

TOOLS AND TECHNIQUES

*“FEEDING RECOMMENDER SYSTEMS
WITH LINKED OPEN DATA”*

Tommaso Di Noia

t.dinoia@poliba.it

Politecnico di Bari

ITALY

Recommender Systems

A definition

Recommender Systems (RSs) are *software tools and techniques providing suggestions for items to be of use to a user.*

[F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. **Recommender Systems Handbook**. Springer, 2011.]

Input Data:

A set of users $U = \{u_1, \dots, u_M\}$

A set of items $X = \{x_1, \dots, x_N\}$

The rating matrix $R = [r_{u,i}]$

Problem Definition:

Given user u and target item i

Predict the rating $r_{u,i}$



The rating matrix

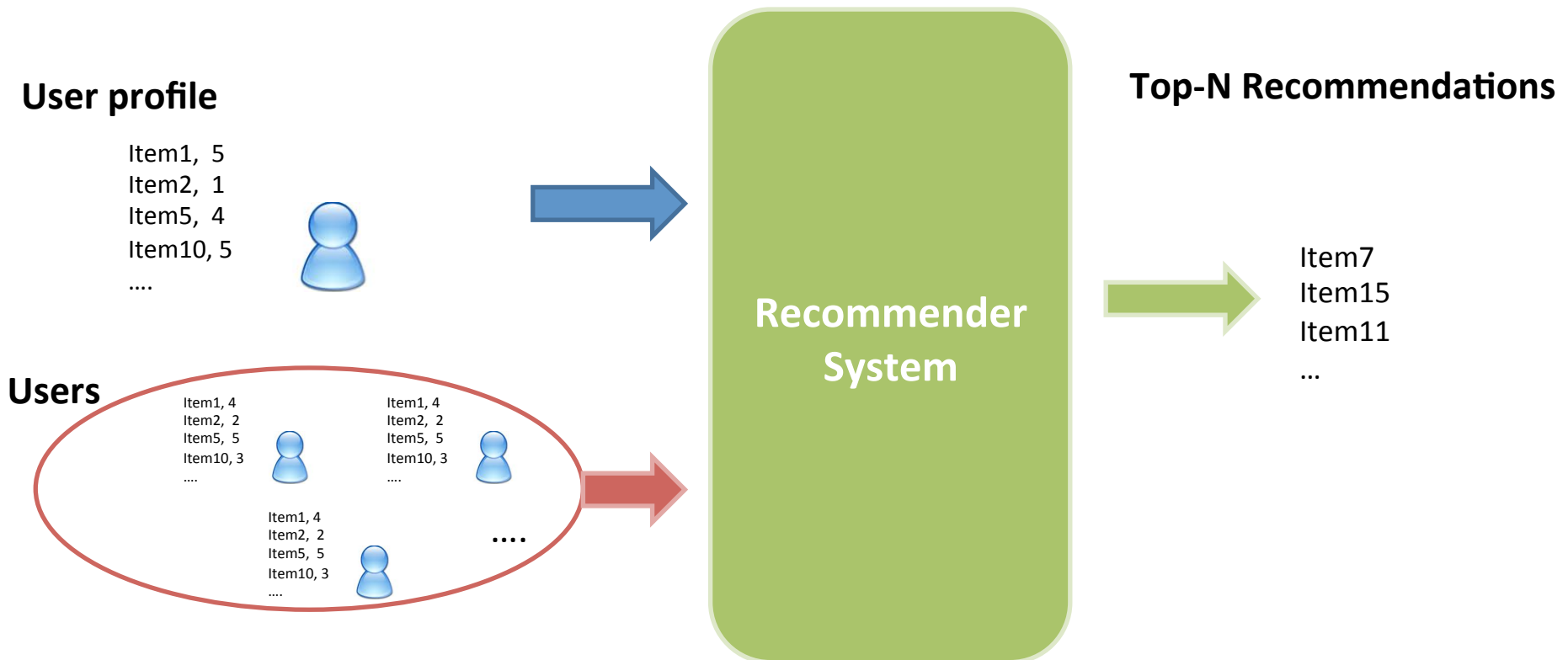
	The Matrix	Titanic	I love shopping	Argo	Love Actually	The hangover
Tommaso	5	1	2	4	3	??
Enrico	2	4	5	3	5	2
Sean	4	3	2	4	1	3
Natasha	3	5	1	5	2	4
Valentina	4	4	5	3	5	2

The rating matrix (in the real world)

	The Matrix	Titanic	I love shopping	Argo	Love Actually	The hangover
Tommaso	5			4	3	??
Enrico	2	4	5		5	
Sean		3		4		3
Natasha	3	5		5	2	
Valentina	4	4	5		5	2

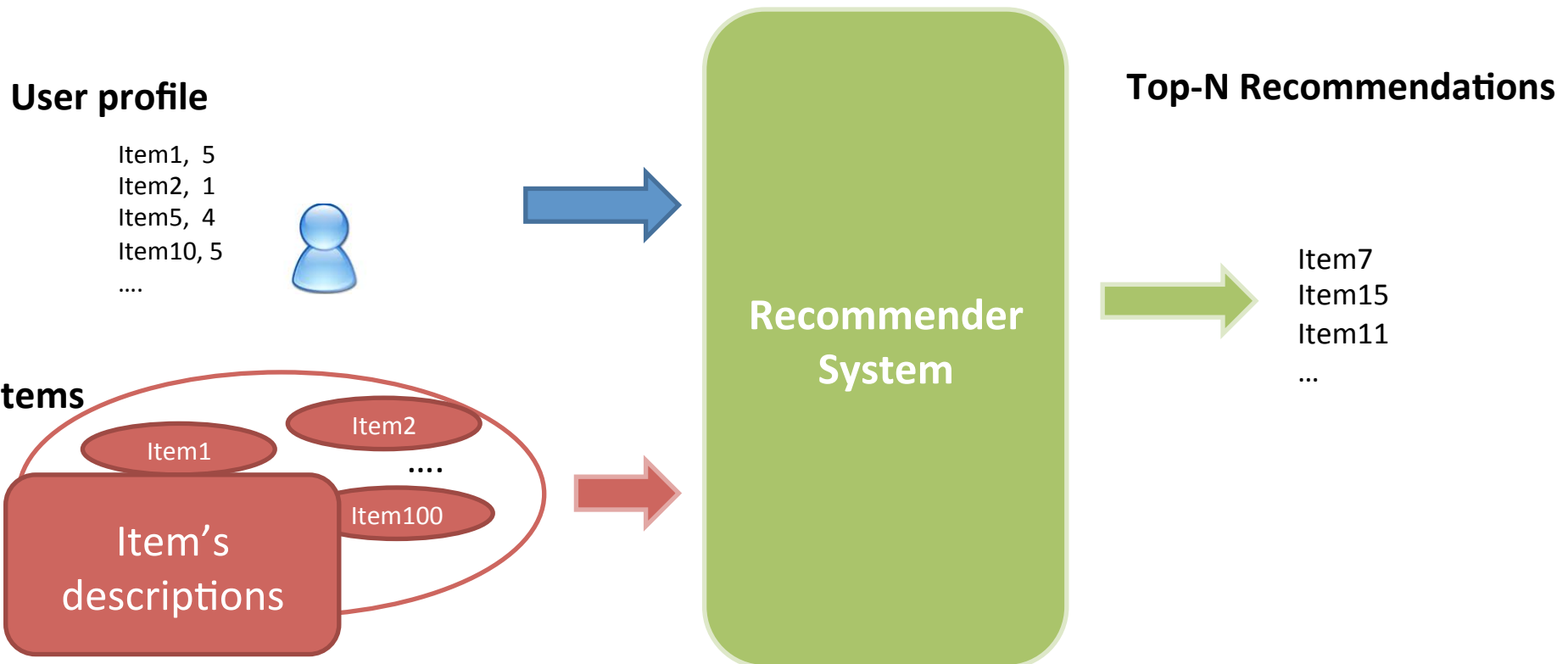
Collaborative Recommender Systems

Collaborative RSs recommend items to a user by identifying other users with a similar profile



Content-based Recommender Systems

CB-RSs recommend items to a user based on their description and on the profile of the user's interests



Knowledge-based Recommender Systems

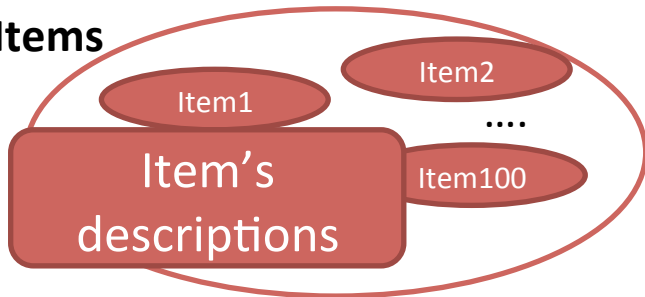
KB-RSs recommend items to a user based on their description and domain knowledge encoded in a knowledge base

User profile

Item1, 5
Item2, 1
Item5, 4
Item10, 5
....



Items



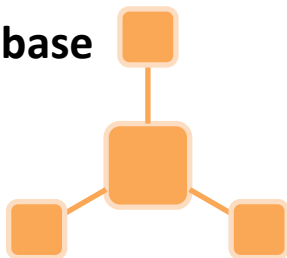
Recommender
System

Top-N Recommendations

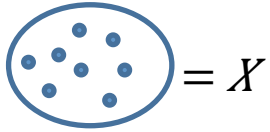


Item7
Item15
Item11
...

Knowledge-base



User-based Collaborative Recommendation



$= X$

Pearson's correlation coefficient

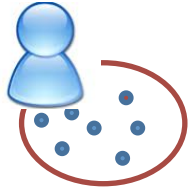
$$sim(u_i, u_j) = \frac{\sum_{x \in X} (r_{u_i, x} - \bar{r}_{u_i}) * (r_{u_j, x} - \bar{r}_{u_j})}{\sqrt{\sum_{x \in X} (r_{u_i, x} - \bar{r}_{u_i})^2} * \sqrt{\sum_{x \in X} (r_{u_j, x} - \bar{r}_{u_j})^2}}$$

Rate prediction

$$\hat{r}(u_i, x) = \bar{r}_{u_i} + \sum_{u_j \in X} sim(u_i, u_j) * (\bar{r}_{u_j} - \bar{r}_{u_i})$$

	The Matrix	Titanic	I love shopping	Argo	Love Actually	The hangover
Tommaso	5	1	2	4	3	??
Enrico	2	4	5	3	5	2
Sean	4	3	2	4	1	3
Natasha	3	5	1	5	2	4
Valentina	4	4	5	3	5	2

Item-based Collaborative Recommendation



$$= X_{u,i}$$

Cosine Similarity

$$sim(x_i, x_j) = \frac{x_i \cdot x_j}{|x_i| \cdot |x_j|} = \frac{\sum_u r_{u,i} \cdot r_{u,j}}{\sqrt{\sum_u r_{u,i}^2} \cdot \sqrt{\sum_u r_{u,j}^2}}$$

Adjusted Cosine Similarity

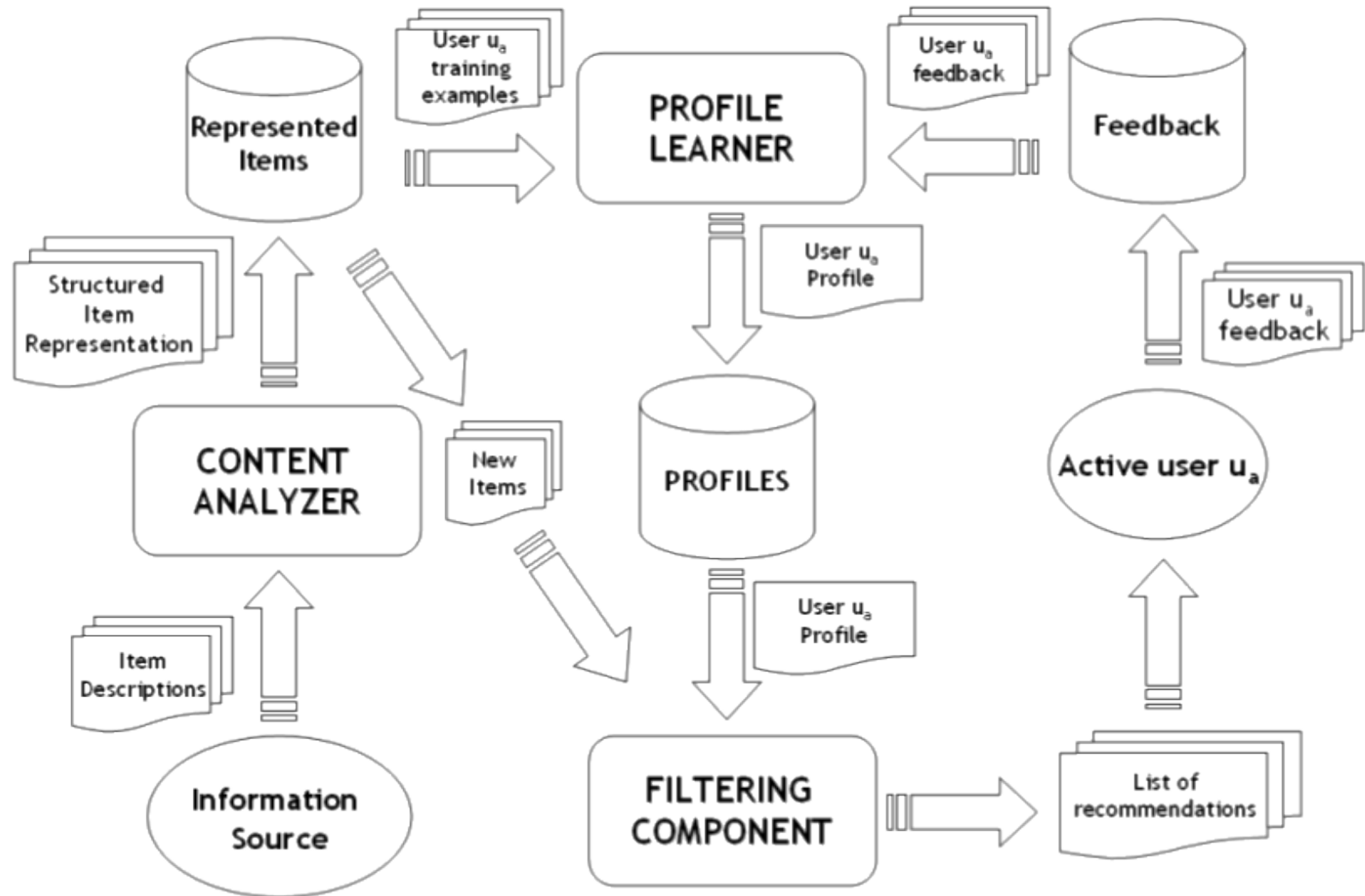
$$sim(x_i, x_j) = \frac{\sum_u (r_{u,i} - \bar{r}_u) \cdot (r_{u,j} - \bar{r}_u)}{\sqrt{\sum_u (r_{u,i} - \bar{r}_u)^2} \cdot \sqrt{\sum_u (r_{u,j} - \bar{r}_u)^2}}$$

Rate prediction

$$\hat{r}(u, i) = \sum_{x \in X_{u,i}} sim(x, x')$$

	The Matrix	Titanic	I love shopping	Argo	Love Actually	The hangover
Tommaso	5	1	2	4	3	??
Enrico	2	4	5	3	5	2
Sean	4	3	2	4	1	3
Natasha	3	5	1	5	2	4
Valentina	4	4	5	3	5	2

Content-Based Recommender Systems



Content-Based Recommender Systems

- Items are described in terms of attributes/features
- A finite set of values is associated to each feature
- Items representation is a (Boolean) vector

Content-Based Recommender Systems

- Compute similarity between items

Jaccard similarity

$$sim(x \downarrow i, x \downarrow j) = |x \downarrow i \cap x \downarrow j| / |x \downarrow i \cup x \downarrow j|$$

Content-Based Recommender Systems

- Compute similarity between items

Cosine similarity

$$\text{sim}(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}$$

Content-Based Recommender Systems

- Compute similarity between items

Cosine similarity and TF-IDF

$TF(v, x) = \# \text{ } v \text{ appears in the description of } x$

$IDF(v) = \log |X| / \# v \text{ appears in } X$

*$x = (TF(v_1, x) * IDF(v_1), \dots, TF(v_n, x) * IDF(v_n))$*

*$sim(x_i, x_j) = x_i \cdot x_j / |x_i| * |x_j|$*

Content-Based Recommender Systems

Rate prediction

$$r(u, i, x') = \frac{\sum_{x \in X} \text{sim}(x, x') * r(x, u, i)}{\sum_{x \in X} \text{sim}(x, x')}$$

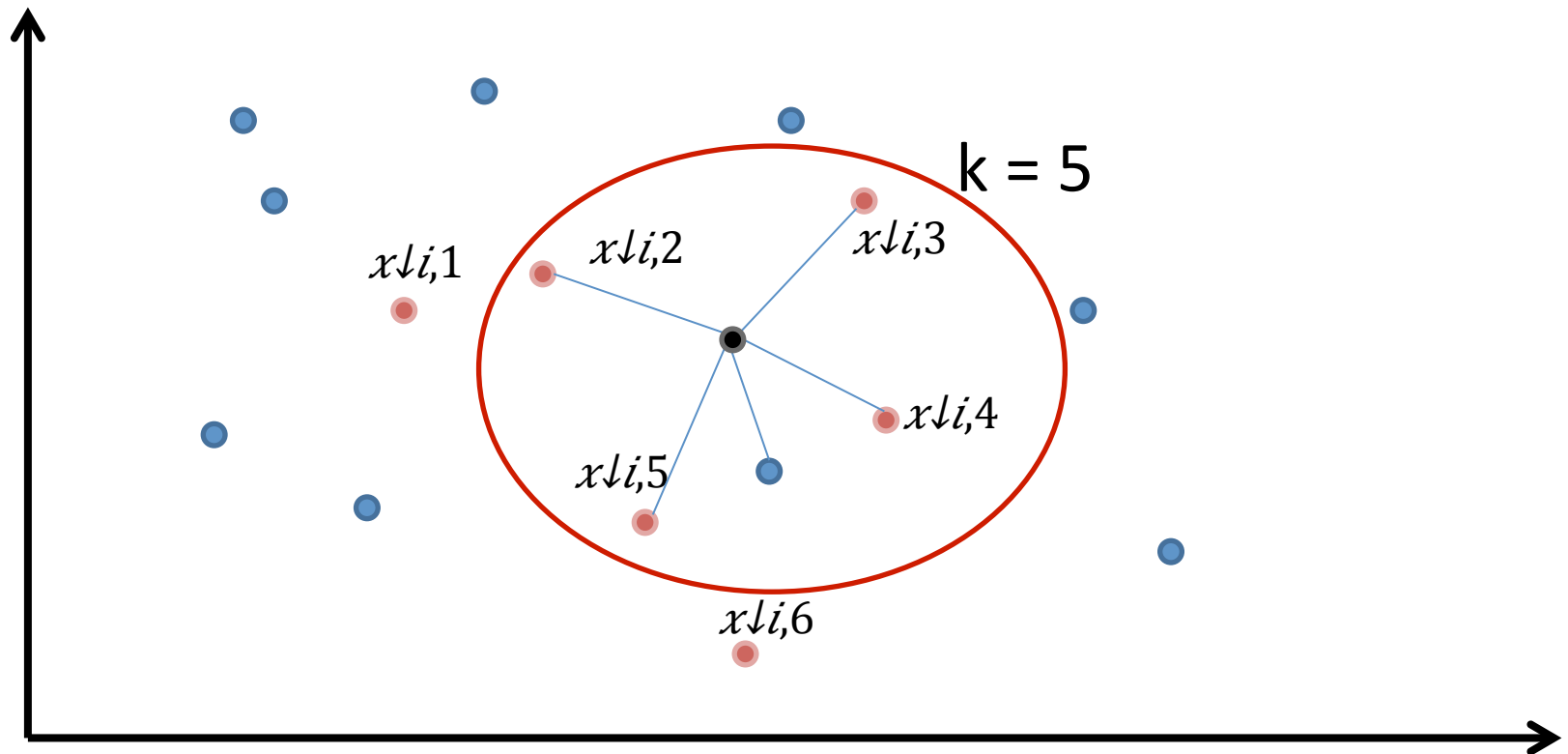
=

Content-Based Recommender Systems

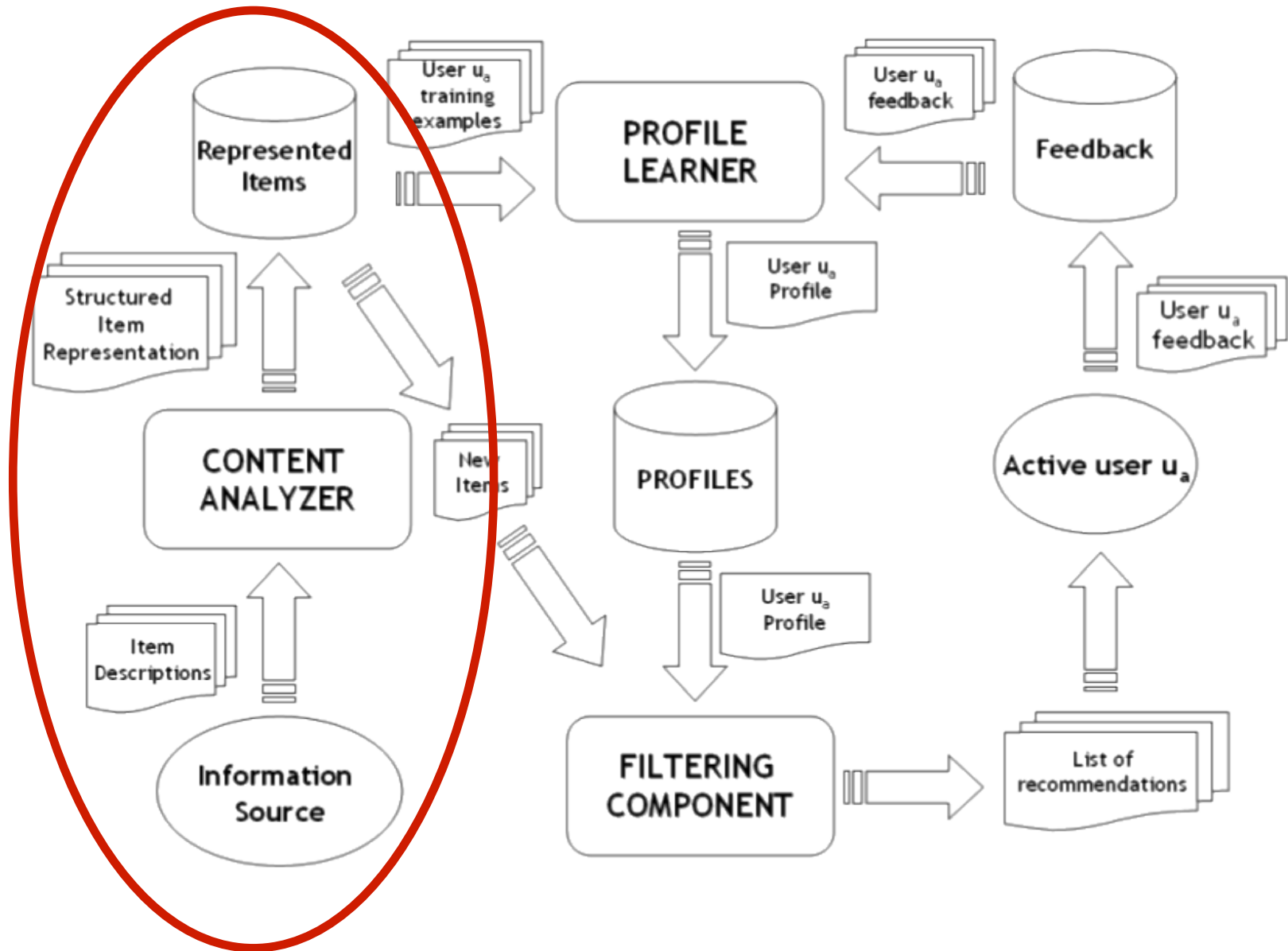
- Nearest neighbors
 - Given a set of items representing the user profile, select their most similar items which are not in the user profile
 - Predict the rate only for the N nearest neighbors

Content-Based Recommender Systems

- Nearest neighbors with kNN



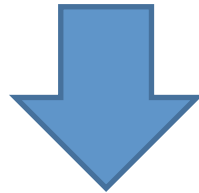
Content-Based Recommender Systems



Main CB RSs Drawback: Limited Content Analysis

No suggestion is available if the analyzed content does not contain enough information to discriminate items the user might like from items the user might not like.*

The quality of CB recommendations are correlated with the quality of the features that are explicitly associated with the items.



Need of domain knowledge!
We need rich descriptions of the items!

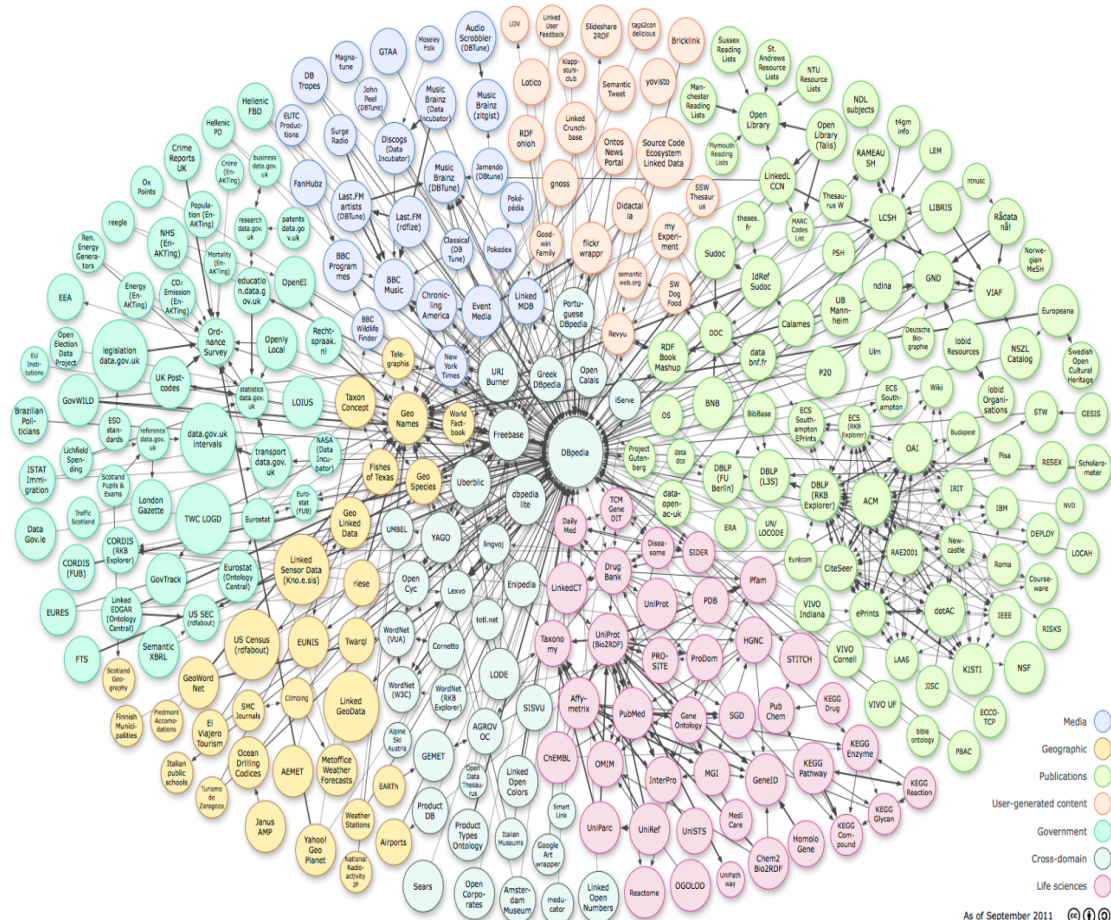
(*) P. Lops, M. de Gemmis, G. Semeraro. Content-based Recommender Systems: State of the Art and Trends. In: P. Kantor, F. Ricci, L. Rokach and B. Shapira, editors, Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners

A solution based on Linked Data




Use Linked Data to mitigate the limited content analysis issue

- Plenty of structured data available
- No Content Analyzer required



FRED: Transforming Natural Language text to RDF/OWL graphs

 **STLab - tools**

Fred Tipalo Aemoo Wikifier

Fred

Enter a text:

In early 1527, Cabeza De Vaca departed Spain as the treasurer of the Narvaez royal expedition to occupy the mainland of North America. After landing near Tampa Bay, Florida on April 15, 1528, Cabeza De Vaca and three other men would be the only survivors of the expedition party of 600 men.

input

Options: ☐ FrameNet roles ☐ NER ☐ Tipalo ☐ WSD

Output format: Graph

Parse

Examples

Miles Davis was an american jazz musician. (Fact)

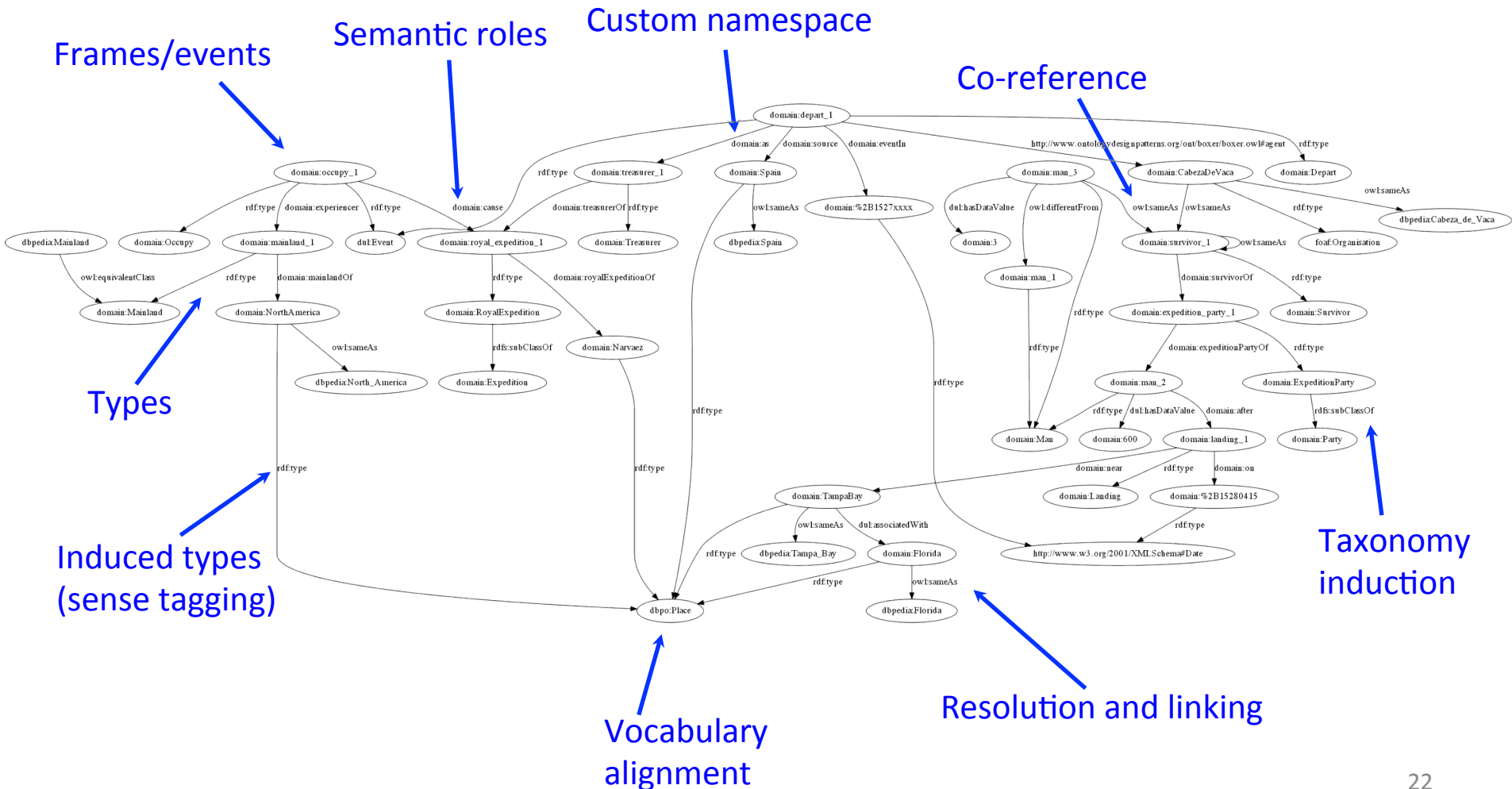
A wind instrument is a musical instrument that contains some type of resonator. (Definition)

Every brass instrument is a musical instrument whose sound is produced by sympathetic vibration of air in a tubular resonator in sympathy with the vibration of the player's lips. (Definition)

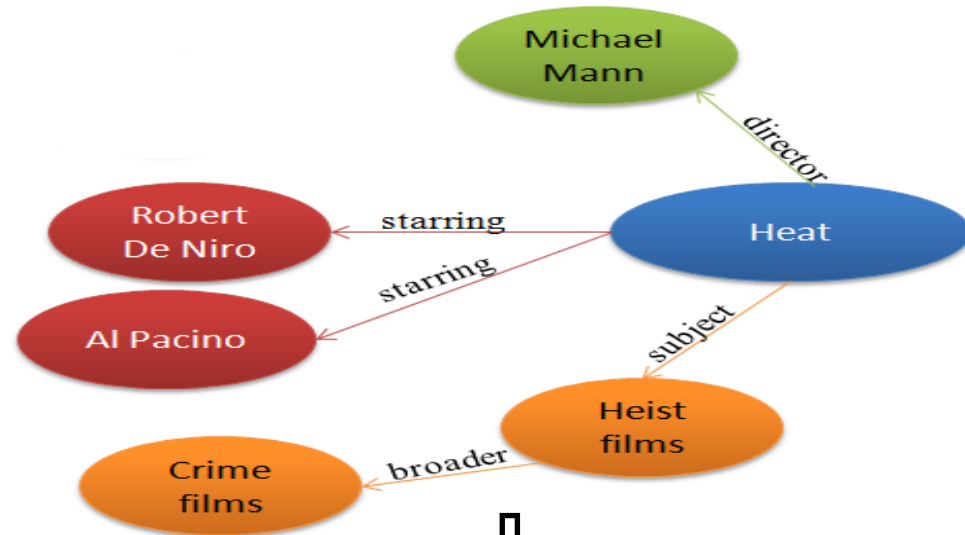
President Barack Obama and European Union leaders huddled in Washington amid growing fears over the future of the euro, which closed greater than 1.3 dollars. (Fact)

The New York Times reported that John McCarthy died. He invented the programming language LISP. (Fact)

FRED: RDF/OWL graph output



Linked Data as structured information source for items descriptions



Rich items descriptions

	<u>STARRING</u>		<u>DIRECTOR</u>	<u>SUBJECT+BROADER</u>	
Heat	Robert DeNiro	Al Pacino	Michael Mann	Heist films	Crime films

Select the domain(s) of your RS

```
SELECT count(?i) AS ?num ?c
WHERE {
    ?i a ?c .
    FILTER(regex(?c, "^http://dbpedia.org/ontology")) .
}
ORDER BY DESC(?num)
```

num	c
956476	http://dbpedia.org/ontology/Agent
763644	http://dbpedia.org/ontology/Person
572728	http://dbpedia.org/ontology/Place
387166	http://dbpedia.org/ontology/PopulatedPlace
348520	http://dbpedia.org/ontology/Settlement
333270	http://dbpedia.org/ontology/Work
277476	http://dbpedia.org/ontology/OrganisationMember
277476	http://dbpedia.org/ontology/SportsTeamMember

Tìpalo: automatic typing of DBpedia entities



STLab - tools

Fred Tipalo Aemoo Wikifier

Tìpalo

Enter a Wikipedia page URI:

e.g., http://en.wikipedia.org/wiki/Wind_instrument

Get types

Examples

[Wind instrument](http://en.wikipedia.org/wiki/Wind_instrument) (http://en.wikipedia.org/wiki/Wind_instrument)

[Pakito](http://en.wikipedia.org/wiki/Pakito) (<http://en.wikipedia.org/wiki/Pakito>)

[Neutron star](http://en.wikipedia.org/wiki/Neutron_star) (http://en.wikipedia.org/wiki/Neutron_star)

[Alter ego](http://en.wikipedia.org/wiki/Alter_ego) (http://en.wikipedia.org/wiki/Alter_ego)

[Lupercal](http://en.wikipedia.org/wiki/Lupercal) (<http://en.wikipedia.org/wiki/Lupercal>)

[Chaise longue](http://en.wikipedia.org/wiki/Chaise_longue) (http://en.wikipedia.org/wiki/Chaise_longue)

[French Revolution](http://en.wikipedia.org/wiki/French_Revolution) (http://en.wikipedia.org/wiki/French_Revolution)

Definition of Chaise longue:

A chaise longue is an upholstered sofa in the shape of a chair that is long enough to support the legs

See the [HTML version](#)



Input: natural language definition of an entity

http://en.wikipedia.org/wiki/Chaise_longue

A *chaise longue* (English /ˈʃeɪzˈlɔːŋ/; [ʃɛzlɔ̃ŋ(ə)] French pronunciation: [ʃɛzlɔ̃ŋ(ə)], "long chair") is an upholstered sofa in the shape of a chair that is long enough to support the legs.

A chaise longue is an upholstered sofa in the shape of a chair that is long enough to support the legs.

DBpedia resource

Extracted type

Inferred superclass

Linked resource

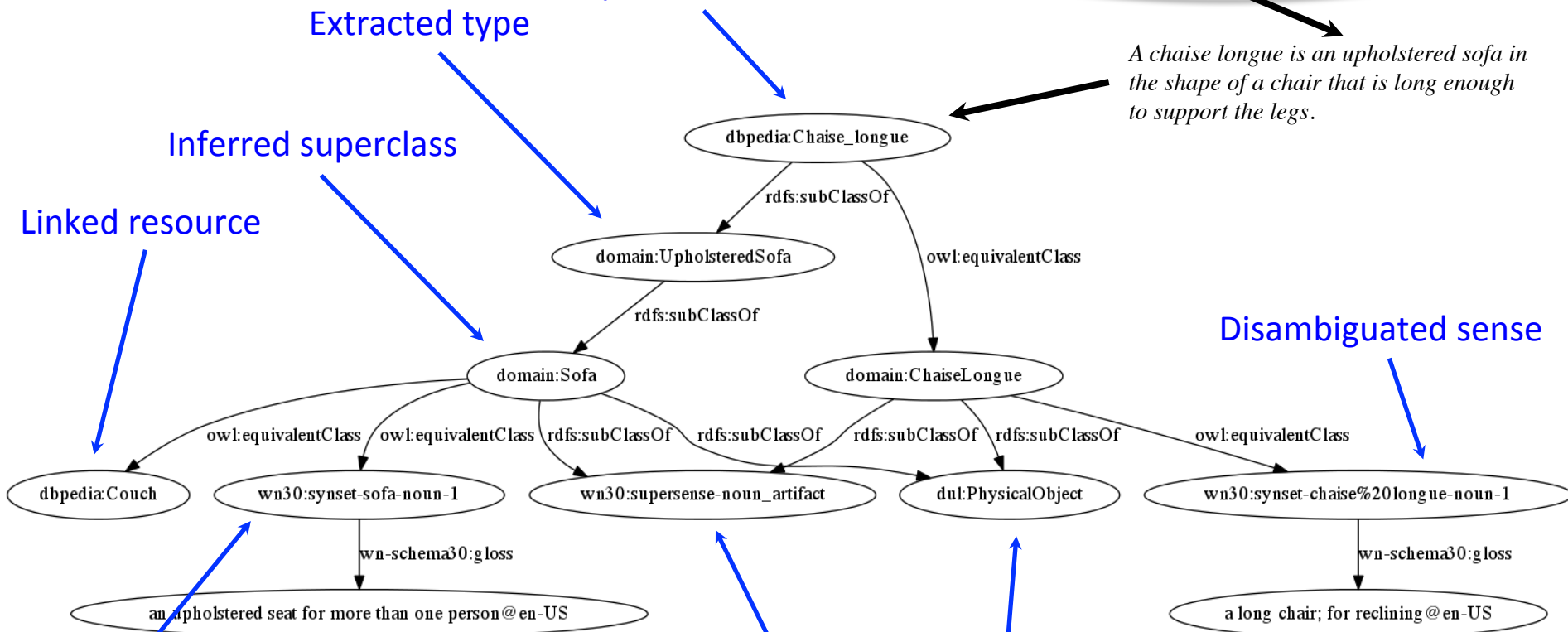
Disambiguated sense

Linking to WN supersenses

Linking to DOLCE

Disambiguated sense

Rich taxonomical data!



Select the sub-graph you are interested in

```
select ?m ?p ?o1
where{
  ?m ?p ?o1 .
  ?m dcterms:subject ?o.
  dbpedia:The_Matrix dcterms:subject ?o.
  ?m a dbpedia-owl:Film.
  filter(regex(?p,"^http://dbpedia.org/ontology/")) .
}
order by ?m
```

Enriching Annotations with Resources From the Web

Watson: find ontologies on the Semantic Web for enriching the representation of your data



What is it? · Submit URI · Website · Blog · APIs

Furniture Search Watson

Found 399 semantic documents - Search Options

- 1- <http://www.talkdieger.com/sioc/www.fullhouses.com/xml> ⓘ
 - ◆ <http://search.dmoz.org/cgi-bin/search?search=furniture>
 - ◆ <http://ma.gnolia.com/people/gilang/tags/furniture>
 - ◆ <http://furnitureindonesia.blogspot.com/2005/10/jepara-furniture-company.html>
 - ◆ <http://antiquefurniture.furniturelogic.com/index.php/2007/07/16/indonesia-furniture-wholesale-antique-reproductions-furniture>
 - ◆ <http://www.webnews.de/kommentare/61064/Indonesia-Furniture-Wholesale-Antique-Reproductions-Furniture-exporter.html>
 - ◆ <http://www.marktd.com/story.php?title=Indonesia-Furniture-Wholesale-Antique-Reproductions-Furniture-exporter-from-Indonesia>
 - ◆ <http://www.nexbig.com/story.php?title=Indonesia-Furniture-Wholesale-Antique-Reproductions-Furniture-exporter-from-Indonesia>
 - ◆ http://digg.com/business_finance/Indonesia_Furniture_Wholesale_Antique_Reproductions_Furniture
 - ◆ <http://www.anindo.com/directories.php>
 - ◆ <http://osx.iusethis.com/app/fullhousesfurniture>[More...](#)
- 2- <http://www.dcs.shef.ac.uk/~angus/daml-oil-workshop/ikea-example/ikea-bt.daml> ⓘ
 - ◆ file:C:\olled\ontologies\kitchen.daml#furniture
- 3- <http://lunarfury.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=clawing+the+furniture>
- 4- <http://www.talkdieger.com/sioc/www.dognameilists.com/blog/xml> ⓘ
 - ◆ <http://www.feedster.com/search/furniture+warehouse>
- 5- <http://jettastar.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=ikea+furniture>
- 6- <http://lucycat.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=scratching+furniture>
- 7- <http://fractal-thought.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=furniture>
- 8- <http://mzt.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=office+furniture>
- 9- <http://rdgadz.livejournal.com/data/foaf> ⓘ
 - ◆ <http://www.livejournal.com/interests.bml?int=furniture>
- 10- <http://reliant.tekknowledge.com/DAML/Economy.daml> ⓘ
 - ◆ <http://paoli.open.ac.uk/watson-cache/2/88f/5f7a/3c1d0/63ea3cc>
 - ◆ <http://reliant.tekknowledge.com/DAML/Mid-level-ontology.daml>



Search interface type: **Simple** Advanced Guru Query Language Documentation

keyword(s)

chaise longue

SEARCH

Group By Dataset: ☐ Sorted by: **relevance**

Quick filters (All options)

Time range:

Any date

Today Yesterday Last week

Last month Last year

Format:

Any format

RDF RDFa MICRODATA

MICROFORMAT XFN HCARD

HCAL ENDAR HLISTING

HRES

ADR

Predict

e.g. 1

Class:

e.g. 1

Ontology: ?

e.g. foaf

Domain: ?

e.g. abc.com OR abc

Sindice search: **chaise longue** found 4,782 documents (in 2.32 seconds)

"chaise longue", chaise longue (TITLE, MICRODATA, RDFa)

2012-07-18 – 22 triples in 3.4 kB

<http://www.solostocks.it/vendita-prodotti/salute-medicina-bellezza/altri-salute-medicina-bellezza/ch...> (Search) Inspect: (Cache) (Live)

Chaise longue (RDFa)

2010-07-15 – 35 triples in 6.8 kB

<http://it.bestshopping.com/prezzi/Chaise-longue.sku=9788881127702%7C.html> (Search) Inspect: (Cache) (Live)

chaise longue (RDF)

2010-02-04 – 9 triples in 1.5 kB

Sindice: find other entities for enriching your knowledge base

Chaise Longue 030CPB Viciani (RDFa)

2010-07-12 – 35 triples in 7.4 kB

<http://it.bestshopping.com/prezzi/Viciani-Chaise-Longue-030CPB.sku=030CPB%7C.html> (Search) Inspect: (Cache) (Live)

Chaise Longue 030CPS Viciani (RDFa)

triples in 7.4 kB

<http://it.bestshopping.com/prezzi/Viciani-Chaise-Longue-030CPS.sku=030CPS%7C.html> (Search) Inspect: (Cache) (Live)

096CPA Viciani (RDFa)

triples in 7.4 kB

<http://it.bestshopping.com/prezzi/Viciani-Chaise-Longue-096CPA.sku=096CPA%7C.html> (Search) Inspect: (Cache) (Live)


096CPB Viciani (RDFa)

<sameAs>

interlinking the Web of Data

The Web of Data has many equivalent URIs.
This service helps you to find co-references
between different data sets.
Enter a known URI, or use Sindice to search first.

<sameAs> 🔍

 🔍

Equivalent URIs for http://dbpedia.org/resource/Chaise_Longue –

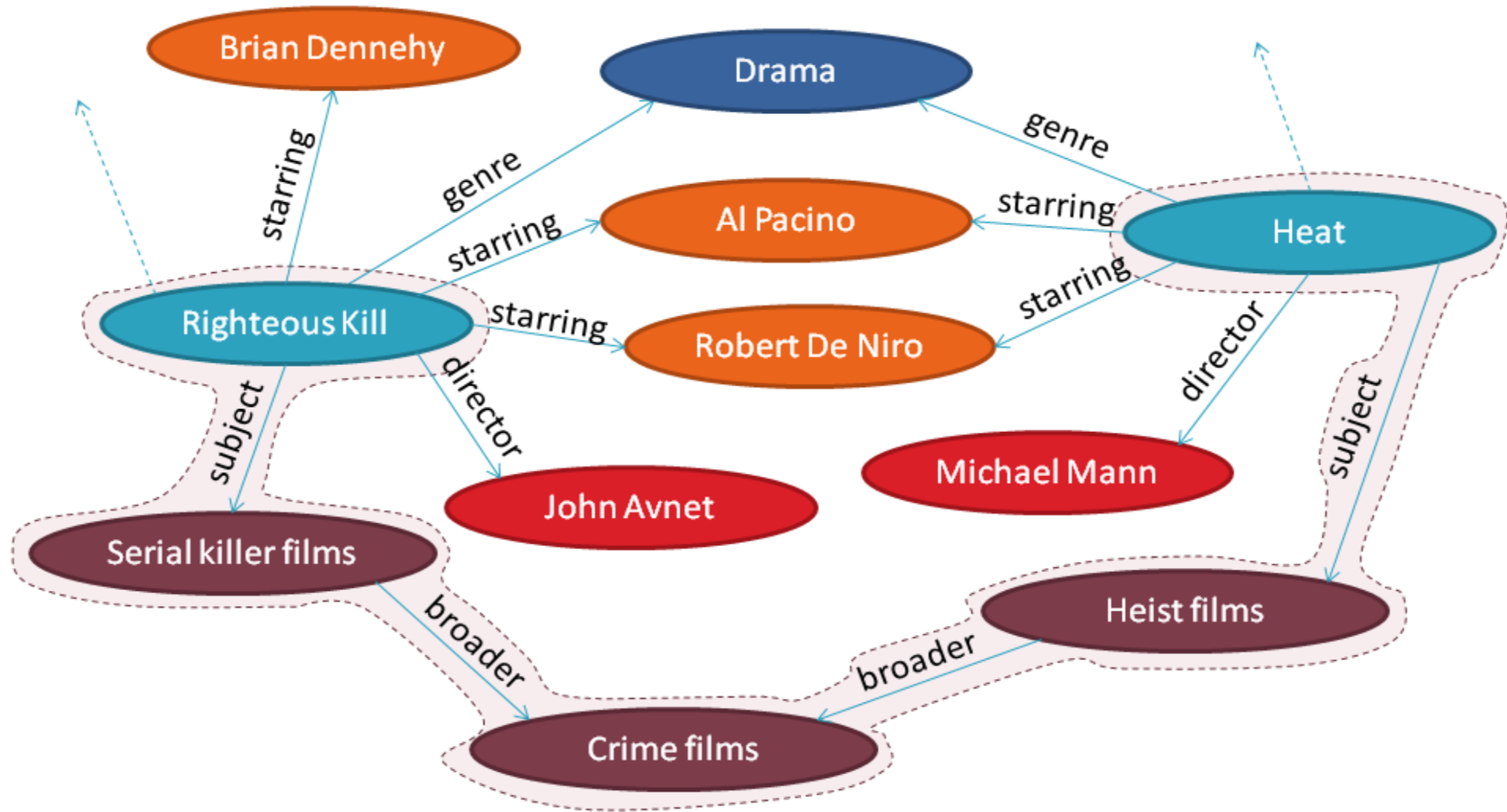
- 1. <http://dbpedia.org/resource/Chaiselongue>
- 2. <http://dbpedia.org/resource/Chaiselongue>
- 3. http://dbpedia.org/resource/Chaise_lounge

Show 20 more

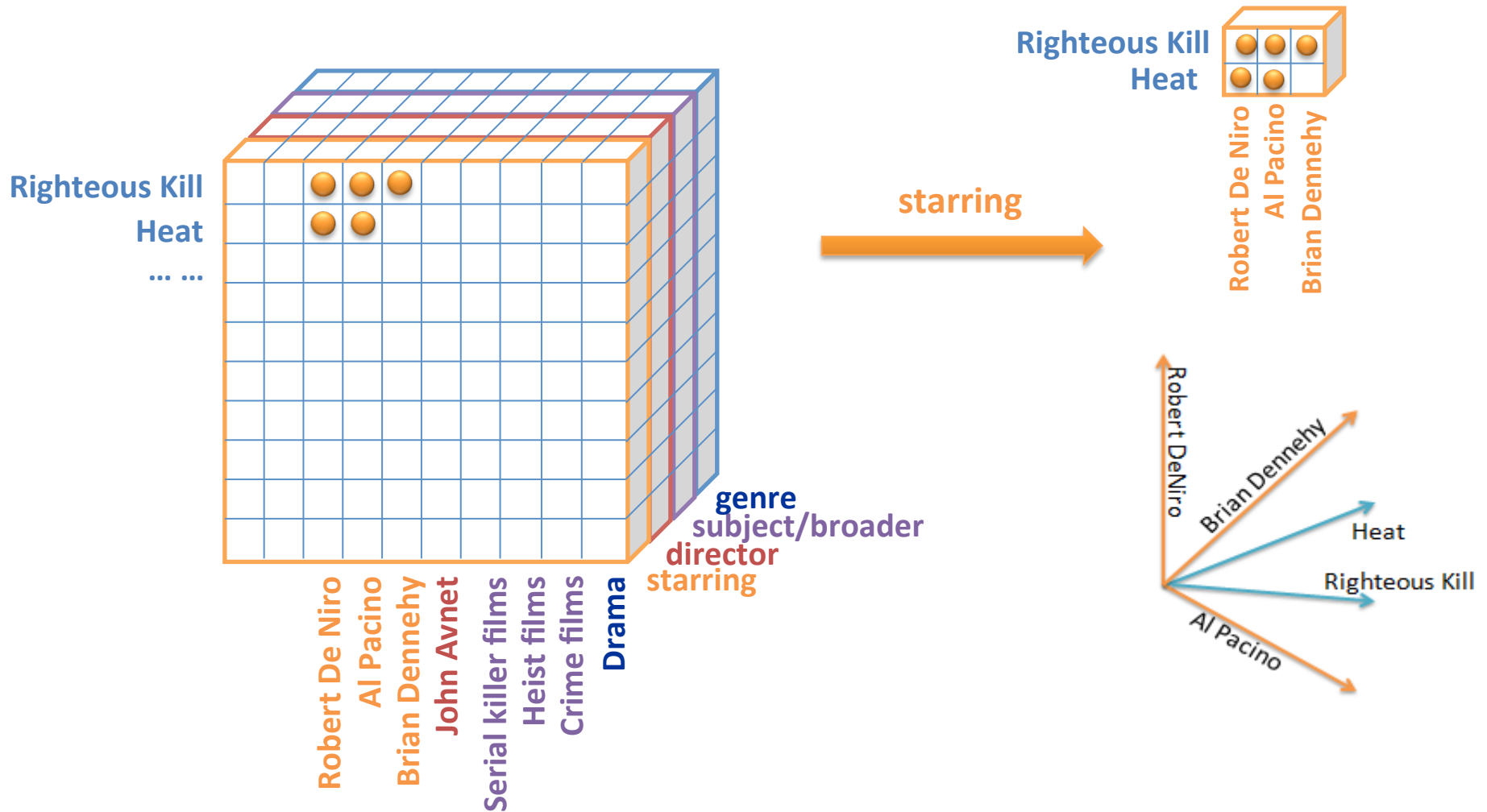
24. <http://sw.opencyc.org/2009/04/07/concept/Mx4rvVjraJwpEbGdrcN5Y29ycA>
rdf+xml · n3 · json · text

SameAs.org: find other entities to link to from your knowledge base

Computing similarity in LOD datasets



Vector Space Model for LOD



Vector Space Model for LOD



Righteous Kill

<u>STARRING</u>	Al Pacino (v1)	Robert De Niro (v2)	Brian Dennehy (v3)
Righteous Kill (m1)	X	X	X
Heat (m2)	X	X	



Heat

$$w_{\downarrow AlPacino, Heat} = tf_{\downarrow AlPacino, Heat} * idf_{\downarrow AlPacino}$$

Righteous Kill (x1)	$w_{v1,x1}$	$w_{v2,x1}$	$w_{v3,x1}$
Heat (x2)	$w_{v1,x2}$	$w_{v2,x2}$	0

Vector Space Model for LOD



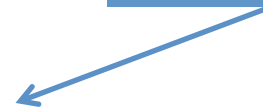
<u>STARRING</u>	Al Pacino (a1)	Robert De Niro (a2)	Brian Dennehy (a3)
Righteous Kill (x1)	0	0	0
Heat (x2)	0	0	

Righteous Kill



Heat

$$tf \in \{0,1\}$$



$$w_{\downarrow AlPacino, Heat} = tf_{\downarrow AlPacino, Heat} * idf_{\downarrow AlPacino}$$

Righteous Kill (m1)	$w_{a1,m1}$	$w_{a2,m1}$	$w_{a3,m1}$
Heat (m2)	$w_{a1,m2}$	$w_{a2,m2}$	0

Vector Space Model for LOD

$$\text{sim} \downarrow \text{starring} (x \downarrow i, x \downarrow j) = w \downarrow v \downarrow 1, x \downarrow i * w \downarrow v \downarrow 1, x \downarrow j + w \downarrow v \downarrow 2, x \downarrow i * w \downarrow v \downarrow 2, x \downarrow j + w \downarrow v \downarrow 3, x \downarrow i * w \downarrow v \downarrow 3, x \downarrow j + \dots$$

$$\alpha \downarrow \text{starring} * \text{sim} \downarrow \text{starring} (x \downarrow i, x \downarrow j)$$

$$\alpha \downarrow \text{director} * \text{sim} \downarrow \text{director} (x \downarrow i, x \downarrow j)$$

$$\alpha \downarrow \text{subject} * \text{sim} \downarrow \text{subject} (x \downarrow i, x \downarrow j)$$

...

=

$$\text{sim} \downarrow \text{starring} (x \downarrow i, x \downarrow j)$$

LOD-based CB RS

Given a user profile defined as: $X_{u,i} = \{ \langle x_i, r_{x_i, u_i} \rangle \}$

We can predict the rating using a **Nearest Neighbor Classifier** wherein the similarity measure is a linear combination of **local property similarities**

$$r(u_i, x') = \sum_{\langle x_i, r_{x_i, u_i} \rangle \in X_{u_i}} \sum_{p \in P} \alpha_p * \text{sim}_p(x_i, x') / P * r_{x_i, u_i}$$

=

Missing in this presentation (not a complete list)

- Learning $\alpha \downarrow p$
- Explanation
 - Quite straight with a content-based approach
- Hybrid approaches
 - Mix a content based approach with a collaborative one
 - Collaborative similarity as a further vector space (property) of the content based approach?
 - Exploit the graph-based nature of both the rating matrix and the RDF graph
- ...

Thank you!

Many thanks to **Vito Claudio Ostuni** and **Roberto Mirizzi**