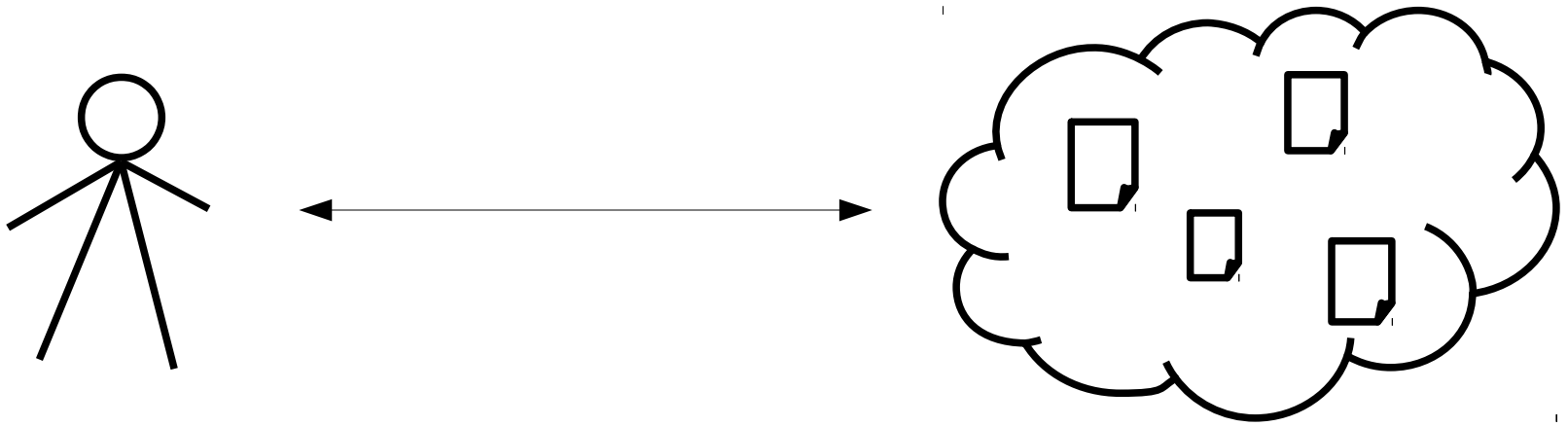
Semantics and Domain Modeling
for Personalized Social Web

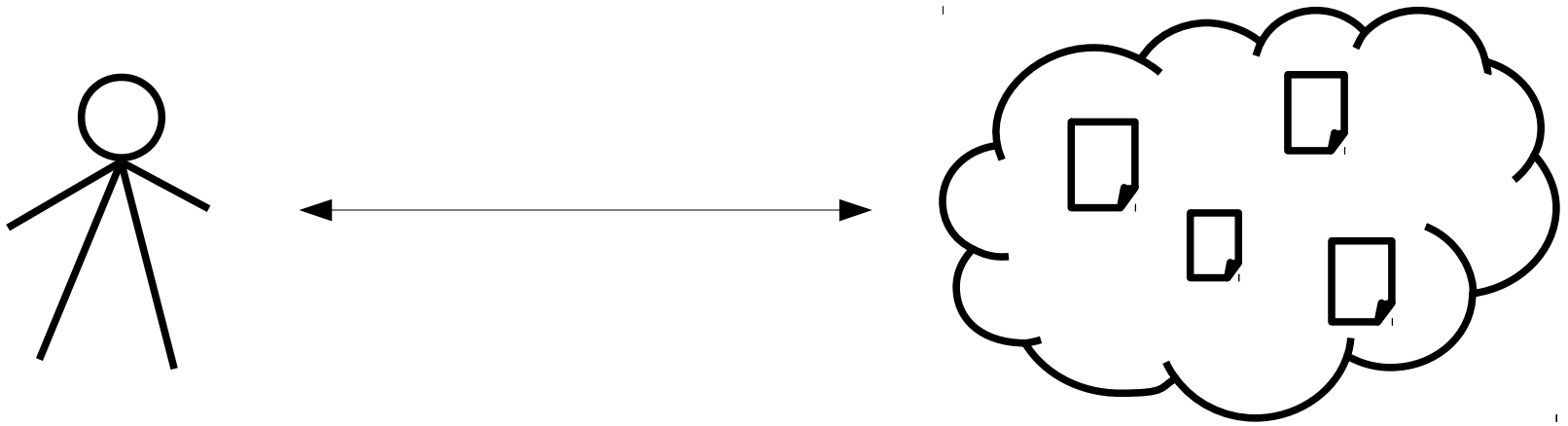
Marián Šimko

15. 10. 2014

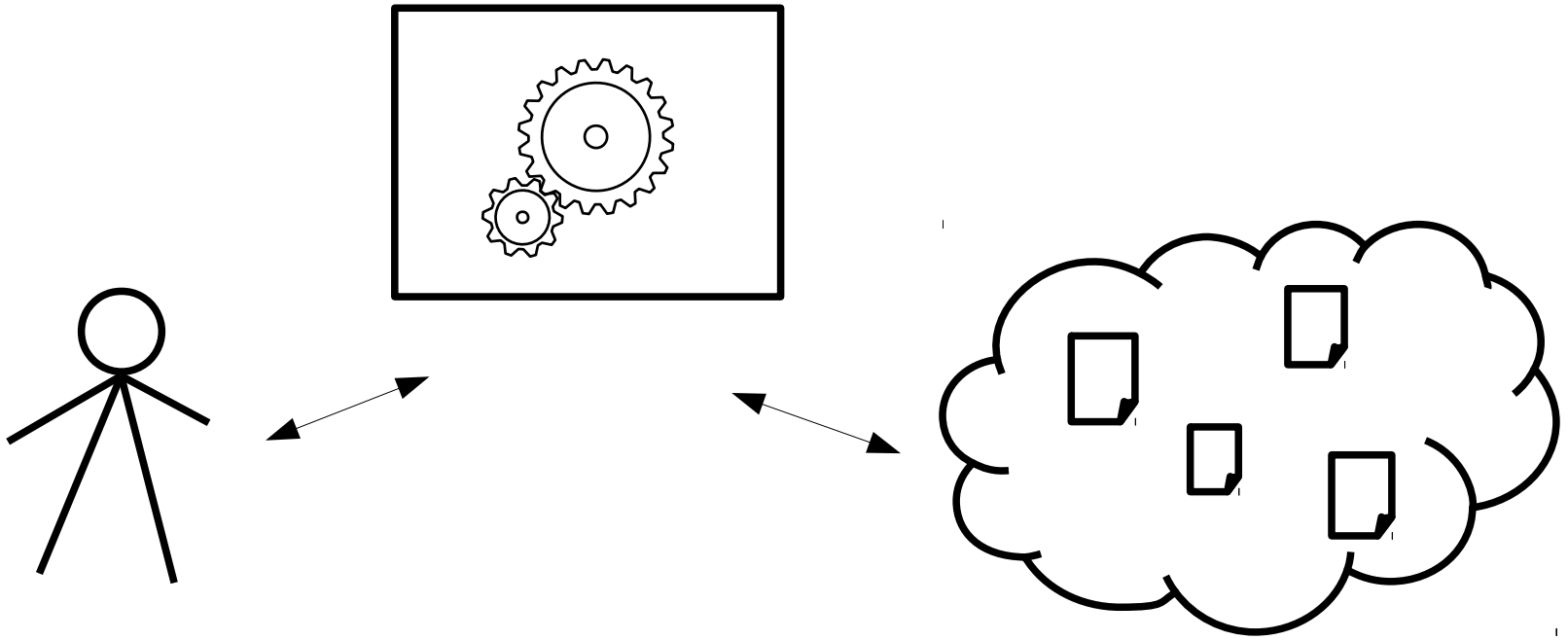
Prístup k informáciám

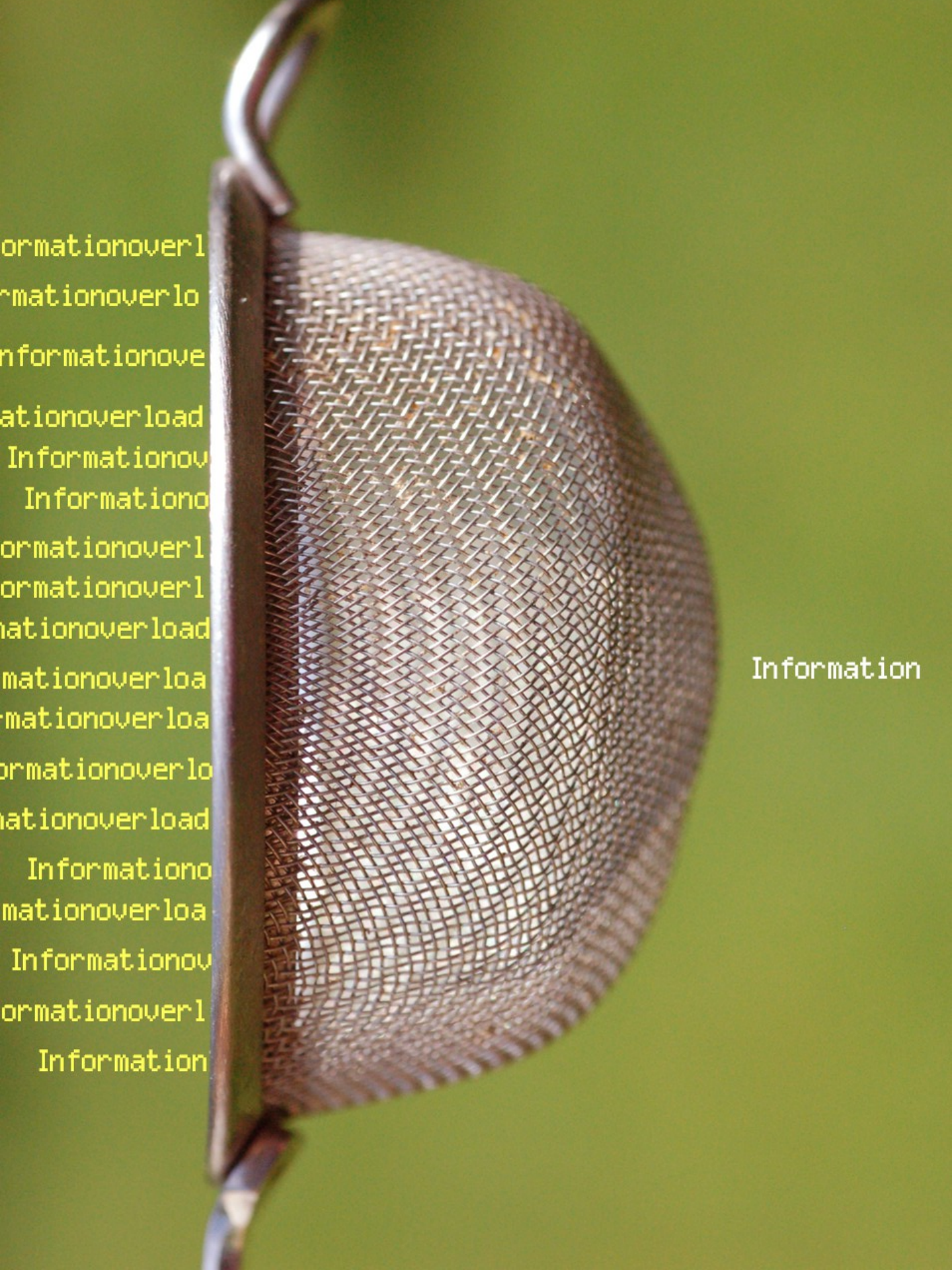


Prístup k informáciám

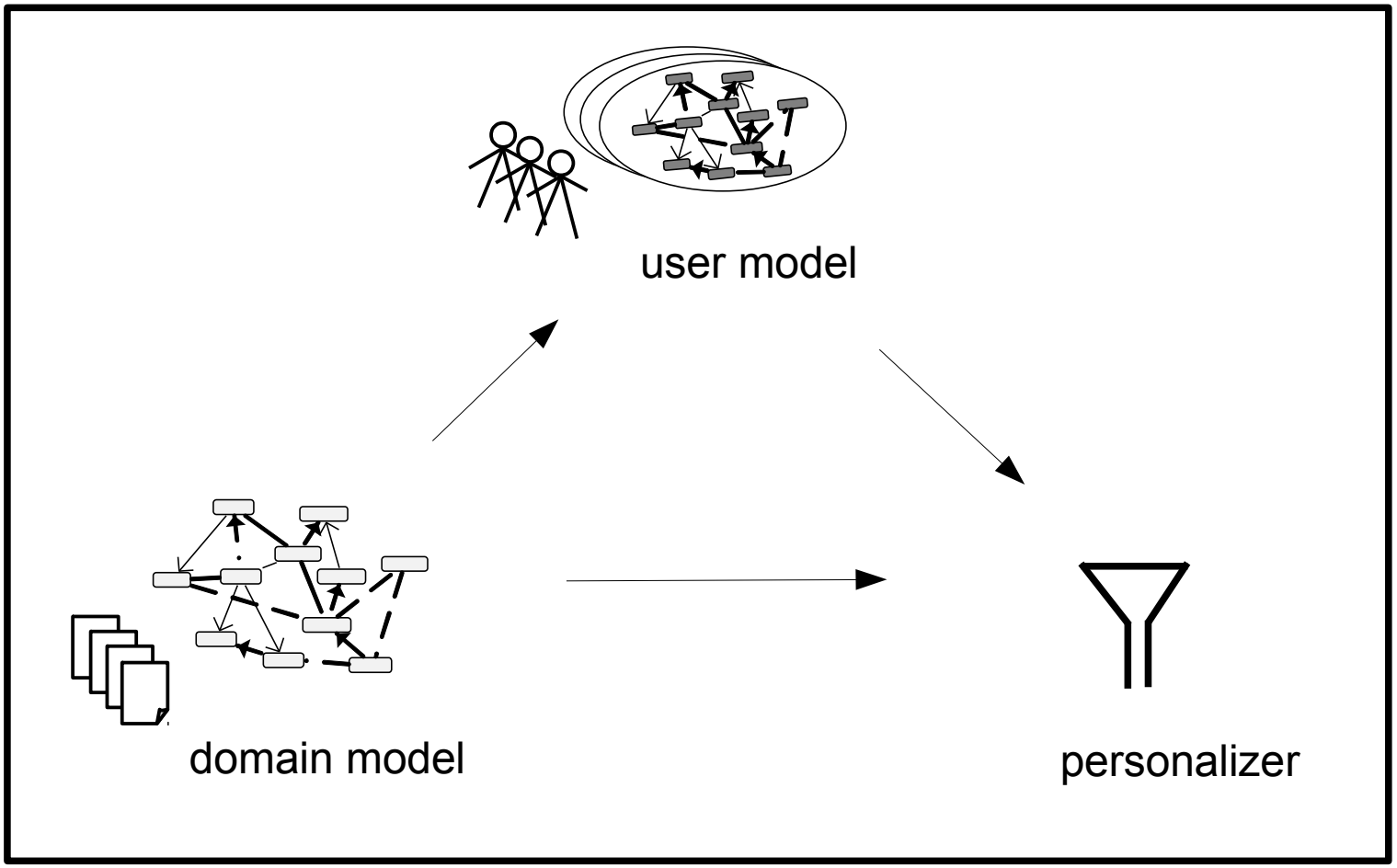


Prístup k informáciám

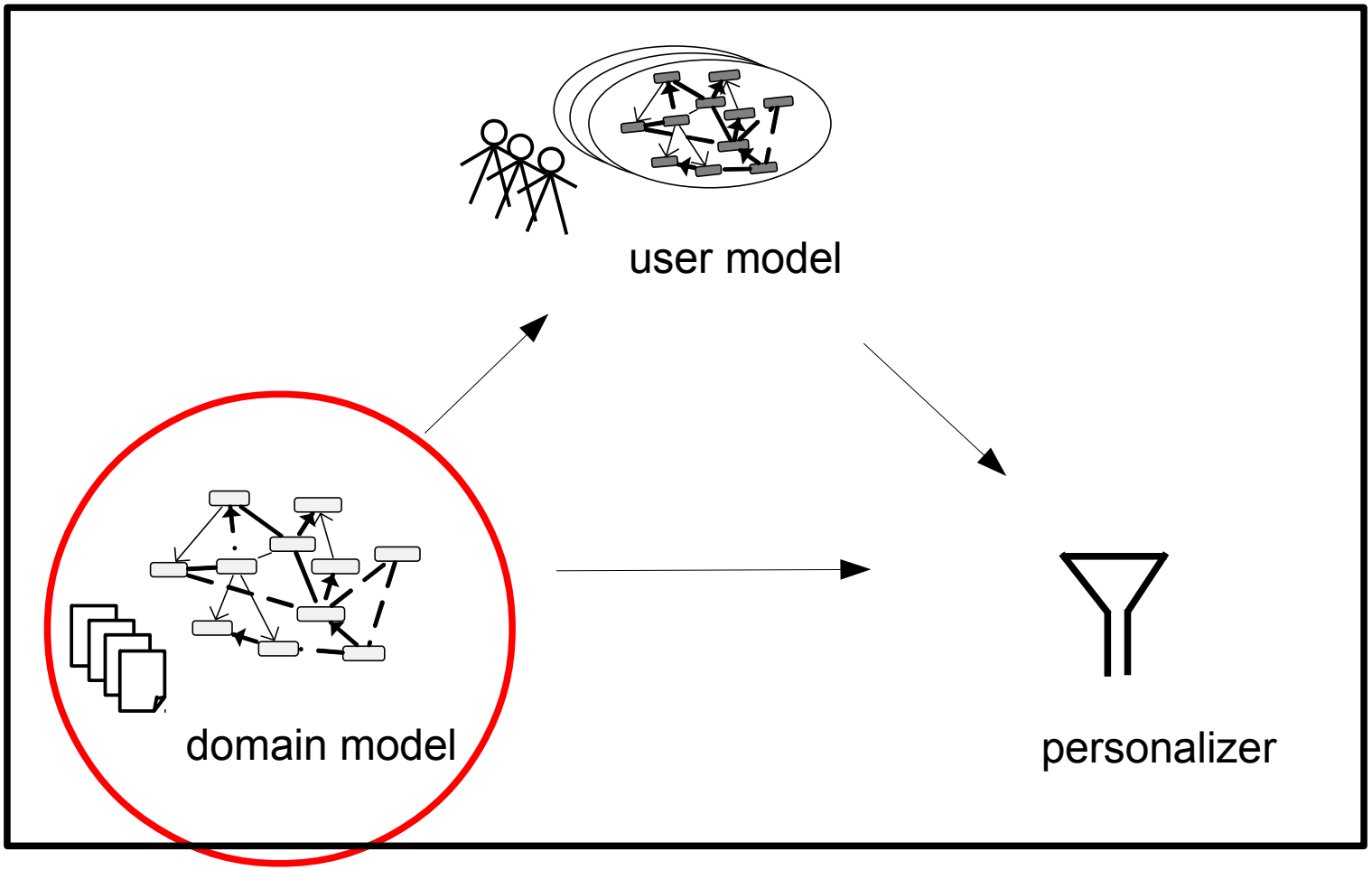



ormationoverl
rmationoverlo
nformationove
ationoverload
Informationov
Informationo
ormationoverl
ormationoverl
nationoverload
mationoverloa
rmationoverloa
ormationoverlo
nationoverload
Informationo
mationoverloa
Informationov
ormationoverl
Information
Information

Personalised Web-based System



Personalised Web-based System



Sémantika

- = význam
- Web so sémantikou (the Semantic Web)

  "man-made woven web of data" that facilitates machines to understand the semantics, or meaning, of information on the World Wide Web
- načo?
  - usudzovanie, odvodzovanie
  - inteligentné vyhľadávanie, odporúčanie

Sémantika

- = význam
- Web so sémantikou (the Semantic Web)

  "man-made woven web of data" that facilitates machines to understand the semantics, or meaning, of information on the World Wide Web
- načo?
  - usudzovanie, odvodzovanie
  - inteligentné vyhľadávanie, odporúčanie

4 எளிய வழிமுறைகளை பிஸ்ஸா மீர்கேரிட்டா

(198 மதிப்பீடுகள்) நல்ல உணவு மூலம் இதழ் சந்தா - £ 3.3 பிரச்சினைகள்

பிரெ: 25 நிமிடங்கள்
சமையல் கலைஞர்கள்: 10 நிமிடங்கள்

எளிதாக

4 உதவுகிறது, 2 பீஸ்ஸாக்கள் செய்கிறது

கூட ஒரு புதிய சமையல்காரர் எங்கள் எளிய படி படிப்படியாக வமிகாட்டி பீஸ்ஸா கலை மாஸ்டர் (மடியம். Bellissimo

சேவை ஒன்றுக்கு ஊட்டச்சத்து

kcalories	புரதம்	காபோவைதரேற்று	கொழுப்பு	வளமாக்கி	இழை	சர்க்கரை	உப்பு
431	19g	59g	15g	7G	3g	2g	1.87g

கூடுதல் தகவல்

பீஸ்ஸாக்கள் வேகவைக்காத

எளிதாக / மடங்காக பாதியாக

சைவம்

என் நல்ல உணவு சேமி

அச்சு

தேவையான பொருட்கள்

அடிப்படை

300g வலுவான ரொட்டி மாவு

1 தேக்கரண்டி உடனடி ஈஸ்ட் (ஒரு பாக்கெட்டை அல்லது ஒரு தொட்டி இருந்து)

1 தேக்கரண்டி உப்பு,

1 டீஸ்பூன் ஆலிவ் எண்ணெய், மற்றும் drizzling கூடுதல்

தக்காளி சாஸ்

100 நொறுக்கப்பட்ட தக்காளி

சில புதிய துளசி அல்லது 1 தேக்கரண்டி காய்ந்த

முறை

1. அடிப்படை செய்ய: ஒரு பெரிய கிண்ணத்தில் மாவு வைத்து, பின்னர் ஈஸ்ட் மற்றும் உப்பு உள்ள அசை. நன்கு, 200ml வெதுவெதுப்பான தண்ணீர் மற்றும் ஆலிவ் எண்ணெய் ஊற்ற மற்றும் ஒரு மென்மையான, மிகவும் ஈரமான மாவை வரை ஒரு மர கரண்டியால் சேர்ந்து கொண்டு வர வேண்டும். சிறிது floured மேற்பரப்பில் மீது திரும்ப மென்மையான வரை 5 நிமிடங்கள் பிசைந்து. ஒரு தேநீர் துண்டு உறை மற்றும் ஒதுக்கியது. நீ விரும்பினால் உயரும் மாவை விட்டு

மாற்று சமையல்

அல்டிமேட் பீஸ்ஸா மார்க்ஹெரிடா

(30 மதிப்பீடுகள்)

மிக எளிய மார்கரட்டா பீஸ்ஸா

(13 மதிப்பீடுகள்)

இலகுவான பிஸ்ஸா மார்கரட்டா

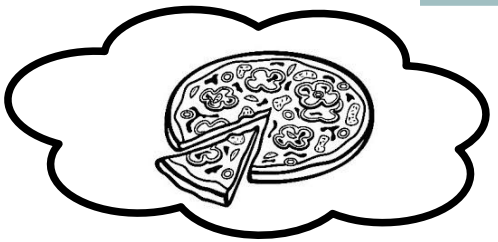
Ako reprezentovať znalosti, vedomosti o niečom?

Základná entita = Koncept, pojem

pojem -jmu *m.* forma myslenia odrážajúca podst. vlastnosti predmetov, javov objektívnej skutočnosti:

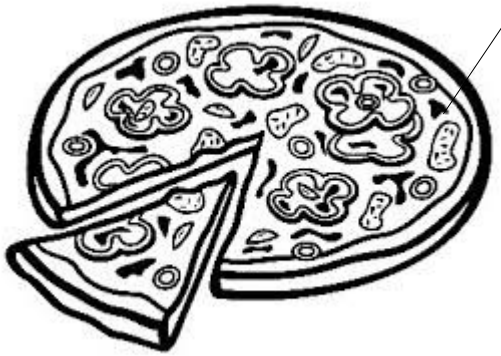
všeobecný, jedinečný p.; *myslíme v p-och a súdoch*;

filoz. výsledok rozumového st. poznania skutočnosti (predmetov, javov ap.)



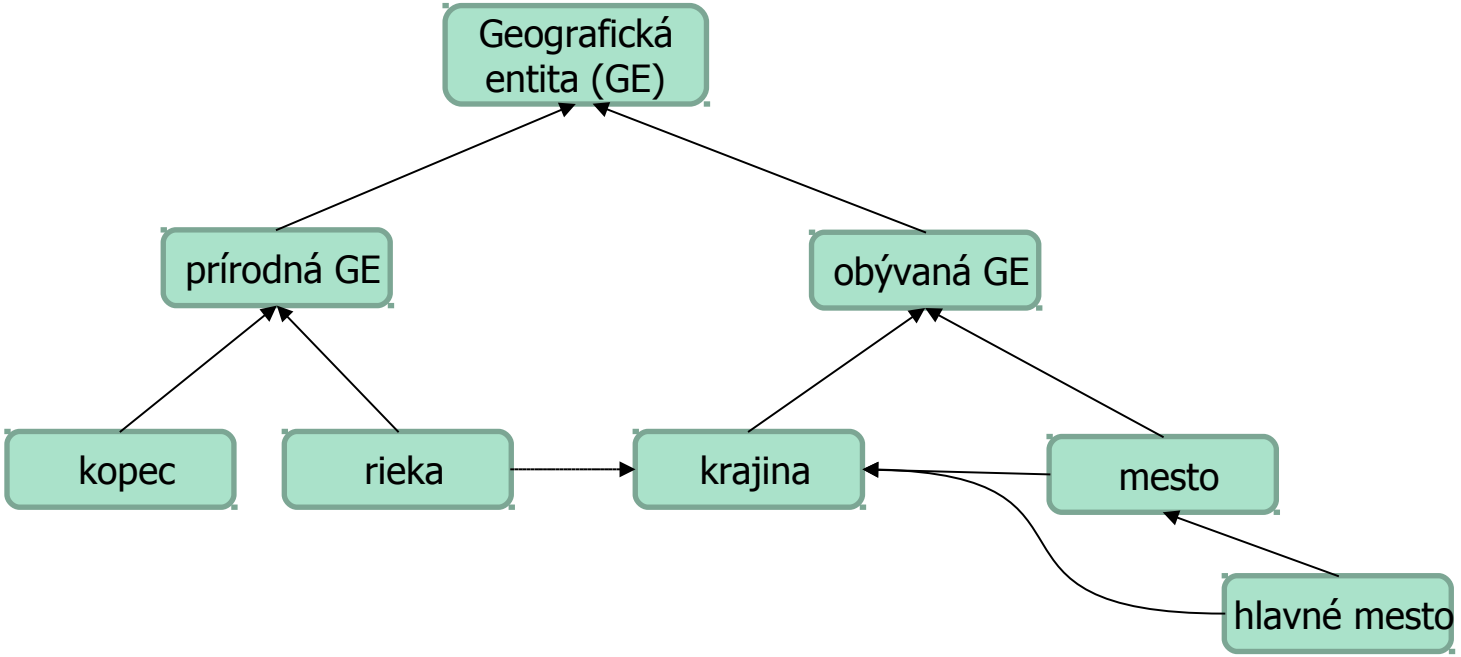
koncept
(pojem, designát)

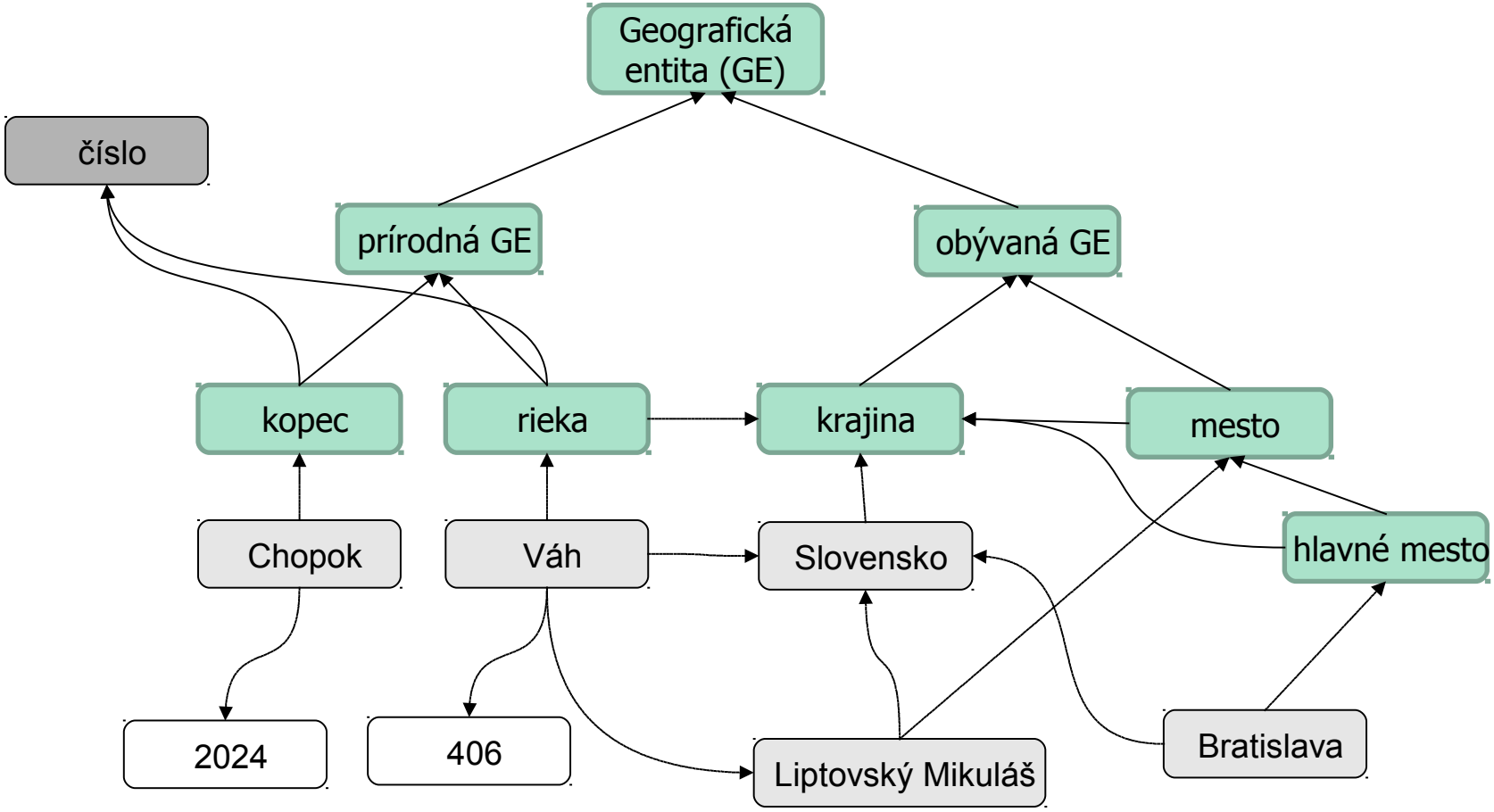
objekt
(vec, denotát)

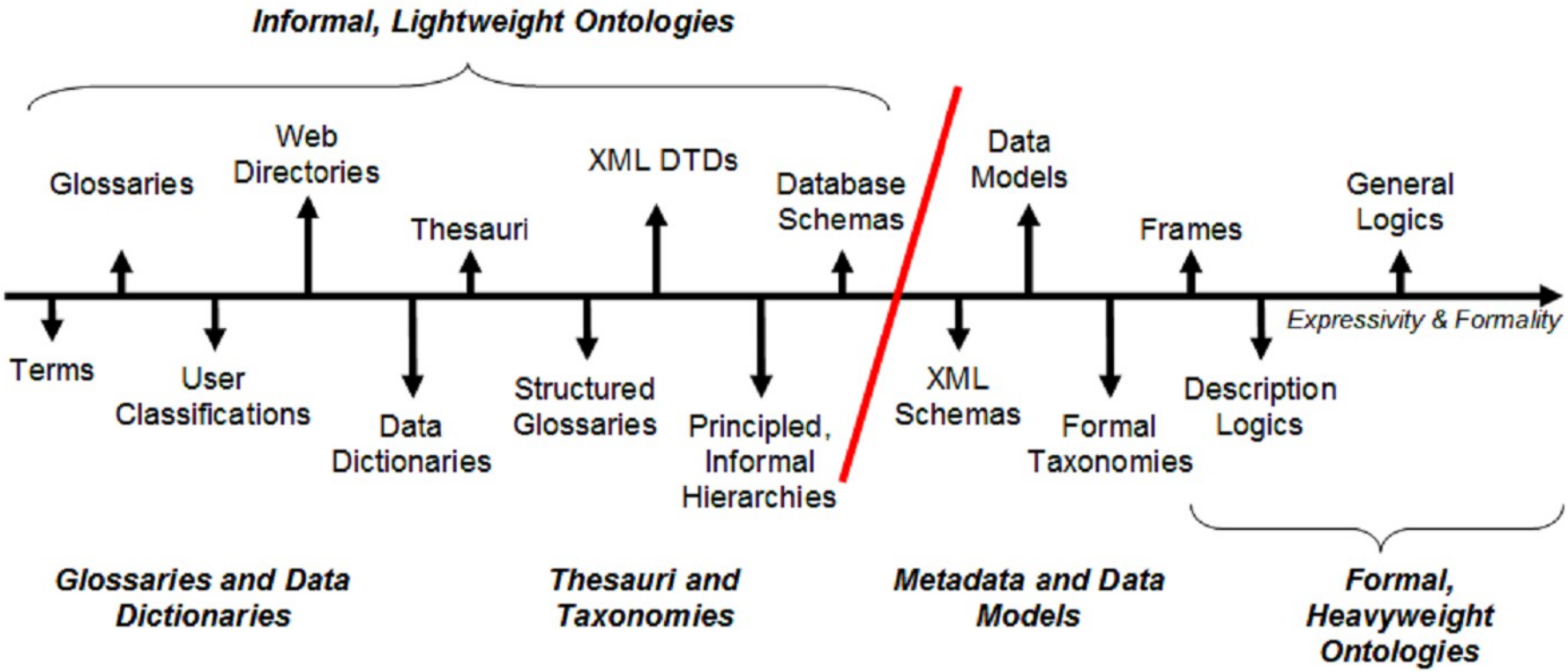


pizza

symbol
(slovo, designátor)


Geografická entita (GE)
prírodná GE
obývaná GE
kopec
rieka
krajina
mesto
hlavné mesto


Geografická entita (GE)
číslo
prírodná GE
obývaná GE
kopec
rieka
krajina
mesto
Chopok
Váh
Slovensko
hlavné mesto
2024
406
Liptovský Mikuláš
Bratislava

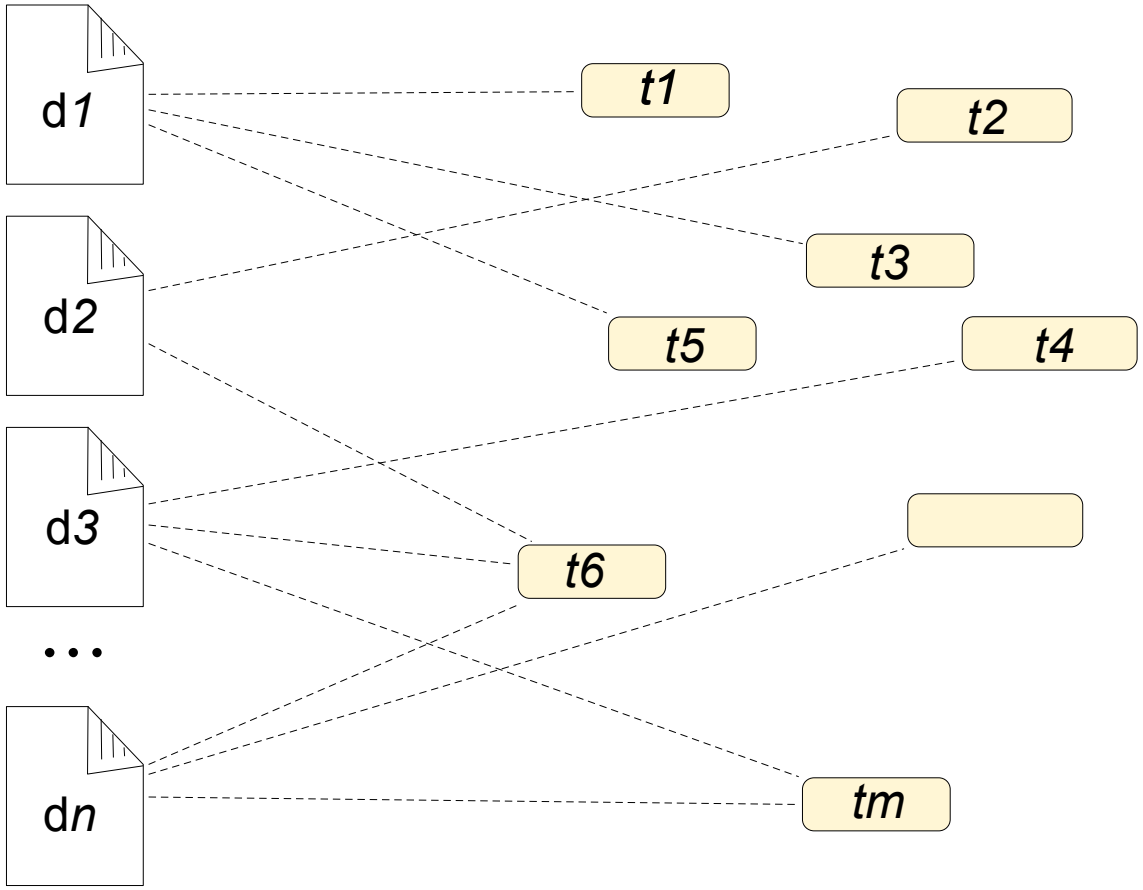


(Wong et al., 2011)

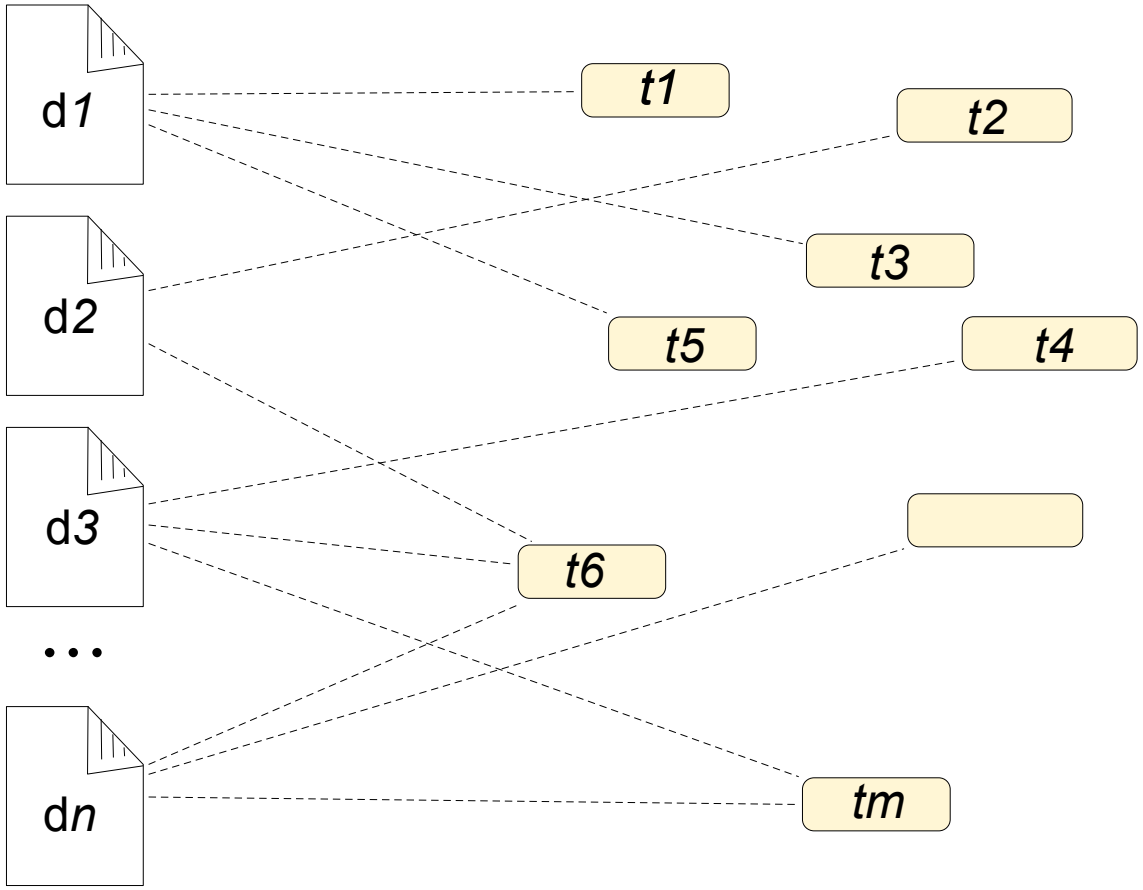
Lightweight ontológie

- Termy
  - jedno a viac-slovné výrazy
  - *pizza*
- Slovníky
  - Termy s výkladom, vysvetlením
  - *pizza* - A baked Italian dish of a thinly rolled bread dough crust typically topped before baking with tomato sauce, cheese and other ingredients such as meat, vegetables or fruit

Termy



Slovníky



t1 = a sauce made of tomatos

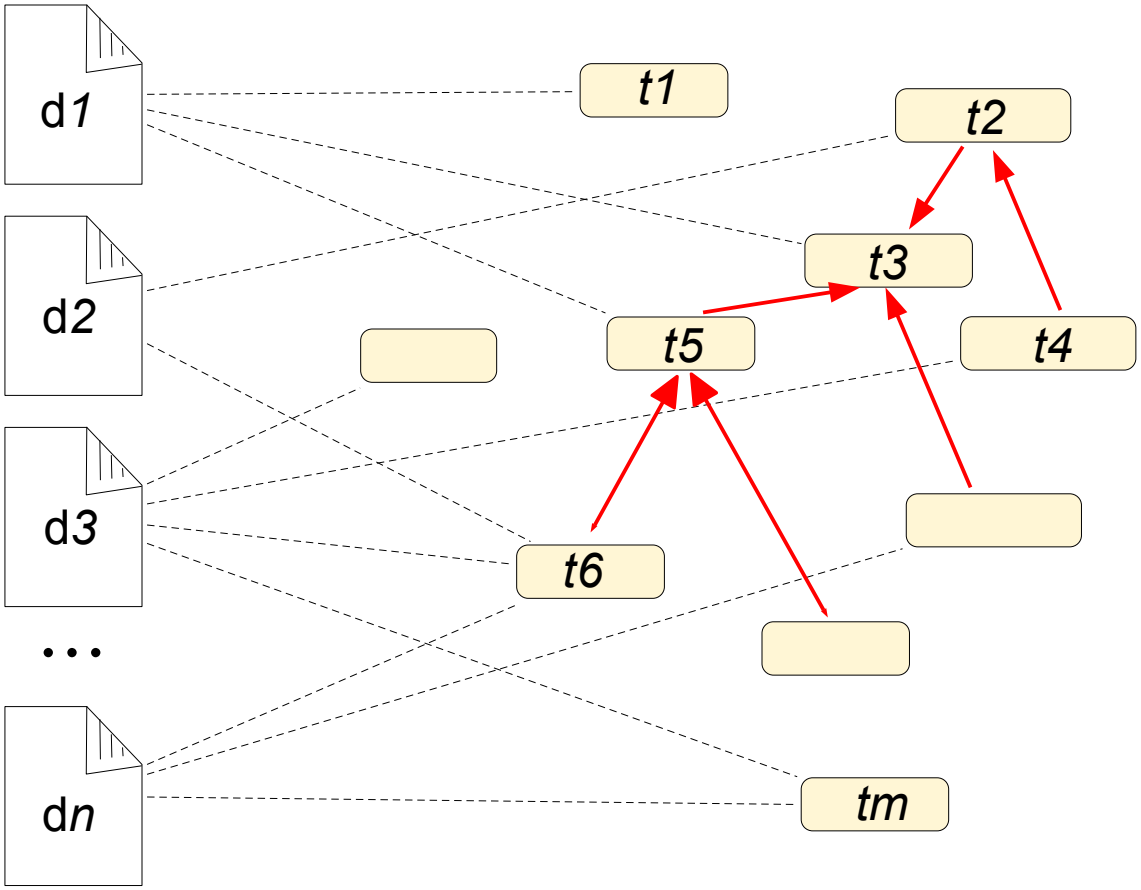
t2 = a baked italian dish …

t3 = ...

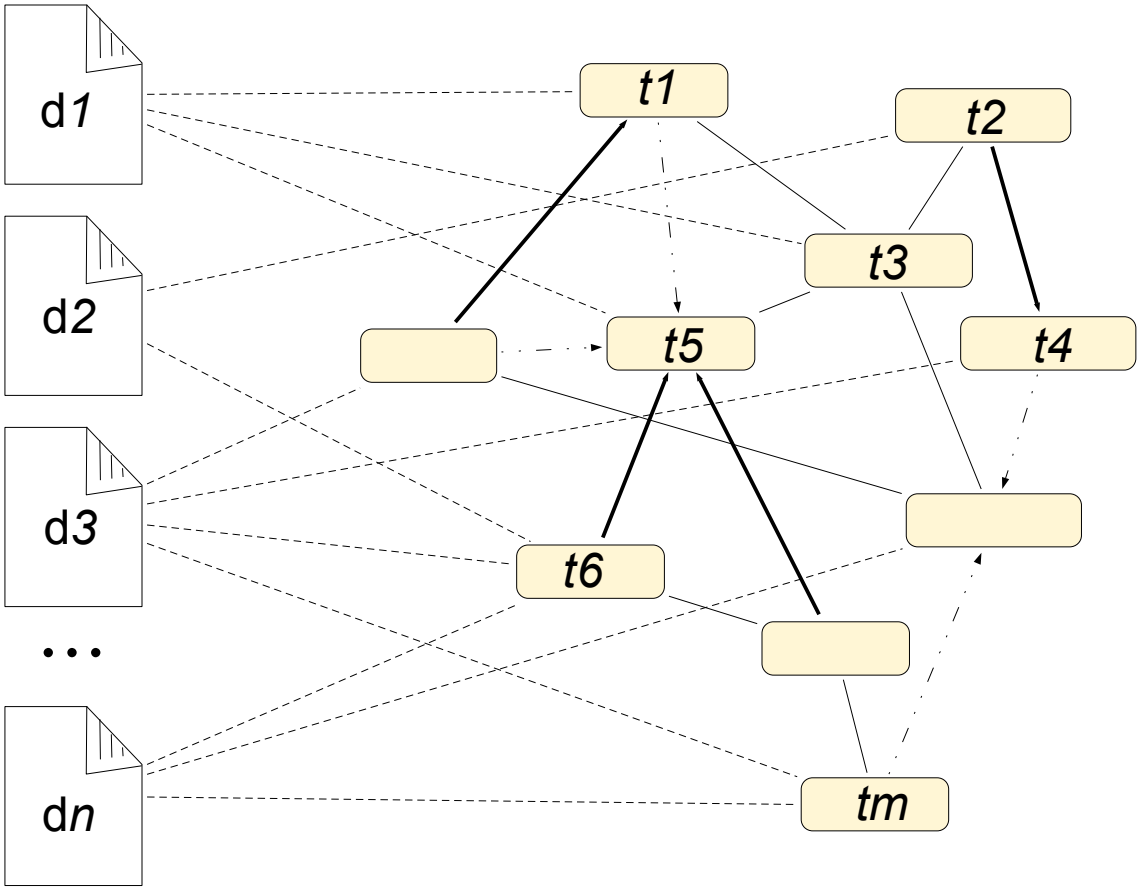
Lightweight ontológia

- Taxonómie
  - Hierarchické vzťahy typu is-a
  - *food*, *pizza*, *margherita*
- Tezaury
  - Slovníky, navyše s:
  - Hierarchickými vzťahmi: všeobecnosť (hypero-/hyponymá), časť celku (mero-/holonymá)
  - Vzťahmi ekvivalencie
  - Vzťahmi paradigmatickej podobnosti (relatedness)

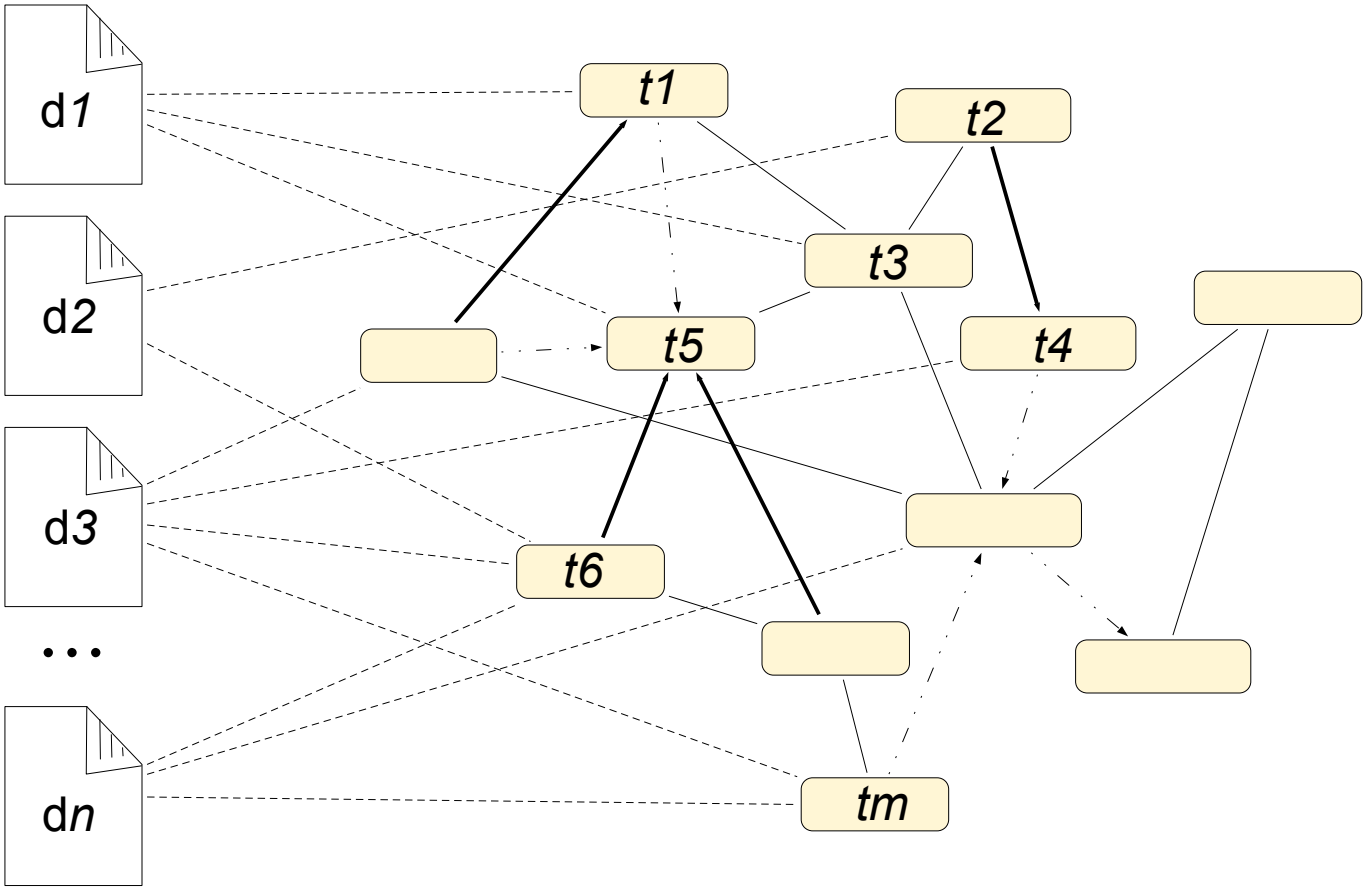
Taxonómie

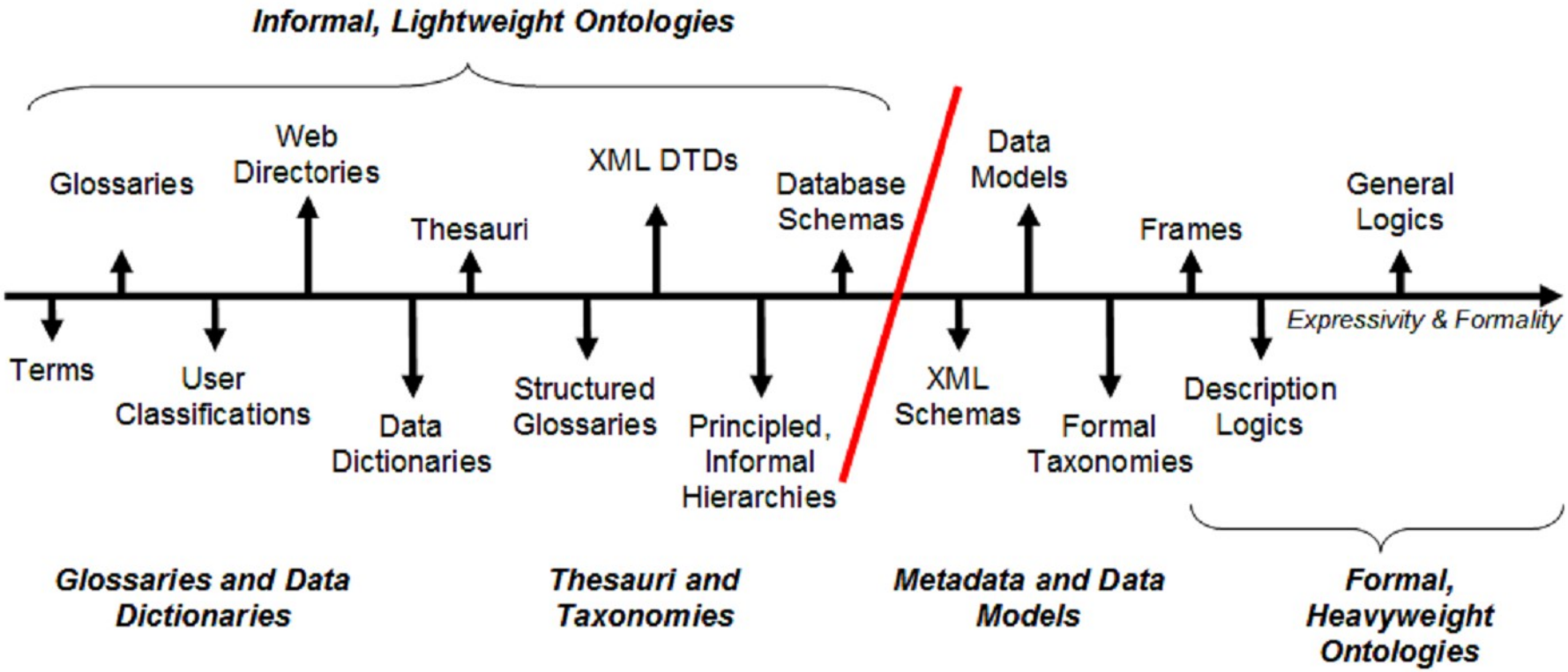


Tezaury



Ľahká ontológia





(Wong et al., 2011)

Heavyweight ontológia

$$O = \{C, \leq_C, R, \sigma_R, \leq_R, A, \sigma_A, T\}$$

- four disjoint sets C, R, A and T whose elements are called concept identifiers, relation identifiers, attribute identifiers and data types, respectively,
- a semi-upper lattice $\leq_c$ on C with top element $root_c$, called concept hierarchy or taxonomy,
- function σ_R: $R \rightarrow C^+$ called relation signature,
- a partial order $\leq_R$ on R, called relation hierarchy, where $r_1 \leq_R r_2$ implies $|\sigma_R(r_1)| = |\sigma_R(r_2)|$ and $\pi_i(\sigma_R(r_1)) \leq_c \pi_j(\sigma_R(r_2))$, for each $1 \leq i \leq |\sigma_R(r_1)|$, and
- a function σ_A: $A \rightarrow C \times T$, called attribute signature,
- a set T of datatypes such as strings, integers, etc.

(Cimiano, 2006)

Heavyweight ontológia

- Definition 2 (Domain and Range)
- Definition 3 ($\mathscr{L}$-axiom System)
- Definition 4 (Lexicon)
- Definition 5 (Knowledge Base (KB))
- Definition 6 (Instance Lexicon)
- Definition 7 (Extension)
- Definition 8 (Intension)

(Cimiano, 2006)

Koncept – trojica

(i(c), [c], Ref_C(c) **)**

- i(c) – intenzionálny opis
- [c] – extenzionálny opis
- Ref_C(c) – lexikálna realizácia v korpuse

Heavyweight ontológia

$\forall x(country(x) \rightarrow \exists y\ capital_of(y,x) \wedge \forall z(capital_of(z,x) \rightarrow y=z))$

disjoint(river, mountain)

capital_of $\leq_R$ located_in

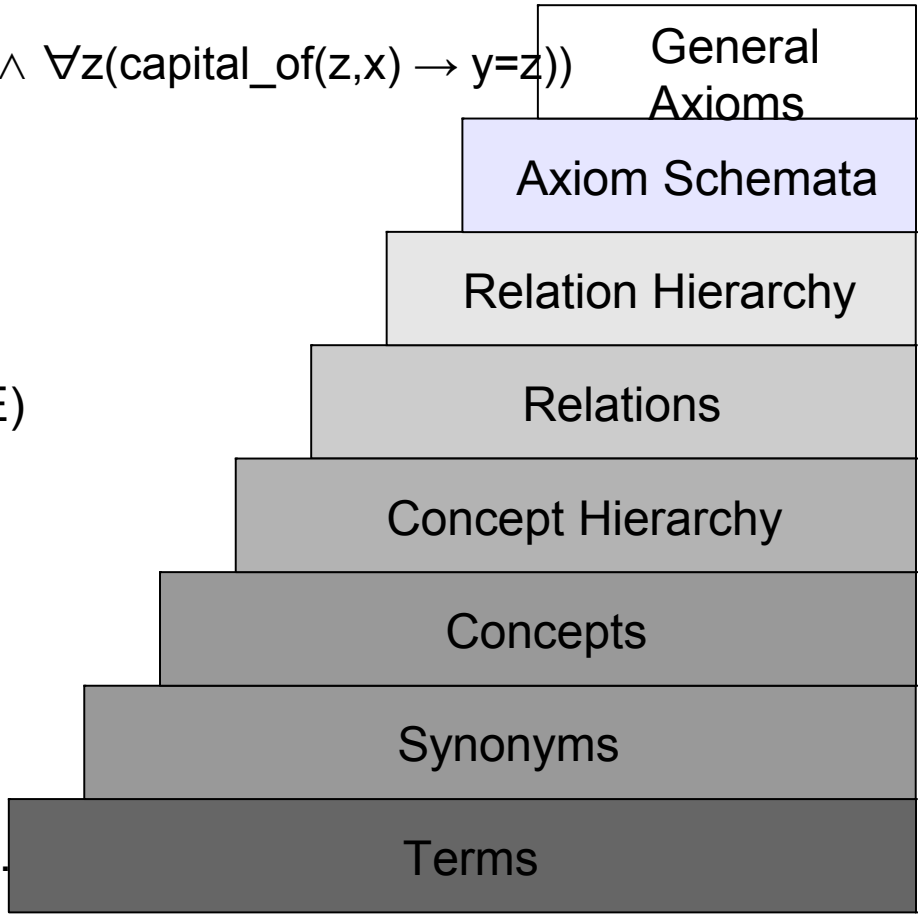
flow_through(dom:river, range:GE)

capital $\leq_C$ city, city $\leq_C$ GE

c:= country := <i(c), ||c||, $Ref_C(c)$>

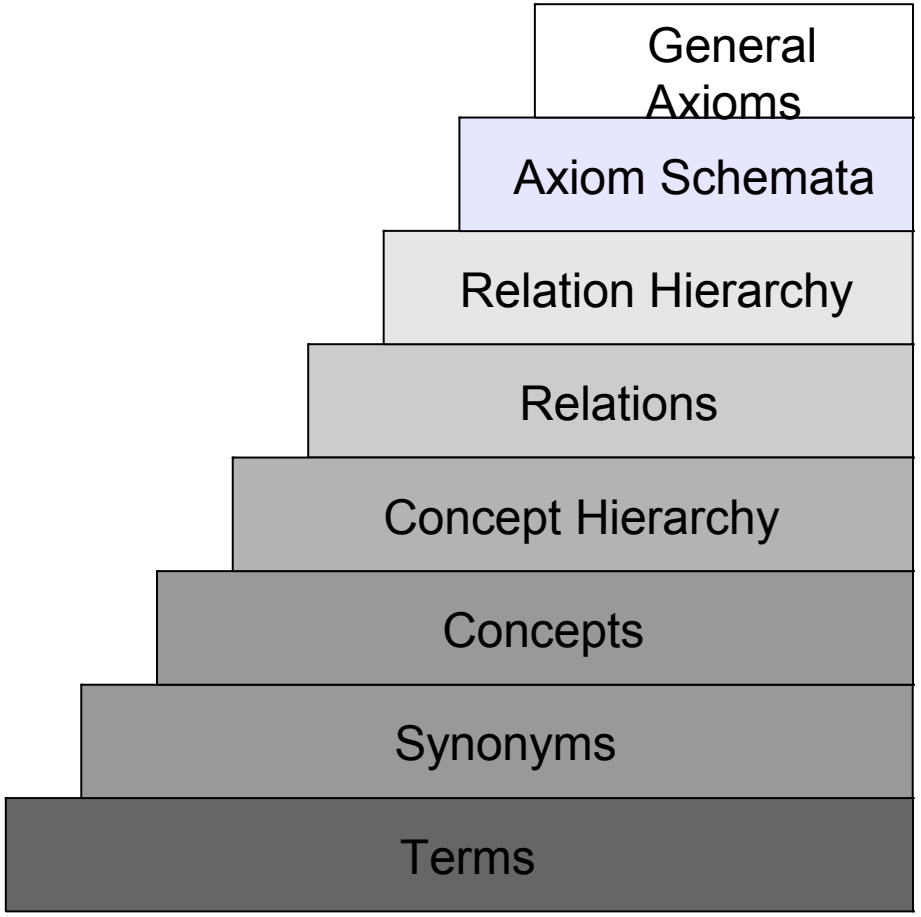
{country, nation}

river, coutry, nation, city, capital,...



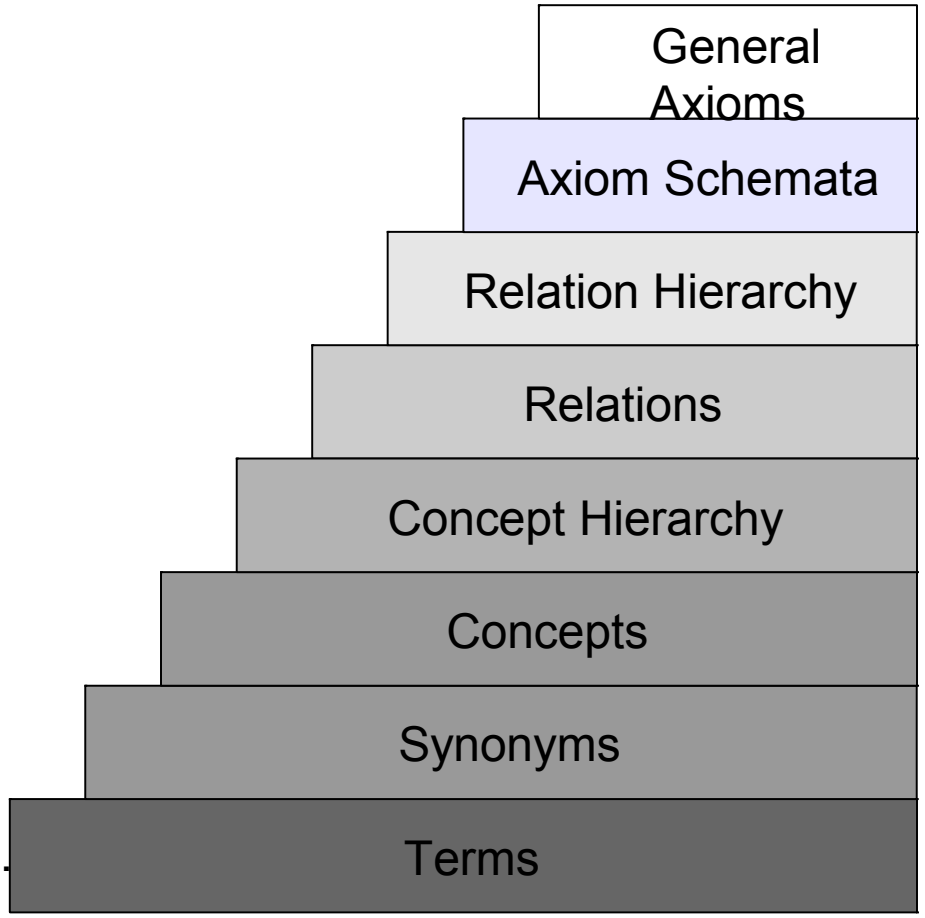
(Cimiano, 2006)

Heavyweight ontológia



(Cimiano, 2006)

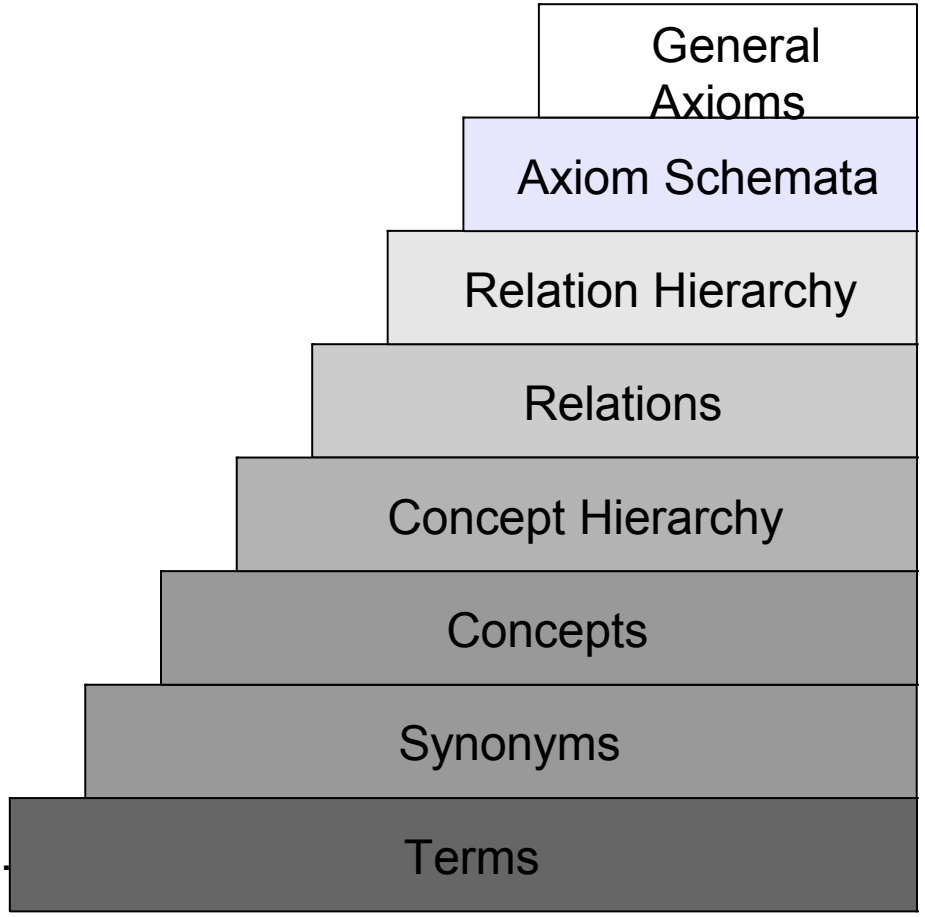
Heavyweight ontológia



river, coutry, nation, city, capital,...

(Cimiano, 2006)

Heavyweight ontológia

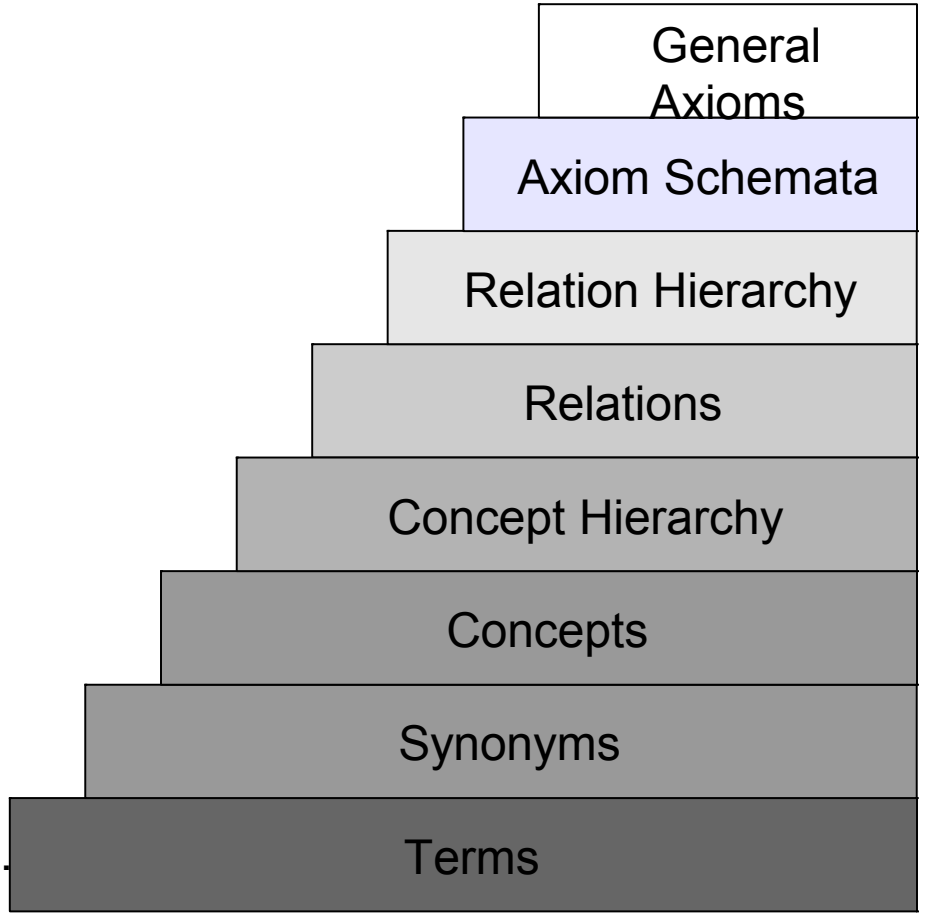


{country, nation}

river, coutry, nation, city, capital,...

(Cimiano, 2006)

Heavyweight ontológia



c:= country := <i(c), ||c||, $Ref_C(c)$>

{country, nation}

river, coutry, nation, city, capital,...

(Cimiano, 2006)

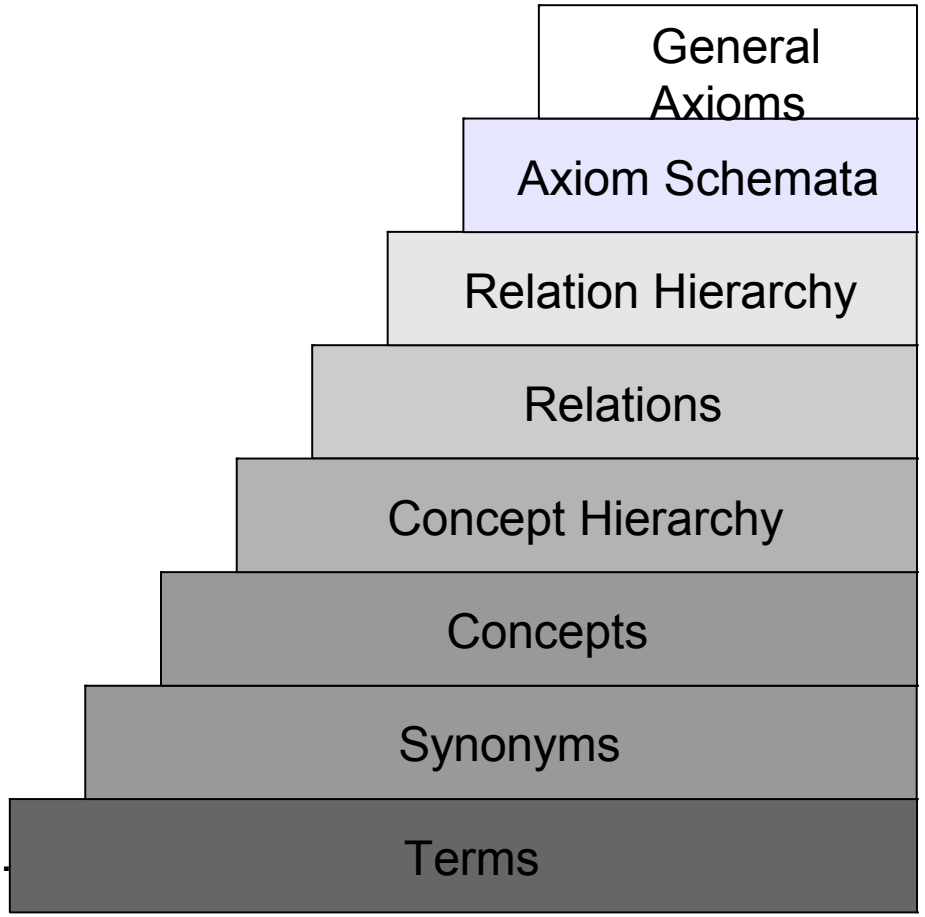
Heavyweight ontológia

capital $\leq_c$ city, city $\leq_c$ GE

c:= country := <i(c), ||c||, $Ref_c(c)$>

{country, nation}

river, coutry, nation, city, capital,...



(Cimiano, 2006)

Heavyweight ontológia

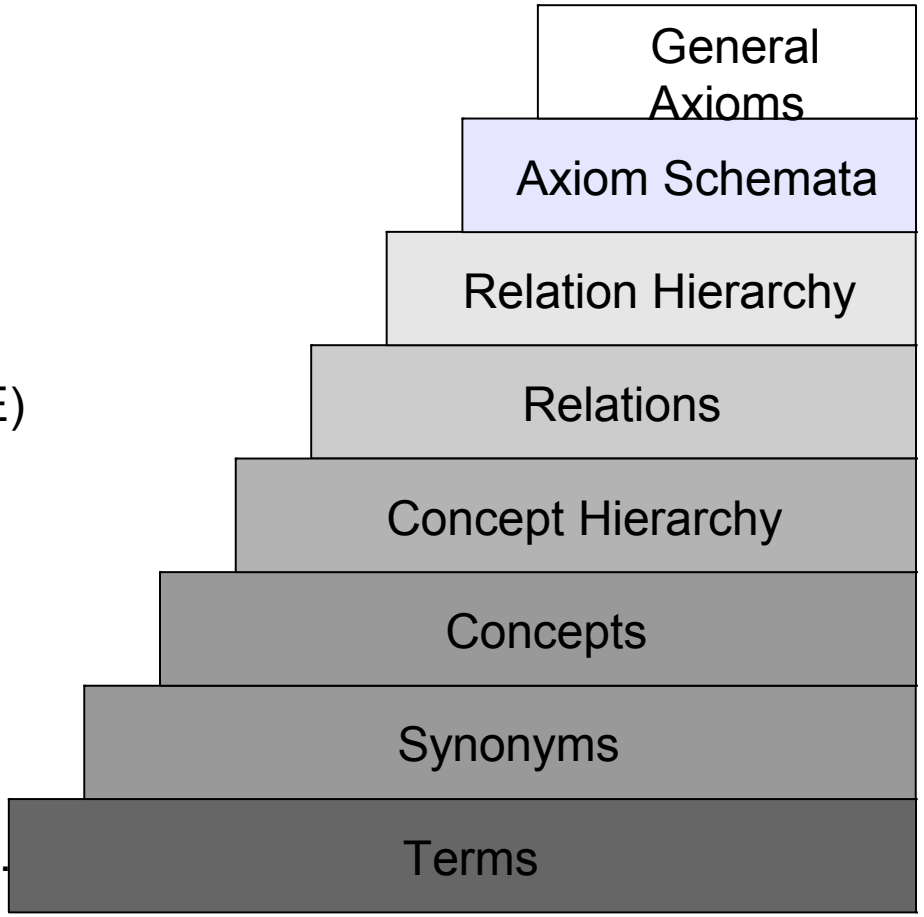
flow_through(dom:river, range:GE)

capital $\leq_C$ city, city $\leq_C$ GE

c:= country := <i(c), ||c||, $Ref_C(c)$>

{country, nation}

river, coutry, nation, city, capital,...



(Cimiano, 2006)

Heavyweight ontológia

capital_of $\leq_R$ located_in

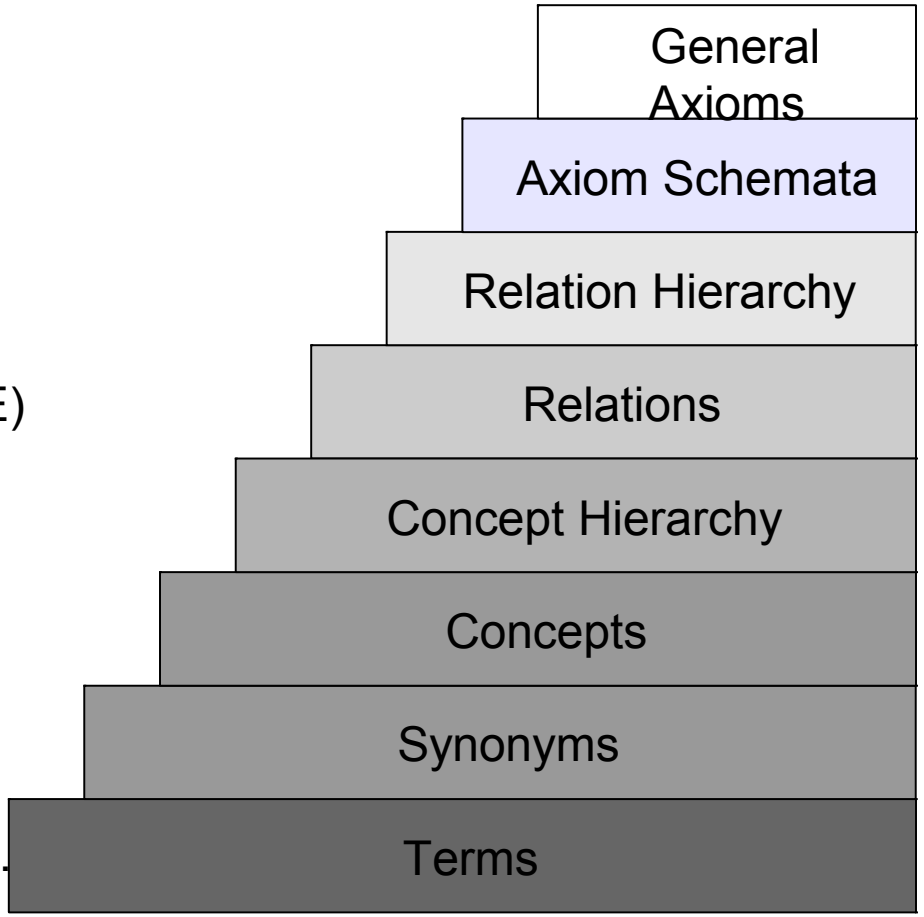
flow_through(dom:river, range:GE)

capital $\leq_C$ city, city $\leq_C$ GE

c:= country := <i(c), ||c||, $Ref_C(c)$>

{country, nation}

river, coutry, nation, city, capital,...



(Cimiano, 2006)

Heavyweight ontológia

disjoint(river, mountain)

capital_of $\leq_R$ located_in

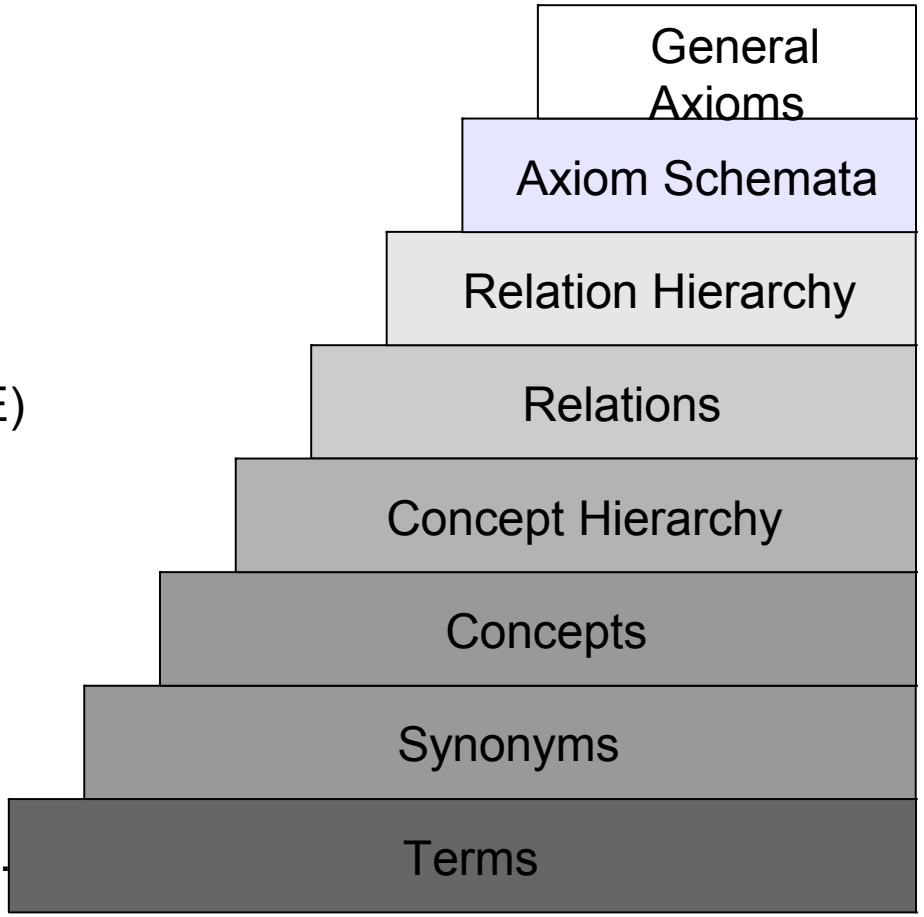
flow_through(dom:river, range:GE)

capital $\leq_C$ city, city $\leq_C$ GE

c:= country := <i(c), ||c||, $Ref_C(c)$>

{country, nation}

river, coutry, nation, city, capital,...



(Cimiano, 2006)

Heavyweight ontológia

$\forall x(country(x) \rightarrow \exists y\ capital_of(y,x) \wedge \forall z(capital_of(z,x) \rightarrow y=z))$

disjoint(river, mountain)

capital_of $\leq_R$ located_in

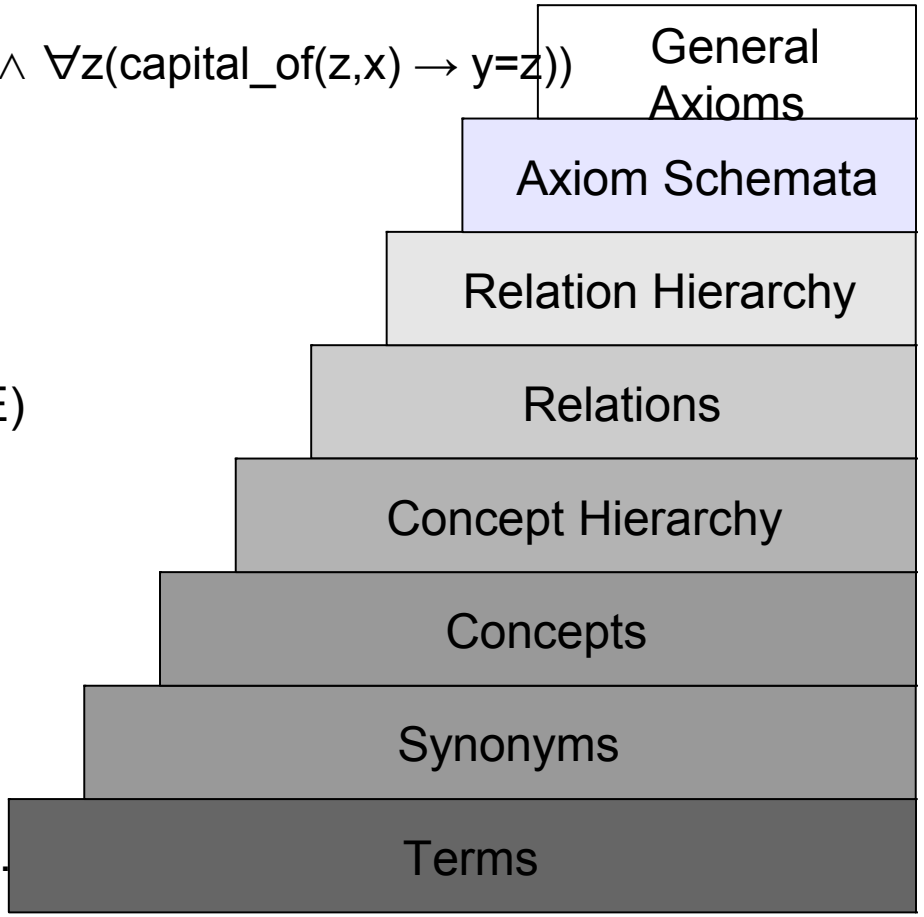
flow_through(dom:river, range:GE)

capital $\leq_C$ city, city $\leq_C$ GE

c:= country := <i(c), ||c||, $Ref_C(c)$>

{country, nation}

river, coutry, nation, city, capital,...



(Cimiano, 2006)

Porovnanie	light	vs.	heavy
• expresivita	--		+
• konzistentnosť	--		+
• zostrojiteľnosť	+		--
• rýchlosť (použitia)	+		--
• transparentnosť	--		+
• inštanciovateľnosť	+		--
• interoperabilita	--		+

Štandardy pre zápis dát so sémantikou

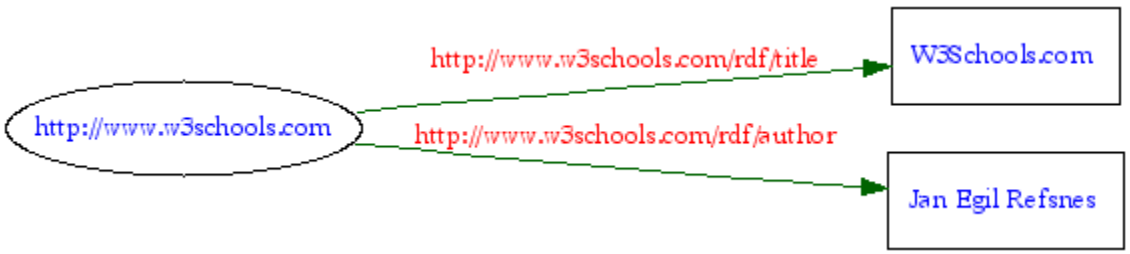
Štandardy

- RDF
  - Resource Description Framework
- RDFS
  - RDF Schema
- OWL
  - Web Ontology Language
  - OWL Lite
  - OWL DL
  - OWL Full

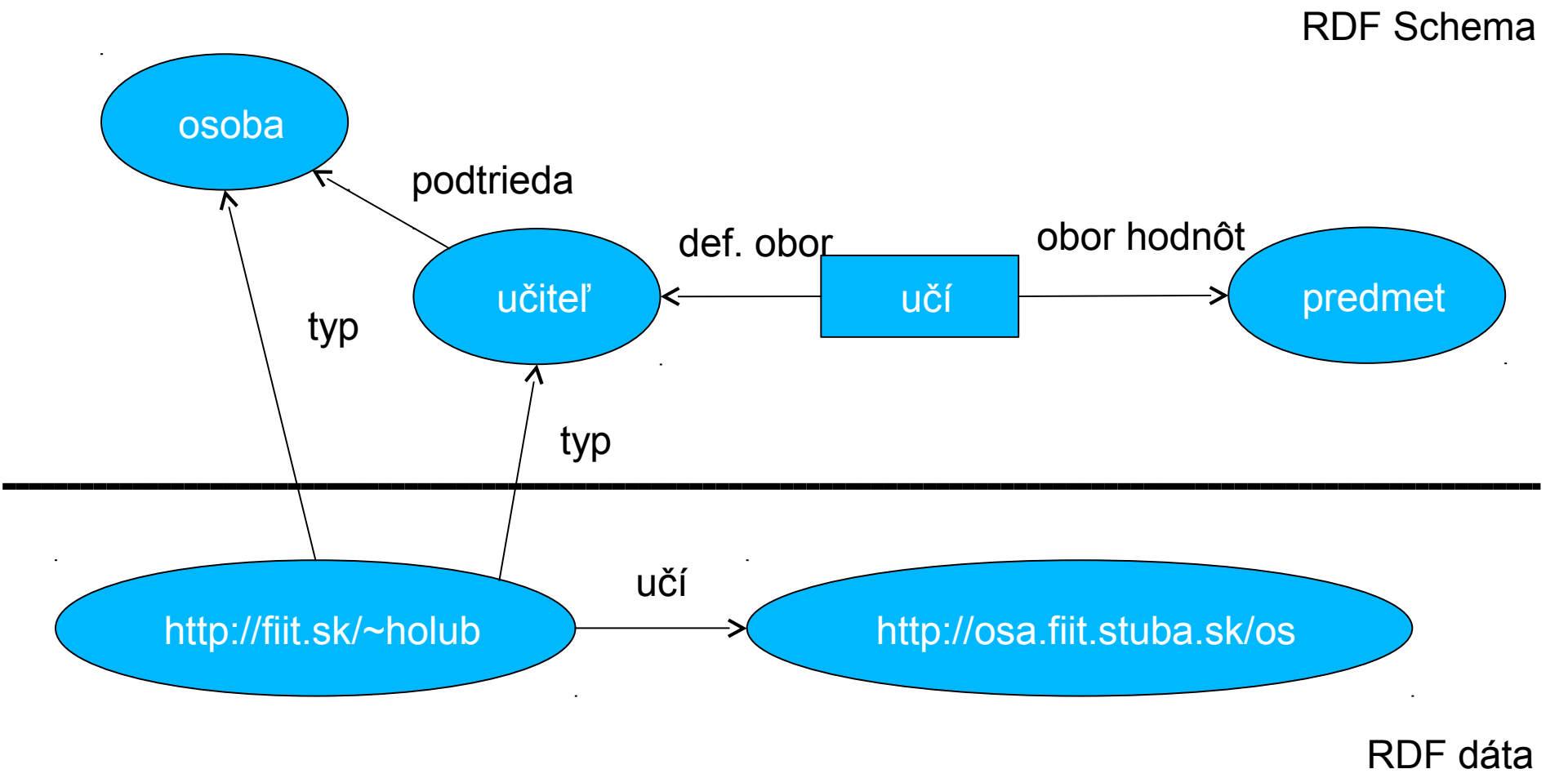
RDF

- Subjekt-predikát-objekt

```
<rdf:RDF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:si="http://www.w3schools.com/rdf/">
   <rdf:Description rdf:about="http://www.w3schools.com">
      <si:title>W3Schools.com</si:title>
      <si:author>Jan Egil Refsnes</si:author>
   </rdf:Description>
</rdf:RDF>
```



RDF + RDFS


RDF Schema
osoba
podtrieda
učiteľ
def. obor
učí
obor hodnôt
predmet
typ
typ
http://fiit.sk/~holub
učí
http://osa.fiit.stuba.sk/os
RDF dáta

RDF(S) – problémy

- doména a rozsah sa nedajú obmedziť
- nedá sa obmedziť existencia a kardinalita
  - každá osoba má práve jednu matku typu osoba
- nedajú sa definovať tranzitívne, inverzné, symetrické tvrdenia
- príliš nepodporuje odvodzovanie, dokazovanie

Štandardy – OWL

OWL Lite

(pod)triedy, indivíduá
(pod)vlastnosti, doména, rozsah
konjunkcia
(ne)rovnosť

RDF Schema

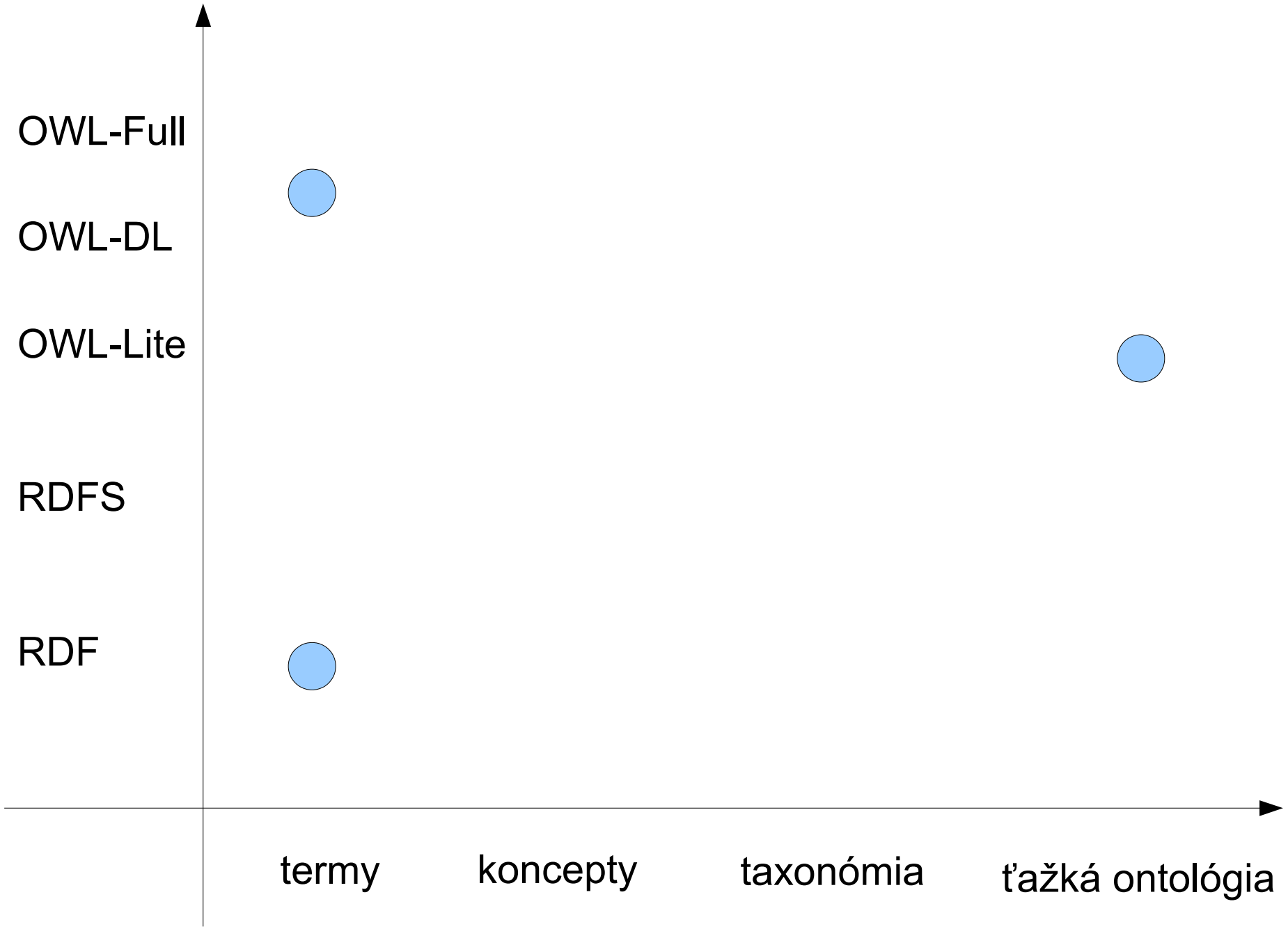
kardinalita 0/1
dátové typy
inverzia, tranzitivita, symetria

OWL DL

negácia, disjunkcia
plná kardinalita
vymenované typy

OWL full

meta triedy

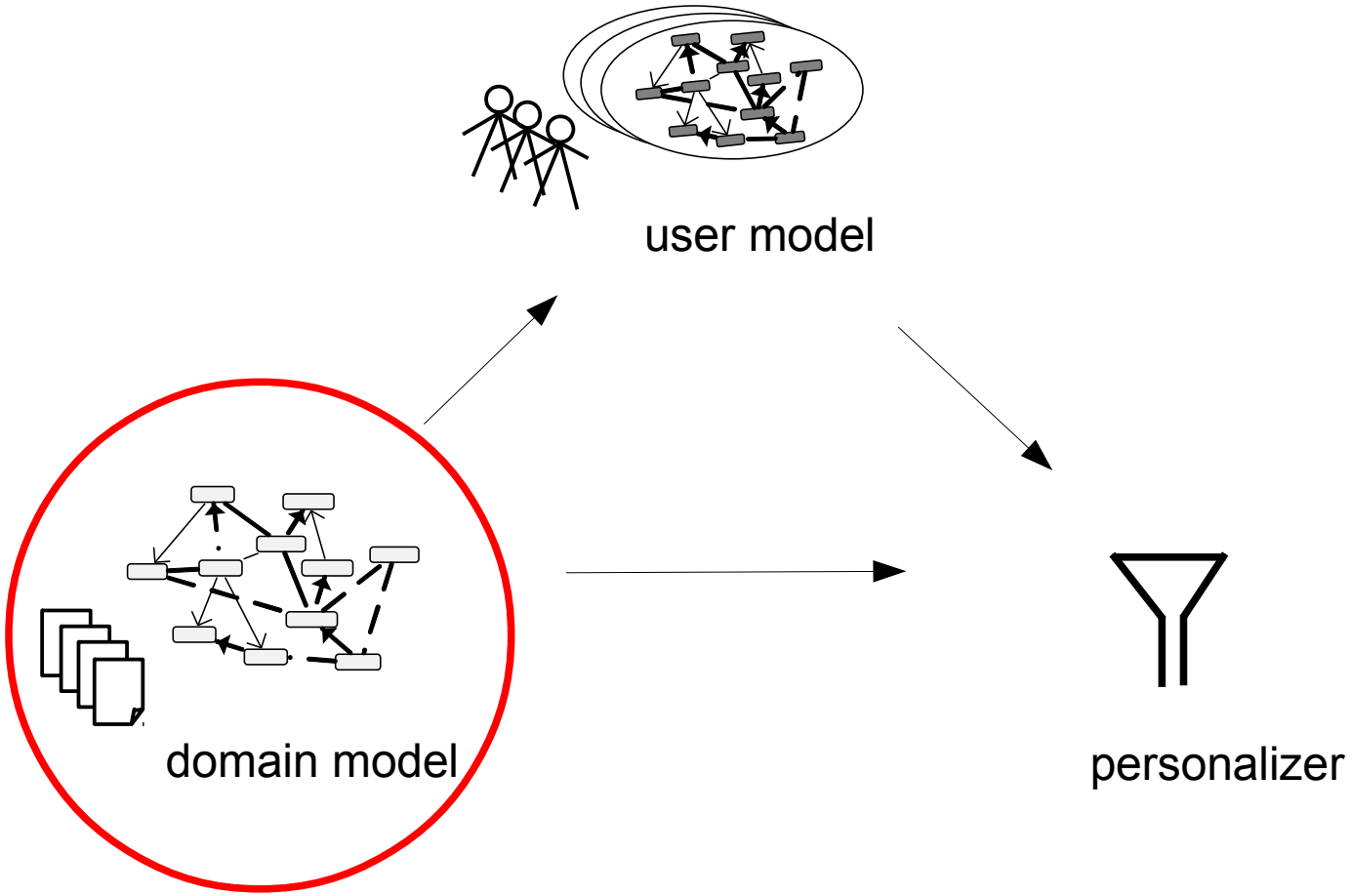

OWL-Full
OWL-DL
OWL-Lite
RDFS
RDF
termy
koncepty
taxonómia
ťažká ontológia

Doménový model

- reprezentácia problémovej oblasti (domény) slúžiaca na opis zdrojov, ktoré sú predmetom ďalšieho, zvyčajne pokročilého spracovania (napr. odporúčanie, inteligentné vyhľadávanie)

- ľahká sémantika vs. ťažká sémantika

Modelovanie domény – úlohy

- reprezentácia
  - aká granularita
  - aká expresivita
- automatizácia tvorby
  - spracovanie obsahu, analýza prepojení, používania
  - ontology learning, NLP, NER, LSA, ...

Personalised Web-based System