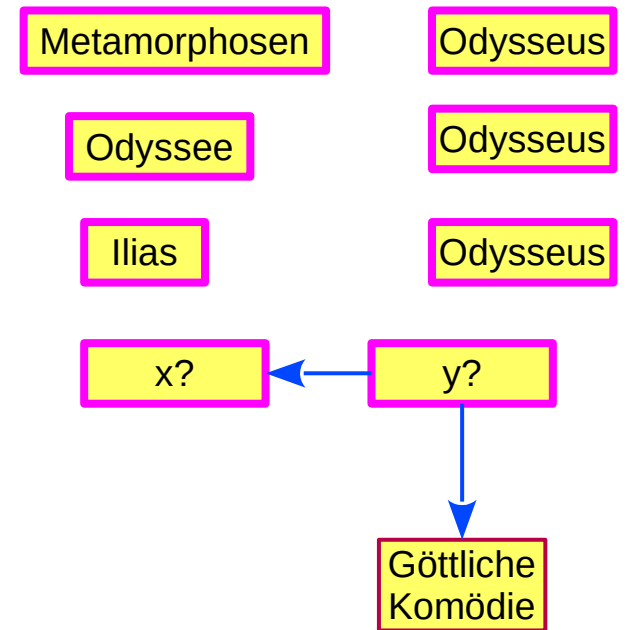






Figur in   *Werk*
 Autor von   *Werk*

Suche

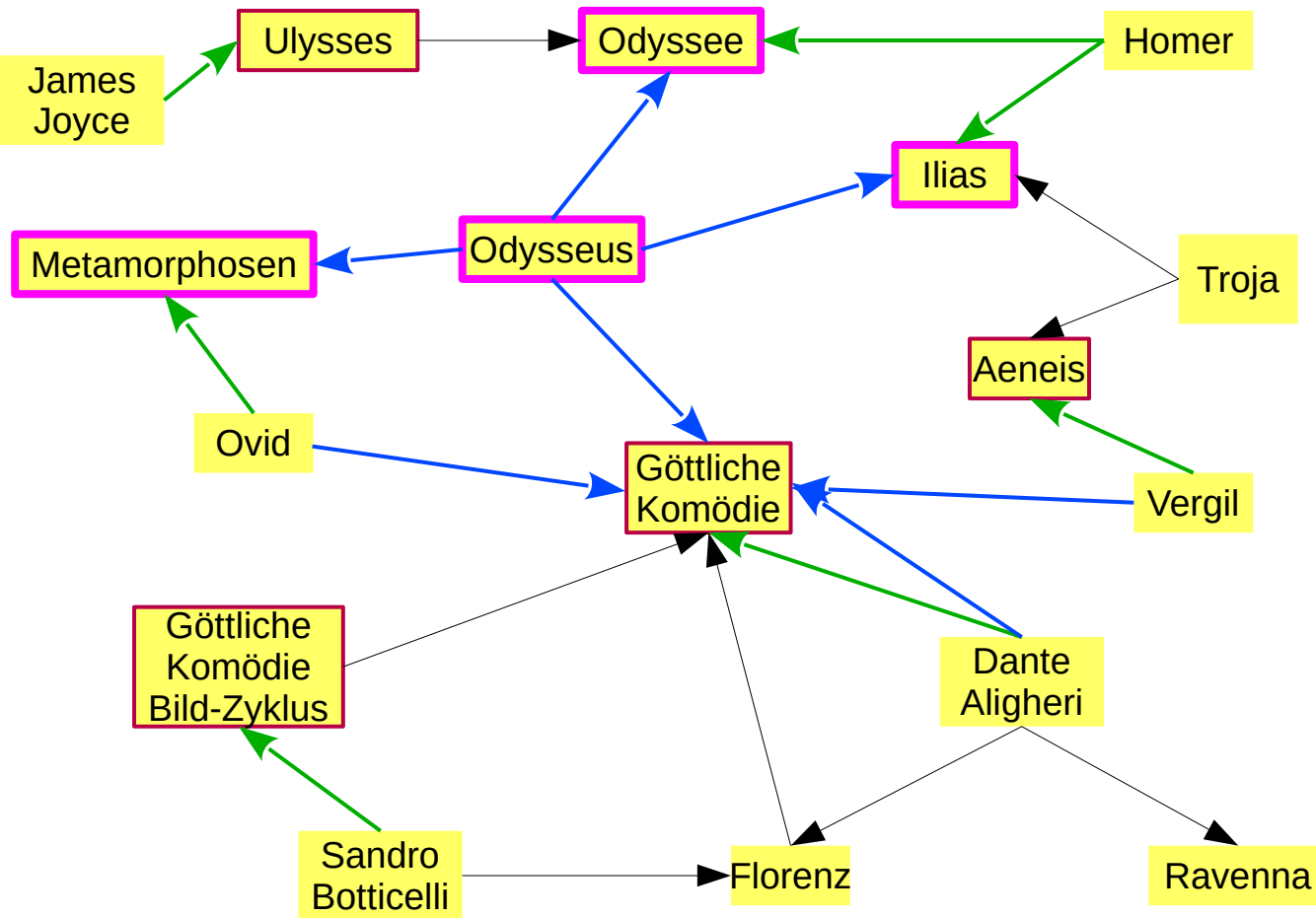


Welche Werke teilen eine Figur mit der Göttlichen Komödie?







Figur in   *Werk*
 Autor von   *Werk*

Suche

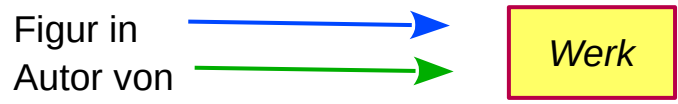
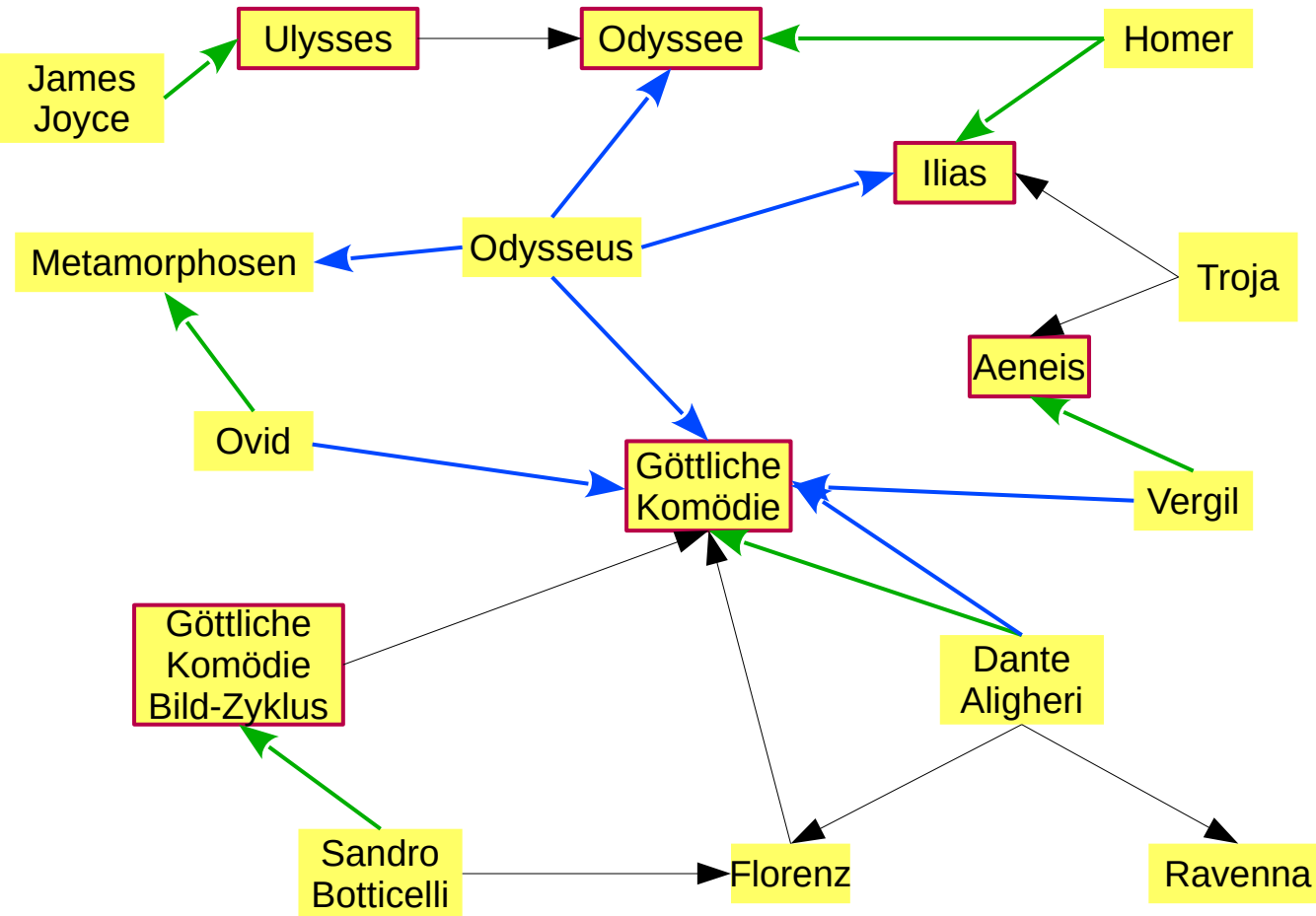


Welche Werke teilen eine Figur mit der Göttlichen Komödie?

- | | |
|---------------|----------|
| Metamorphosen | Odysseus |
| Odyssee | Odysseus |
| Ilias | Odysseus |

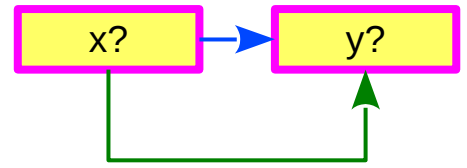
Figur in   *Werk*
 Autor von   *Werk*

Suche

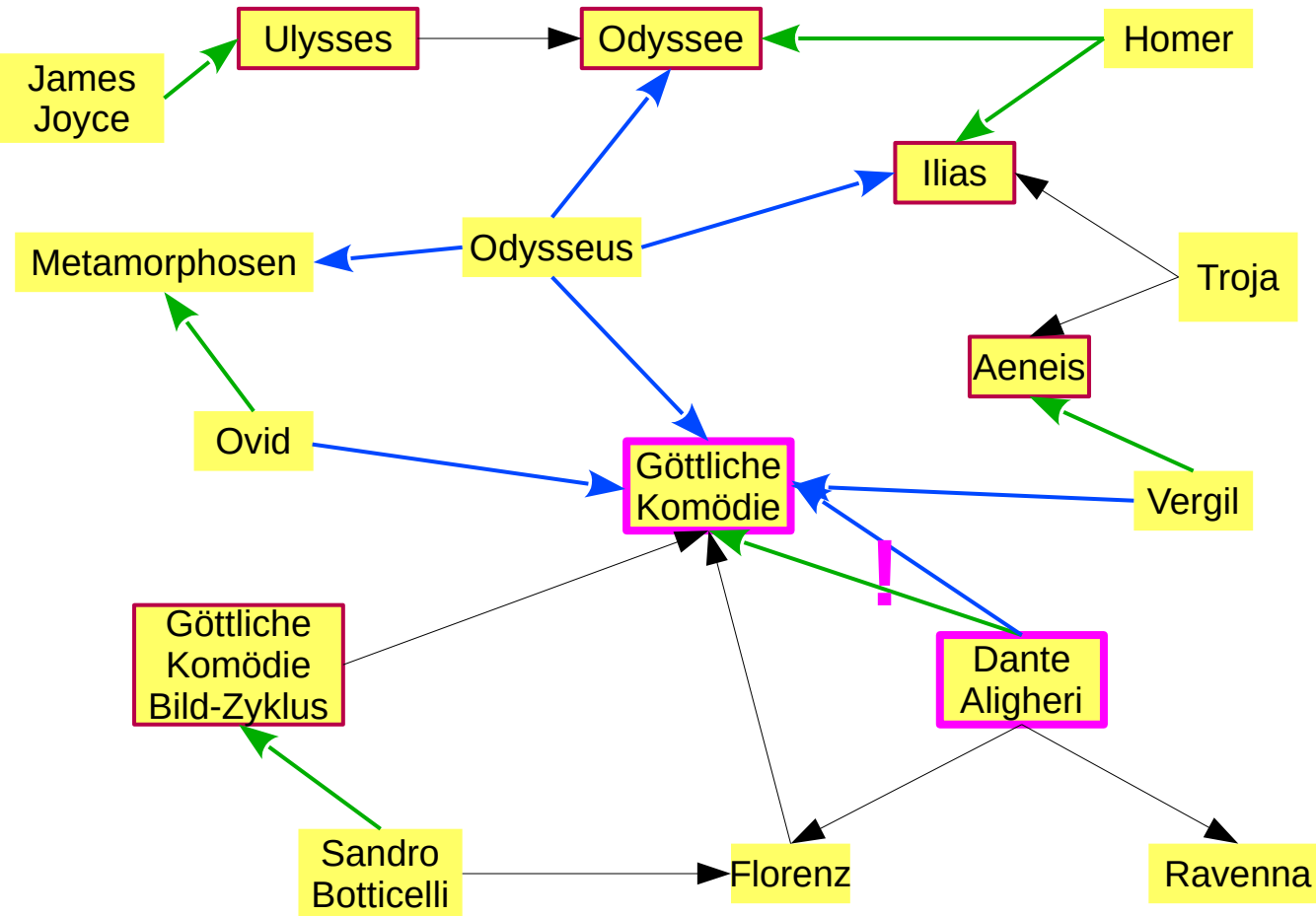


Beispiel 3

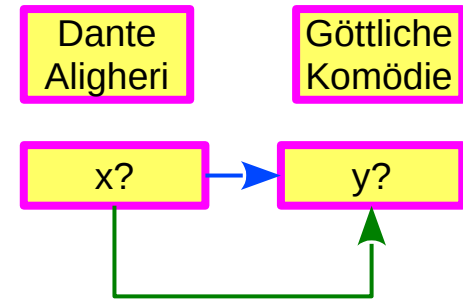
Welcher Autor ist Figur in seinem eigenen Werk?







Suche

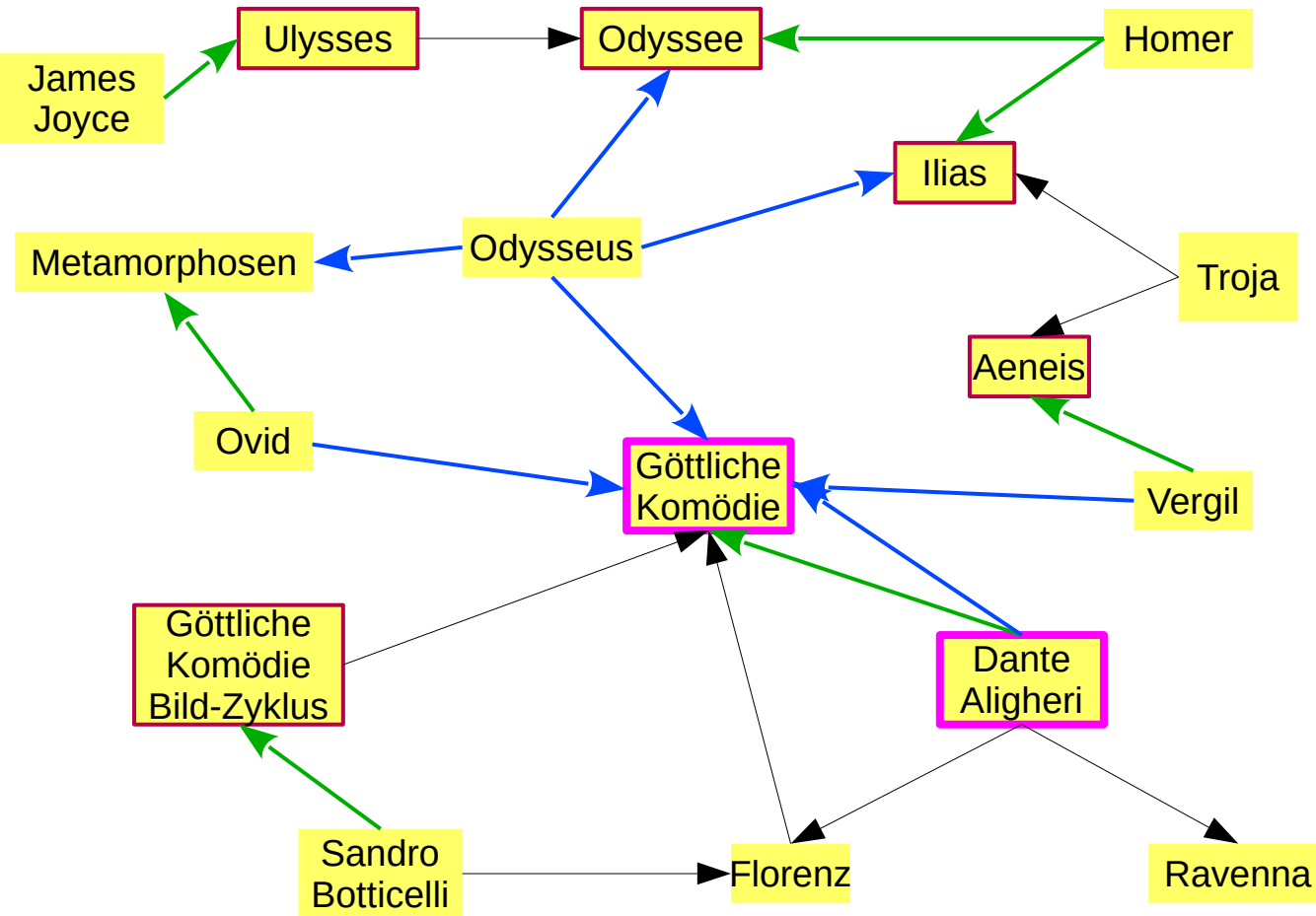


Welcher Autor ist Figur in seinem eigenen Werk?



Figur in   *Werk*
 Autor von   *Werk*

Suche



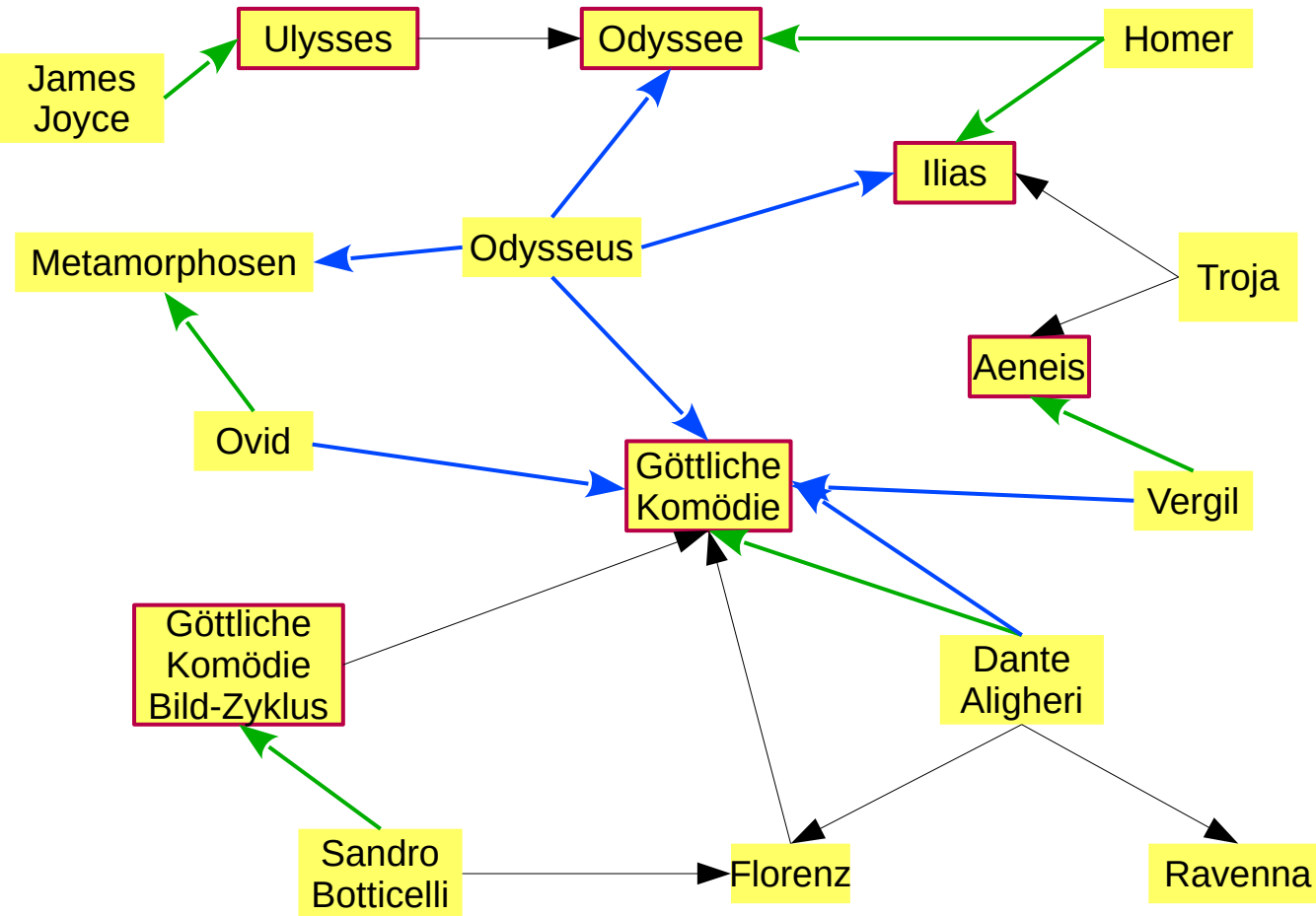
Welcher Autor ist Figur in seinem eigenen Werk?

Dante Aligheri

Göttliche Komödie

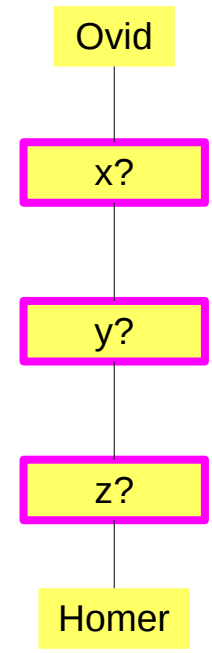


Suche

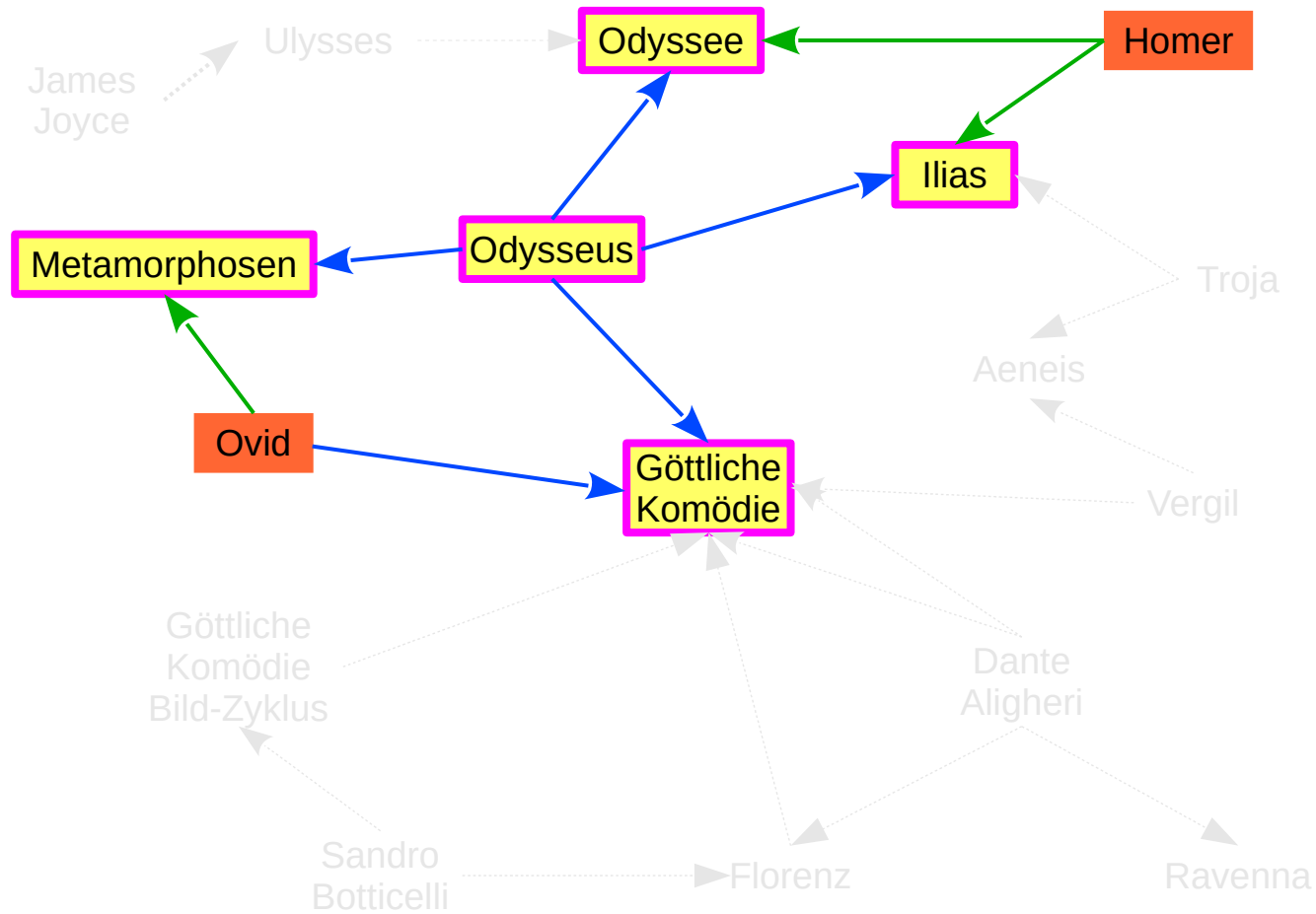


Beispiel 4

Welche Verbindung besteht zwischen Homer und Ovid (über max. drei Knoten)?



Suche



Figur in
Autor von



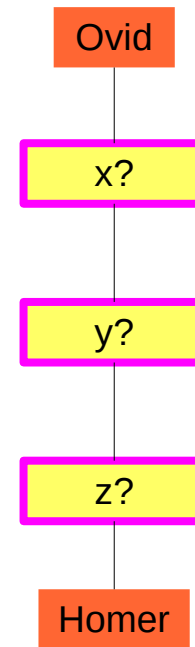
Werk

→ ReIFinder:

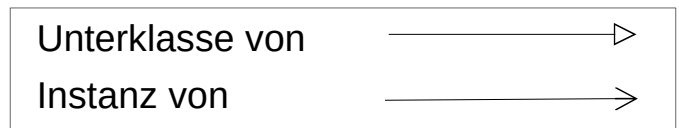
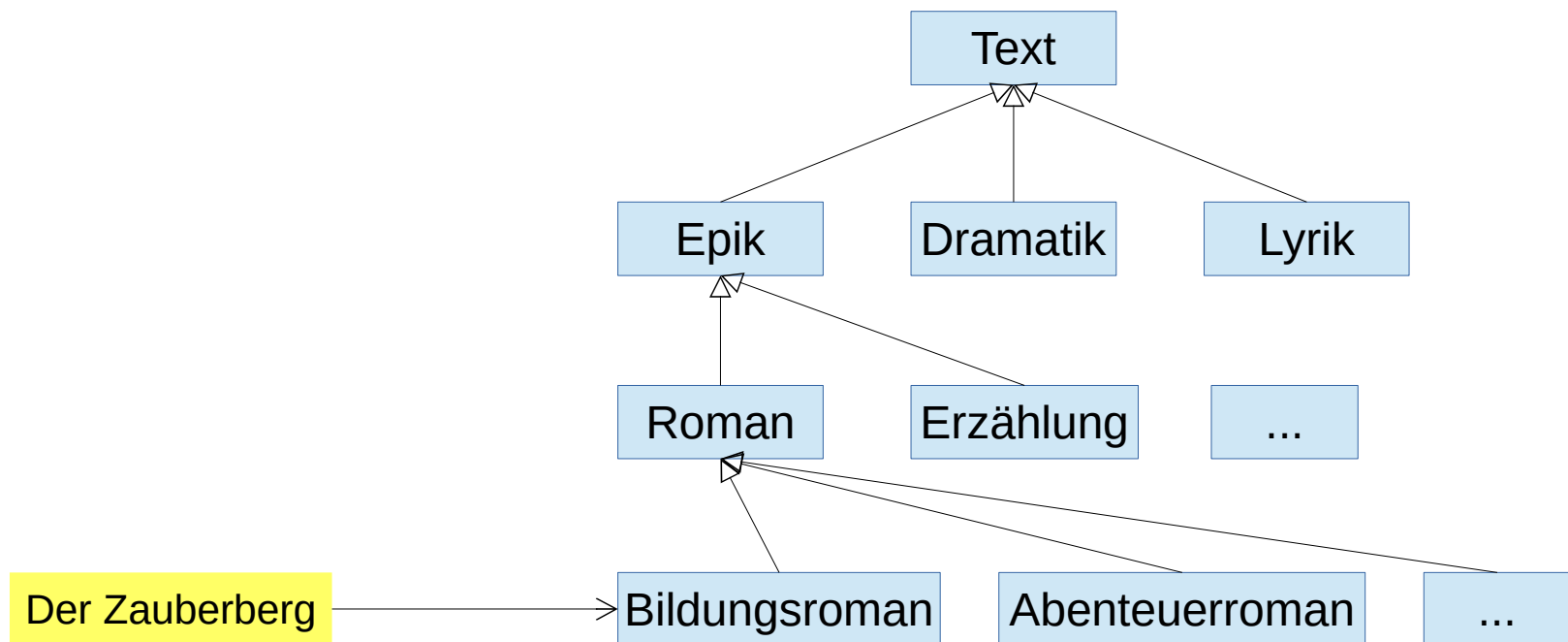
<http://www.visualdataweb.org/refinder/refinder.php>

Beispiel 4

Welche Verbindung besteht zwischen Homer und Ovid (über max. drei Knoten)?

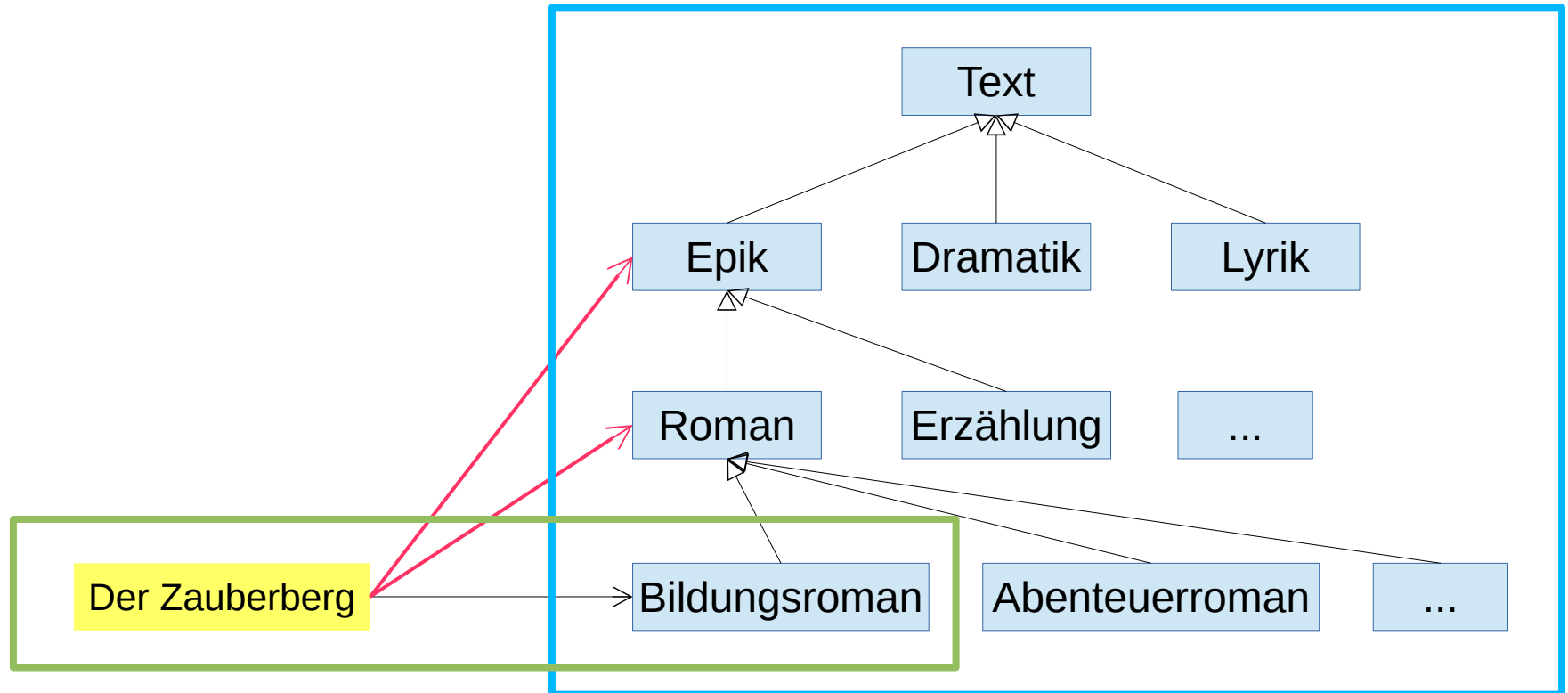


Reasoning



Reasoning

Schema



explizite Fakten

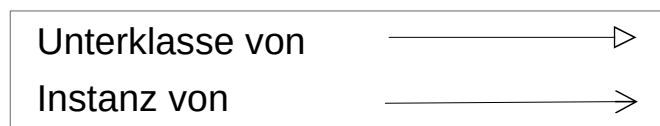
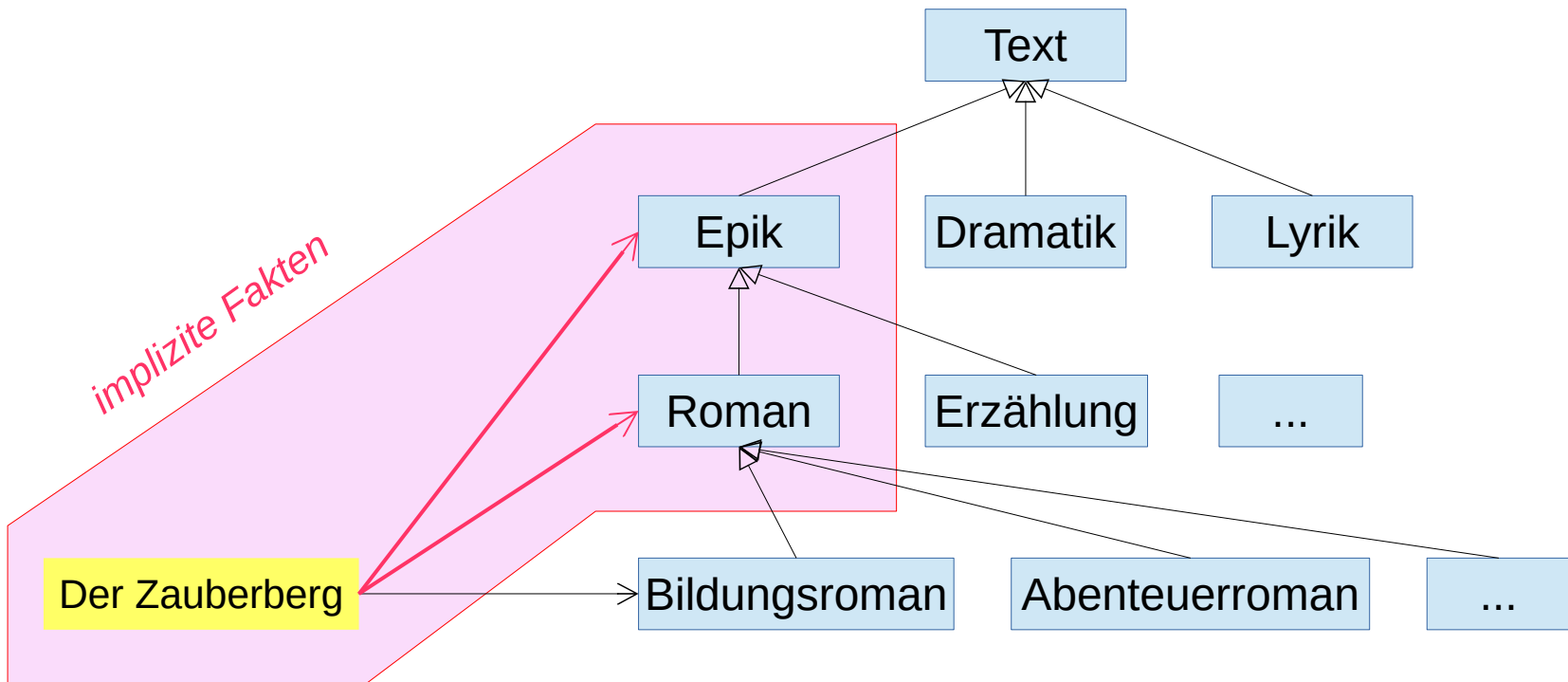


explizite Relationen



implizite Relationen

Reasoning



explizite Relationen



implizite Relationen



WIKIDATA

Wikidata

Wikidata is a free, collaborative, multilingual, secondary database, collecting structured data to provide support for Wikipedia, Wikimedia Commons, the other wikis of the Wikimedia movement, and to anyone in the world.

- **Free.** The data in Wikidata is published under the CC 1.0
- **Collaborative.** Data is entered and maintained by Wikidata editors who decide on the rules of content creation and management. Automated bots also enter data into Wikidata.
- **Multilingual.** Editing, consuming, browsing, and reusing the data is fully multilingual. Editing in any language is possible and encouraged.
- **A secondary database.** Wikidata records not just statements, but also their sources, and connections to other databases. (→ diversity of knowledge available; → verifiability)
- **Collecting structured data.** High degree of structured organization allows for easy reuse of data by Wikimedia projects and third parties, and enables computers to process and “understand” it.
- **Support for Wikimedia wikis.** Wikidata assists Wikipedia with more easily maintainable information boxes and links to other languages, thus reducing editing workload while improving quality. Updates in one language are made available to all other languages.
- **Anyone in the world.** Anyone can use Wikidata for any number of different ways by using its API.

Wikidata – Historischer Abriss

Das Projekt Wikidata wurde erstmals im September 2004 von Erik Möller vorgeschlagen. Kurz darauf Proof of Concept Implementierung von Magnus Manske.

Beginn der eigentlichen Umsetzung 2012 durch Spenden 1.3 Mill €

(Allen Institute for Artificial Intelligence, Gordon and Betty Moore Foundation, Google Inc.)

Anschluss Finanzierung geplant durch Wikimedia Deutschland für einen Projektplan:

Phase 1:

30.10.2012 offizieller Start von Wikidata mit Freigabe für die Bearbeitung

Anfangs wurden nur Sprachlinks zu Datenobjekten erfasst.

(it: Divina Commedia, de: Göttliche Kommüdie, en: The Divine Comedy, ...)

Phase 2:

Ab 4.2.2013 Aussagen zu Datenobjekten möglich (Dante ist Autor von Divinia Commedia)

Ziel: Informationen auf Infoboxen in den Wikipedia-Sprachversionen zentral in Wikidata zu speichern und diese bei Bedarf in Wikipedia zu verwenden (seit 23.4.2013 für alle Wikipedias zugänglich)

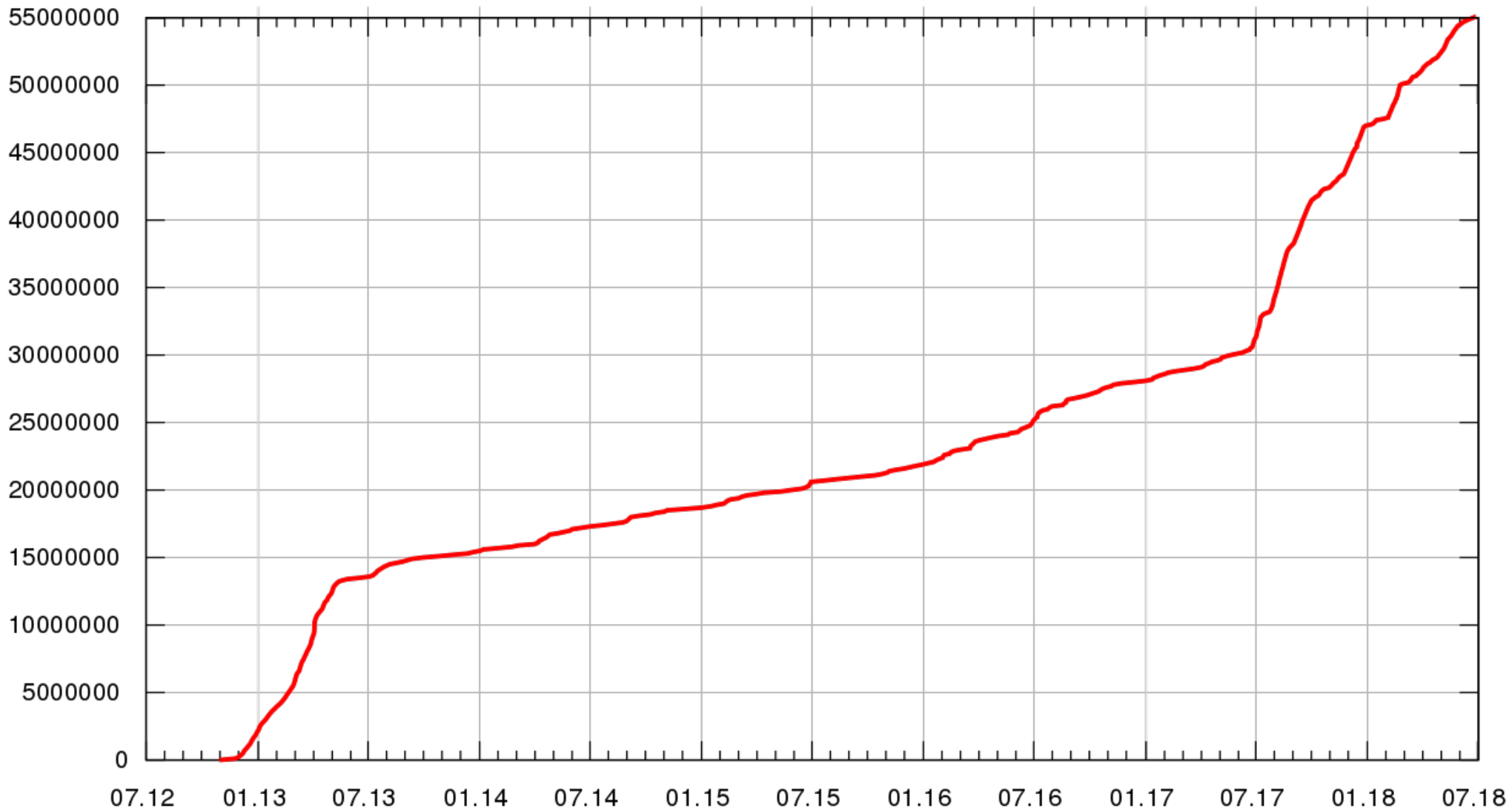
Phase 3:

Automatisiertes erstellen von Listen (Liste aller italienischen Schriftsteller)

Weitere Meilensteine:

- 2013: IBM spendet Preisgeld des Feigenbaum-Preises 2013 für das Watson-Projekt
- 2015: Freebase (Faktendatenbank von Google) wird in Wikidata überführt
- 2016: Bereitstellung des **SPARQL-Endpoints**

Entwicklung der Datenobjektzahlen auf Wikidata



Verbreitung von Wikidata

Eine Umfrage, die der Bibliotheks-Dienstleister OCLC im Jahr 2018 durchgeführt hatte, ergab, dass Wikidata unter den Linked-Data-Projekten, aus denen Daten in eigene Angebote übernommen wurden, mit 41 % der Befragten den Platz fünf einnahm – vor WorldCat und ISNI, aber deutlich nach der Library of Congress und VIAF. Im Vergleich zu 2015 sei Wikidata vor allem für Bibliotheken, Museen und Archive bedeutsamer geworden.

[<https://de.wikipedia.org/wiki/Wikidata#Verbreitung>]

Karen Smith-Yoshimura: *Analysis of 2018 International Linked Data Survey for Implementers*. In: *Code4Lib Journal*. Band 42, 8. November 2018, ISSN 1940-5758 (code4lib.org).

