

Accessing Cultural Heritage using Semantic Web Techniques

Hu Jingwen

overview

- Cultural Heritage interoperability problems
- Why Semantic Web techniques can be relevant
- Porting CH vocabularies to the Semantic Web
- Vocabulary alignment

The Interoperability Problem in Cultural Heritage

- *Trend*: simultaneous access to different collections
The European Library, Memory of the Netherlands
- Problem: how to access seamlessly different collections?
- Traditional solution: using object metadata
For instance subjects coming from controlled vocabularies
But...

Interoperability Problems

- Different formats
- Different metadata schemas
- Different conceptual vocabularies

Interoperability Solutions?

- Different formats

“We have a solution!”

XML as a standard for data exchange

- Different metadata schemas

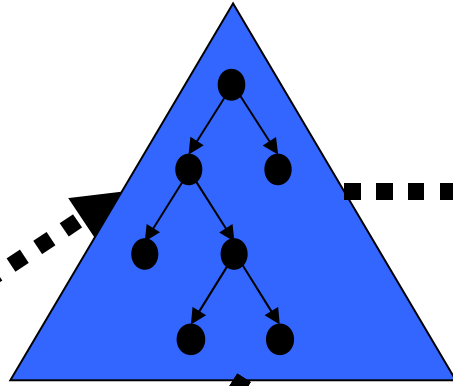
“Something could be used...”

Dublin Core for simple metadata publication & exchange

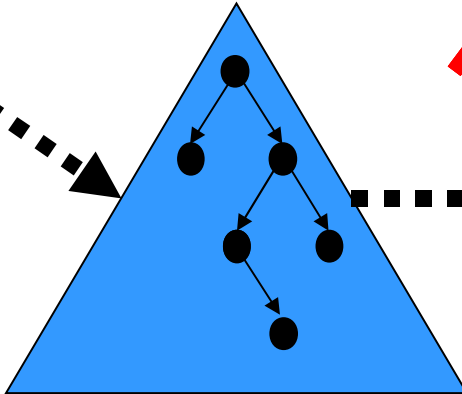
- Different conceptual vocabularies

in concepts identical or similar meanings but different labels or names

What we have



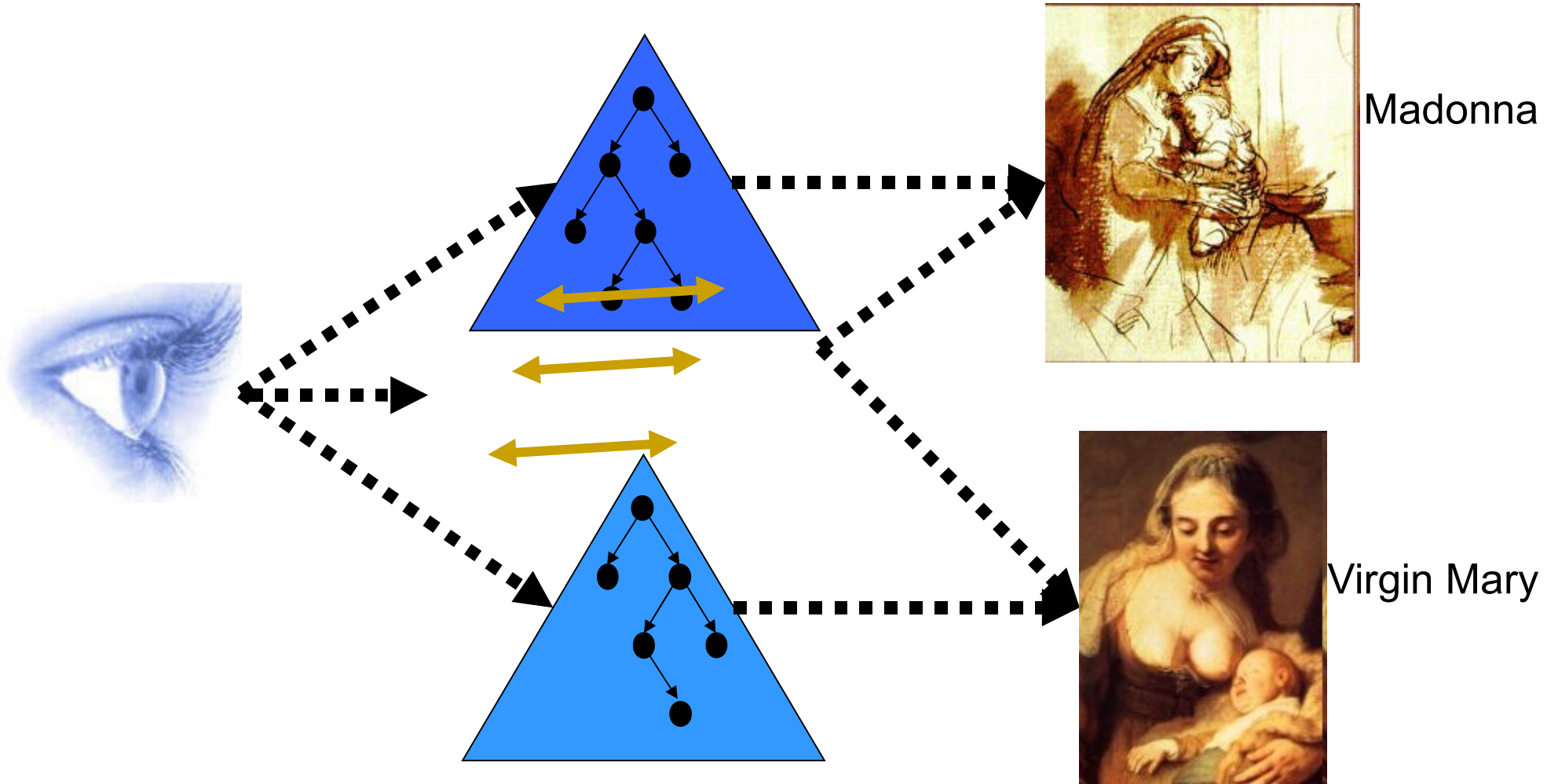
Madonna



Virgin Mary



What we want



With semantic web technique: Any 2 have identical or similar meanings vocabularies will align.

overview

- Cultural Heritage interoperability problems
- [Why Semantic Web techniques can be relevant](#)
- Porting CH vocabularies to the Semantic Web
- Vocabulary alignment

What is the Semantic Web?

- “The Semantic Web is a web of data”
- an evolving development of the World Wide Web
- making it possible for the web to "understand" and satisfy the requests of people and machines to use the web content

SW Problem: The Web for Humans, But the Web for Computers?

Amsterdam

From Wikipedia, the free encyclopedia

For other uses, see Amsterdam (disambiguation).

Amsterdam (help·info) is the capital of the Netherlands, and is located in the south of the North Holland province. The city is known for its historic port, the Rijksmuseum, the red-light district (*de Wallen*), the liberal coffeeshops, and the canals which have led to Amsterdam being termed the "Venice of the North".^{[5][6][7][8]} During the Dutch Golden Age, Amsterdam was one of the most important ports in the world, with innovative developments in trade, and became the leading centre for finance and diamonds.^{[9][10]}

The city was founded in the late 12th century as a small fishing village, and has grown to become the largest city in the Netherlands with a population of 743,027 inhabitants, which includes 177 different nationalities, making Amsterdam the most multicultural city in the world.^[11]

The metropolitan region has a population of 1,021,870 inhabitants and is part of the conglomerate metropolitan area Randstad, with a population of 6,659,300 inhabitants. The name *Amsterdam* is a derivative from *Amstel dam*,^[12] that is, a dam in the river Amstel.

Contents [hide]

- History
- Geography
 - The canals
 - Climate
- Economy
 - Retail
- Demography
- Culture

Amsterdam



Canal houses alongside the Prinsengracht



Flag



Coat of arms

Nickname: Mokum

Motto: Heldhaftig, Vastberaden, Barmhartig
(*Valiant, Determined, Compassionate*)



- A city

- A flag

- The city's location

The Semantic Web (1/2)

- **Pointing at resources**

- *What?* **Knowledge objects**

- everything that we may want to refer to including documents, persons...

- **How? Uniform Resource Identifiers**

- HTTP URLs: <http://www.few.vu.nl/~aisaac/>
<mailto:aisaac@few.vu.nl>

The Semantic Web (2/2)

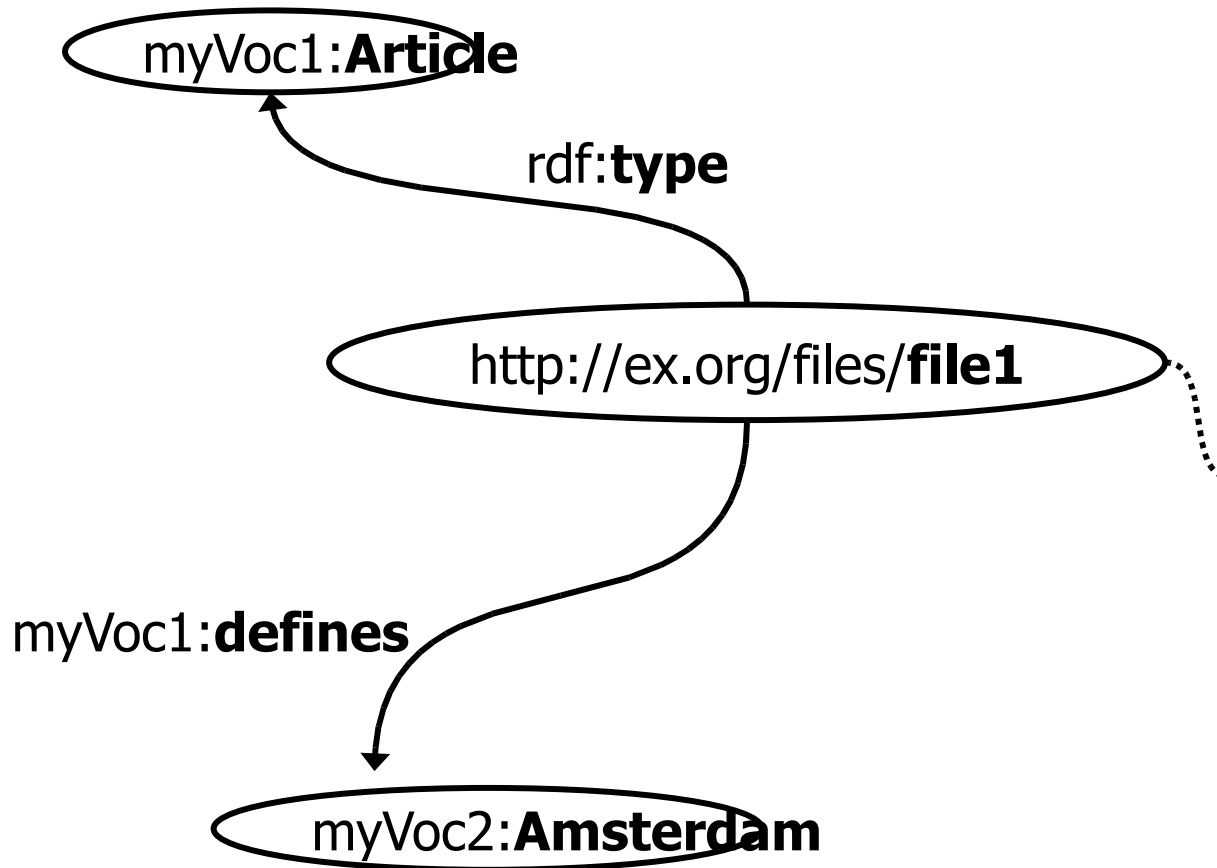
- Pointing at resources: URIs
- Creating structured assertions involving resources

*What? **Typed links** between resources*

*How? **RDF (Resource Description Framework)***

This means that Semantic Web data can merge and operate with resources coming from different information spaces.

Data in an RDF “Graph”



阿姆斯特丹，荷兰首都，人口约74万。

目录 [隐藏]

- 1 历史
 - 1.1 城市建立和初期发展
 - 1.2 黄金时代
 - 1.3 工业革命
 - 1.4 两次世界大战
 - 1.5 战后重建
 - 1.6 现代的城市
- 2 交通
- 3 文化
 - 3.1 體育活動
- 4 旅游景点
- 5 友好城市
- 6 外部链接

历史

阿姆斯特丹的历史最早可以追溯到13世纪时的渔村。人们曾在附近阿姆斯特河（Amstel）上建筑水坝，阿姆斯特丹就得名于此。原来的名字“Amstelredam”，意指“阿姆斯特河水坝”。17世纪是阿姆斯特丹历史上的“黄金时代”。阿姆斯特丹是当时世界上最重要的港口和银行业中心。

城市建立和初期发展

在上个千年之初，一些冒险者垂着由控空的画

阿姆斯特丹 Amsterdam	
	
别名: <i>Amstelredam</i>	
國家	荷蘭
省份	北荷兰省
现任市長	Job Cohen
面積	219平方公里
—城市面積	166平方公里
—土地面積	53平方公里
—水域面積	(2006年1月1日)
人口	743,027 (2006年1月1日)
—人口密度	4,459/平方公里
時區	UTC +1
位置	
—經度	4° 54′ E
—緯度	52° 22′ N
	

<http://www.ned.nl/rep321>

Each statement has a subject resource which is linked to an object resource via a property resource.

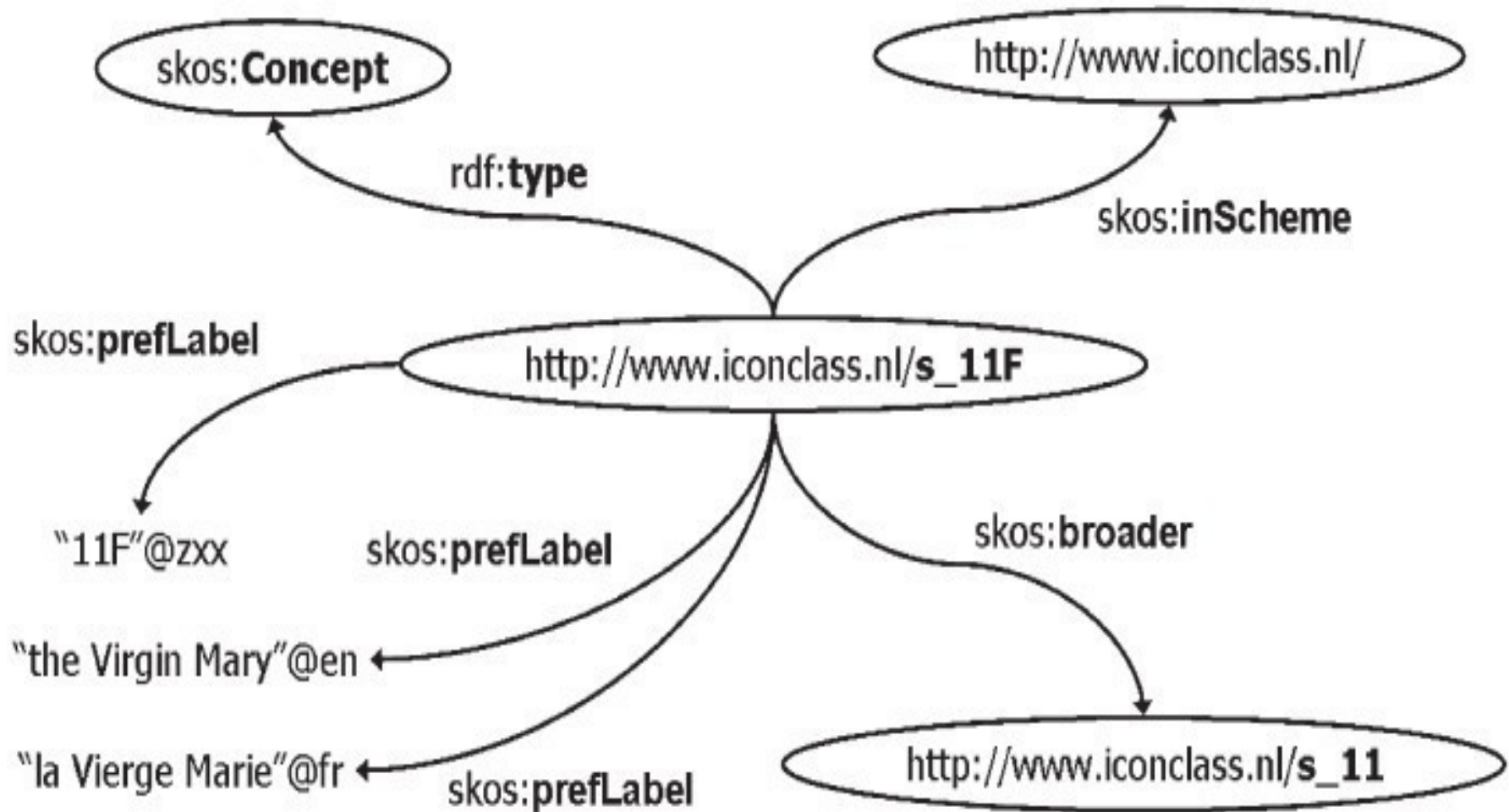
overview

- Cultural Heritage interoperability problems
- Why Semantic Web techniques can be relevant
- **Porting CH vocabularies to the Semantic Web**
- Vocabulary alignment

SKOS (Simple Knowledge Organization System)

- SKOS offers building blocks to represent KOSs in RDF
 - Objects: **Concept** and **ConceptScheme**
 - Lexical properties (multilingual)
 - prefLabel**
 - altLabel**
 - Semantic relations
 - broader, narrower**
 - related**
 - Notes
 - scopeNote**
 - definition**
 - ...

SKOS: Example



overview

- Cultural Heritage interoperability problems
- Why Semantic Web techniques can be relevant
- Porting CH vocabularies to the Semantic Web
- Vocabulary alignment

The semantic interoperability problem

- There is no standard vocabulary
- We don't really want it
different vocabularies for different expertise domains, traditions, tasks

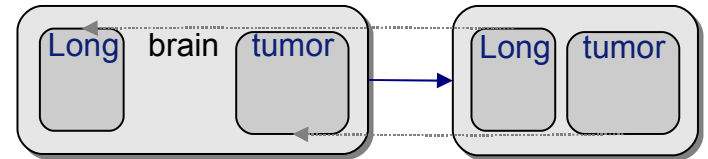
Vocabulary alignment

(a solution for interoperability problem)

- Aim: finding semantic correspondences between vocabulary elements
- Doing it (semi-) automatically
 - Vocabularies are big (tens of thousands concepts)
 - They change

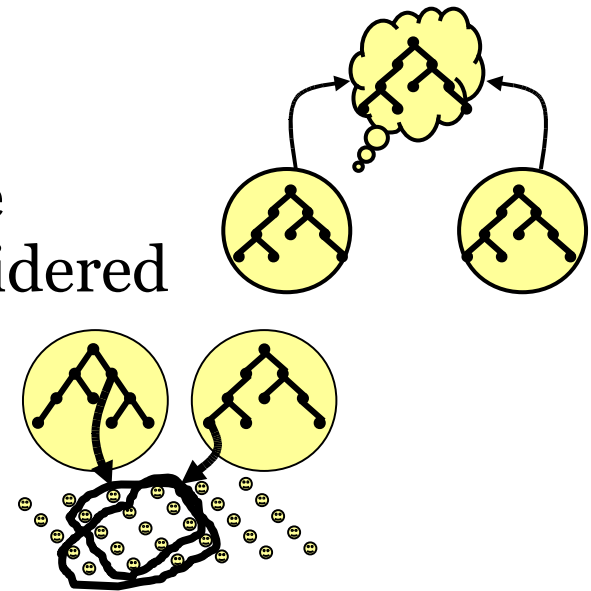
Automatic alignment techniques

- Lexical
lexical materialisations of the concepts are compared to each other



- Structural

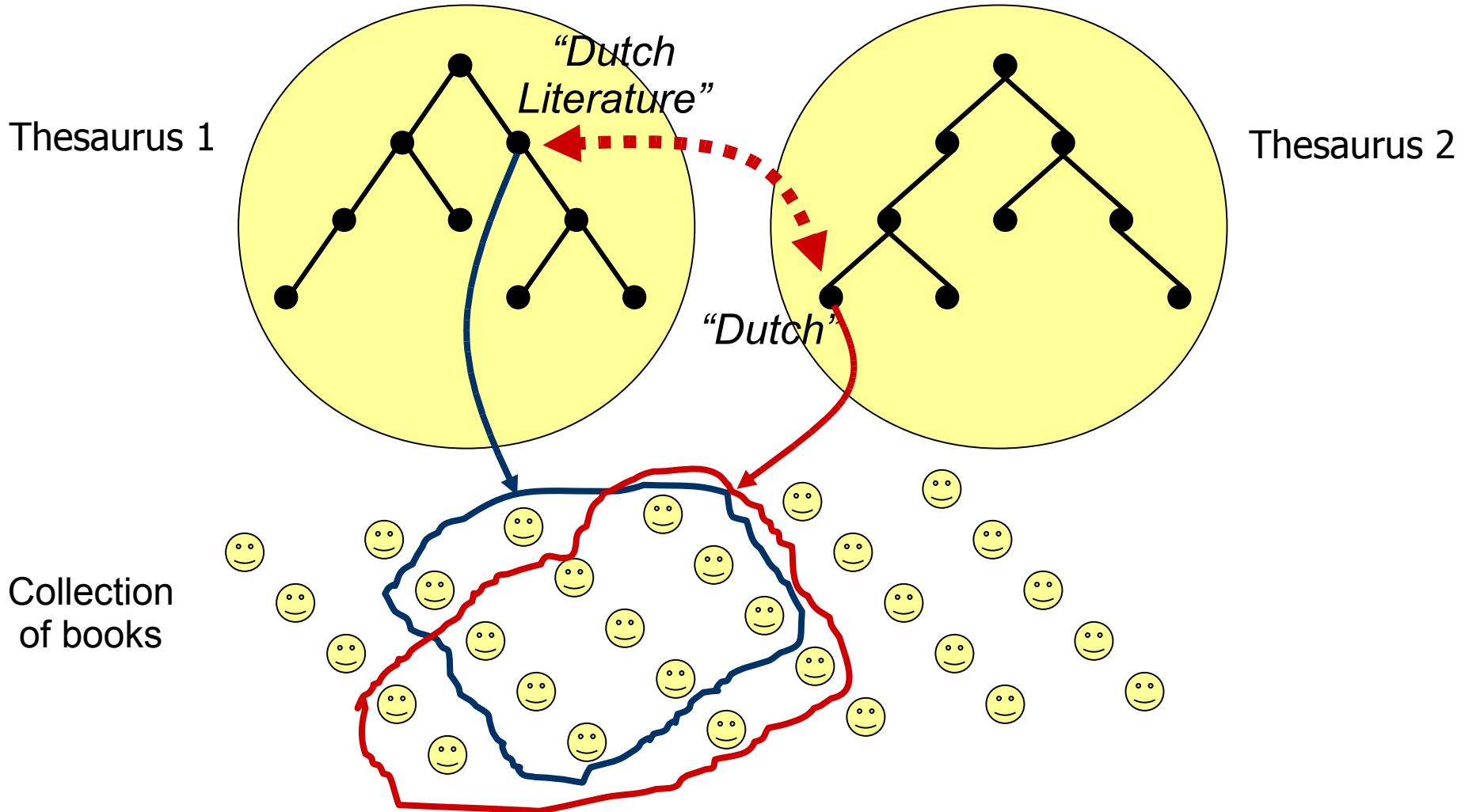
- Background knowledge
rely on knowledge sources: external to the application and the vocabularies being considered



- Extensional
Object information

Extensional Statistical Alignment

- **Object information** A statistical techniques can be used to compare the sets of documents describe by vocabularies. A high degree of overlap between this sets will yield a high similarity between coresponding concepts.



Alignment: still have some disadvantages

- Current techniques are not reliable as unique source of knowledge

Workflow would imply checking/completion by human

Combination of techniques is required

- Alignment is a *difficult* research problem

Summary

Semantic Web techniques

- **Representation** of collections and vocabularies
- **Alignment** of vocabularies

can help solving Cultural Heritage problems

- **Semantic integration**
- **Publication and access**

Thanks!