

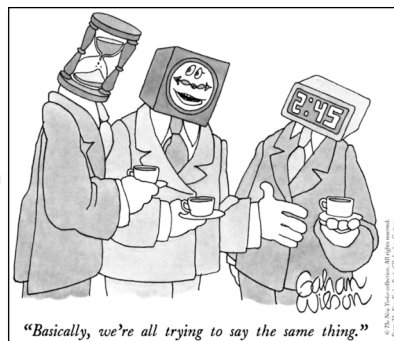
# Semantic Interoperability

With thanks to Jérôme Euzenat, INRIA & LIG, France and  
Natasha Noy, Stanford University, USA

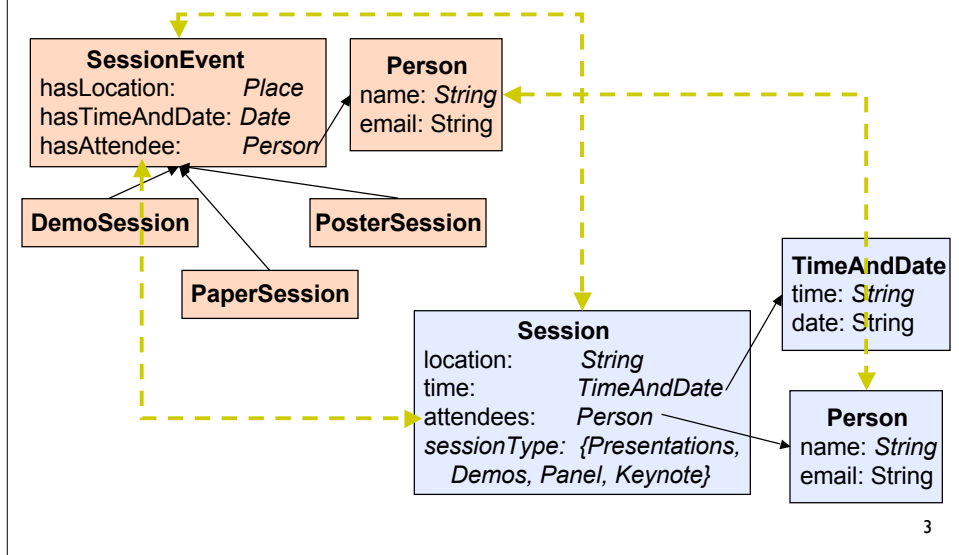


## Context

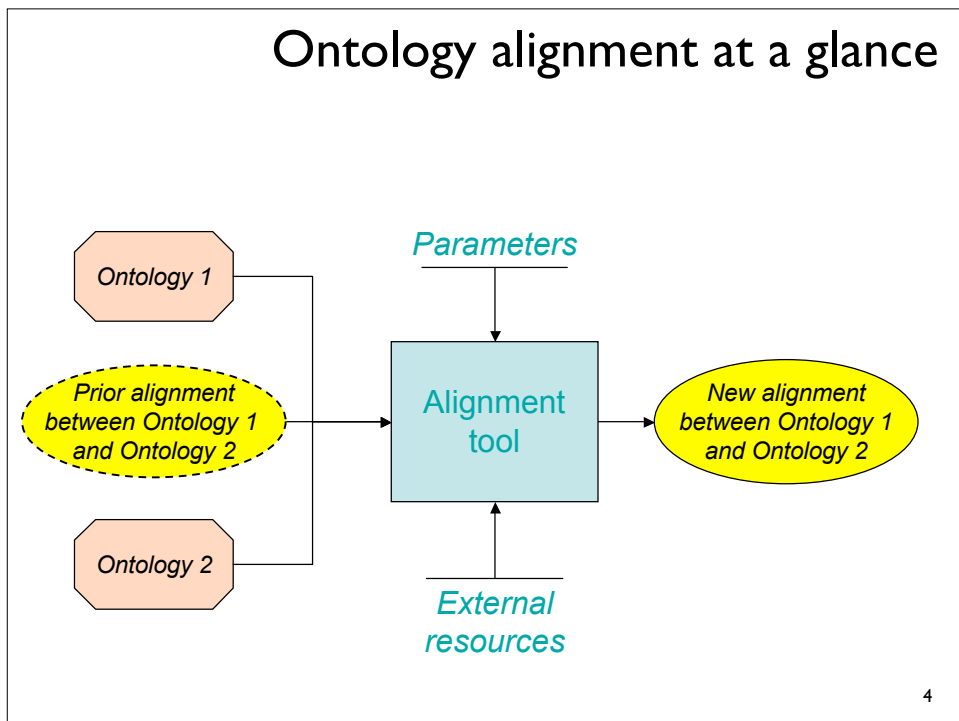
- So far, we've concentrated largely on a single ontology or conceptualisation
- In real applications, particularly in the Semantic Web, this is an unrealistic assumption
  - We will not just be dealing with one person's ontology
  - We will not just be dealing with a common ontology used by several people
  - We will have many different ontologies used by many different people
  - A mess, but a meaningful mess
  - How do we support interoperation between these models?



# Heterogeneous Ontologies: Example



# Ontology alignment at a glance

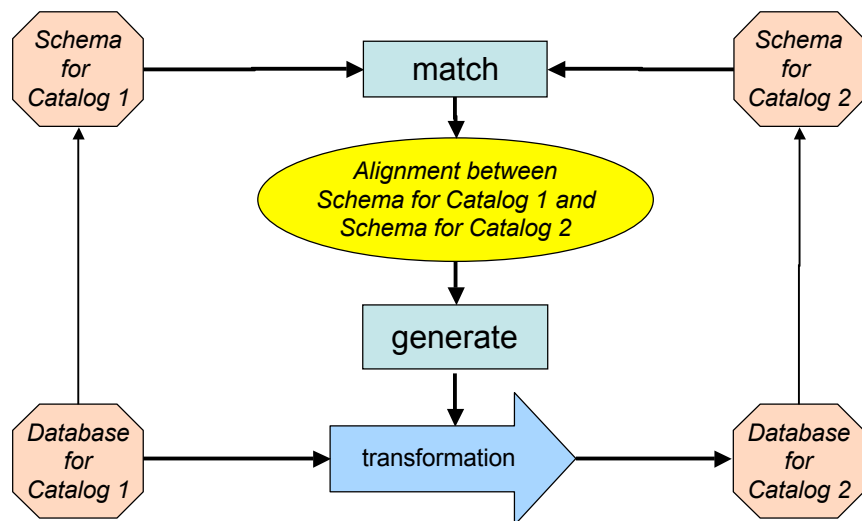


## Why should we learn to deal with this?

- Catalogue integration
- Schema and data integration
- Query answering
- Peer-to-peer information sharing
- Web service composition
- Agent communication
- Data transformation
- Ontology evolution

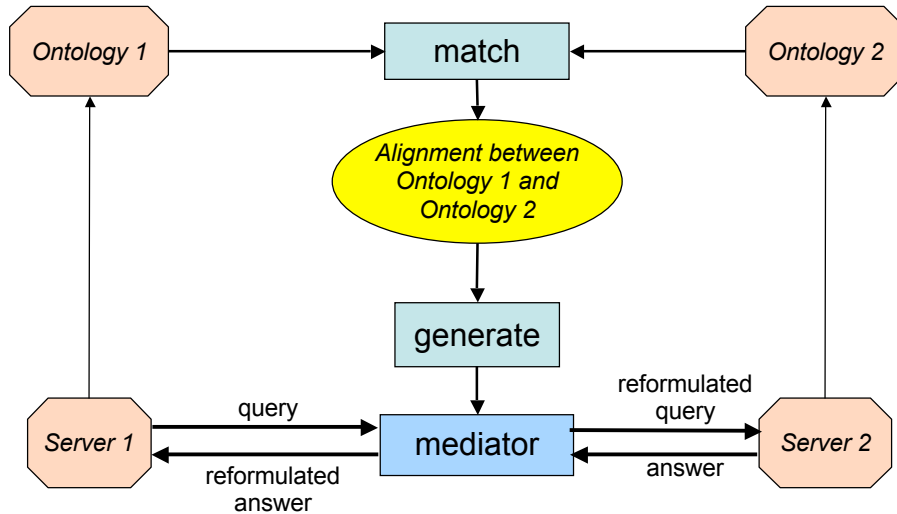
5

## Application: Catalogue integration

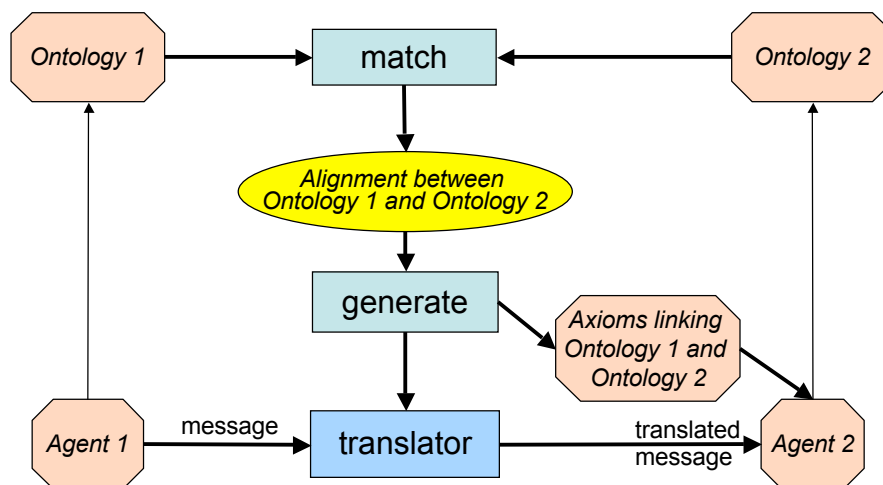


6

## Application: Query answering



## Application: agent communication

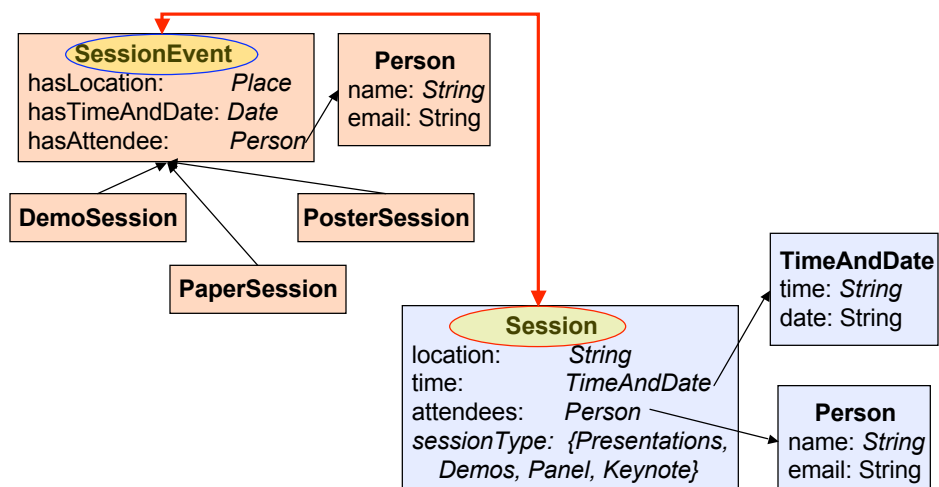


## Why is semantic interoperability difficult?

- There are a number of reasons why mismatches may occur
- Different contexts (databases, ontologies)
- Different representation languages/logics
- Same concept, different names
- Same name, different concepts
- Different approaches to conceptualization (e.g., subclasses versus property values)
- Different levels of granularity
- Different, but overlapping, domains

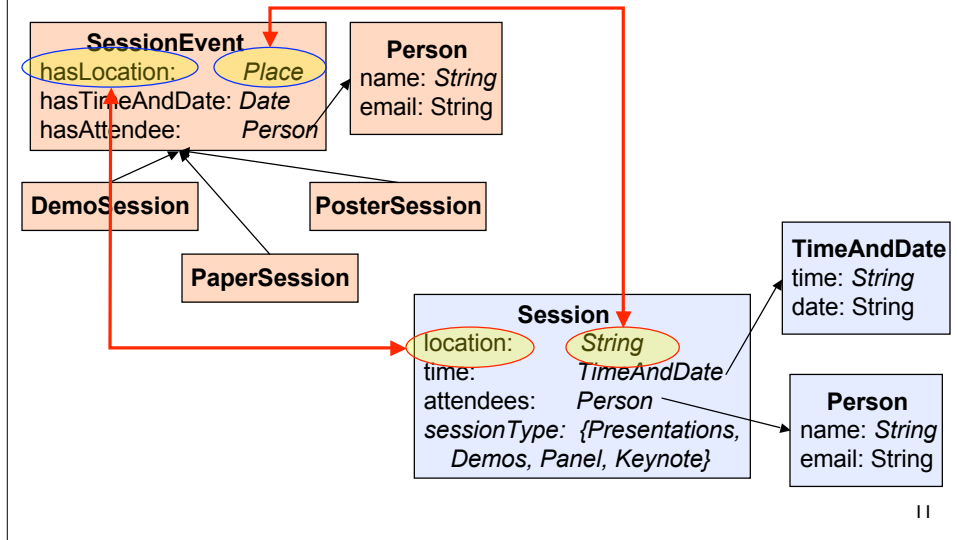
9

## Why is semantic interoperability difficult?

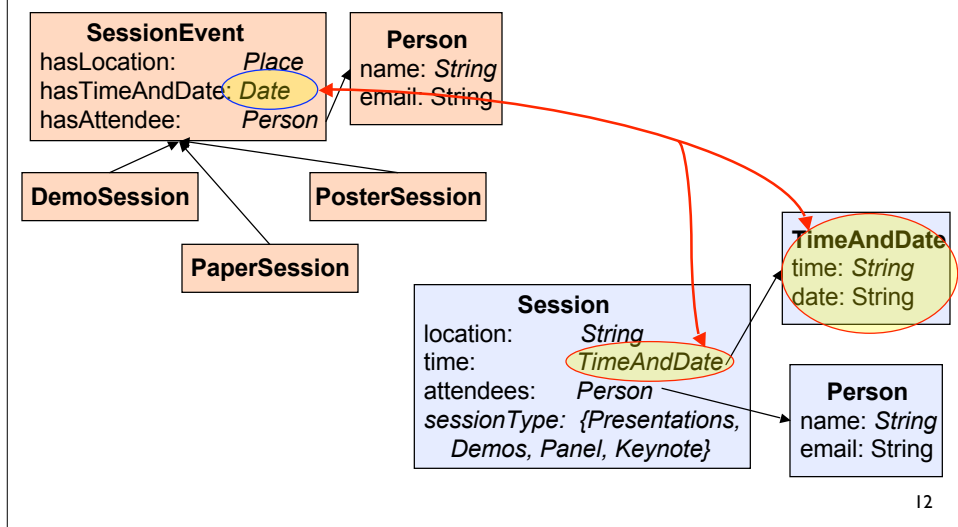


10

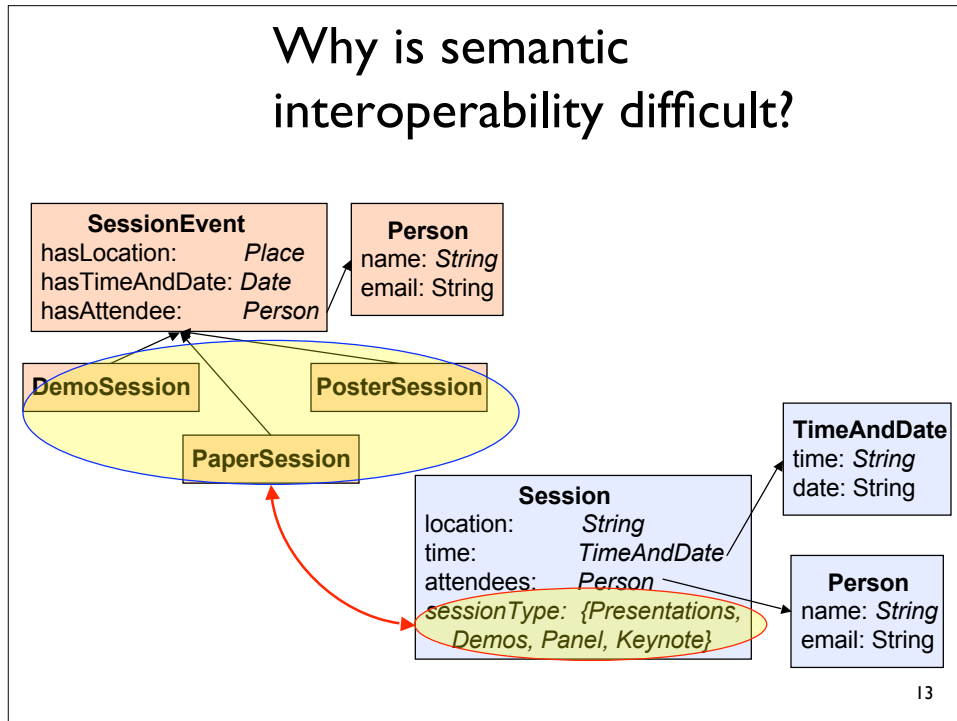
# Why is semantic interoperability difficult?



# Why is semantic interoperability difficult?



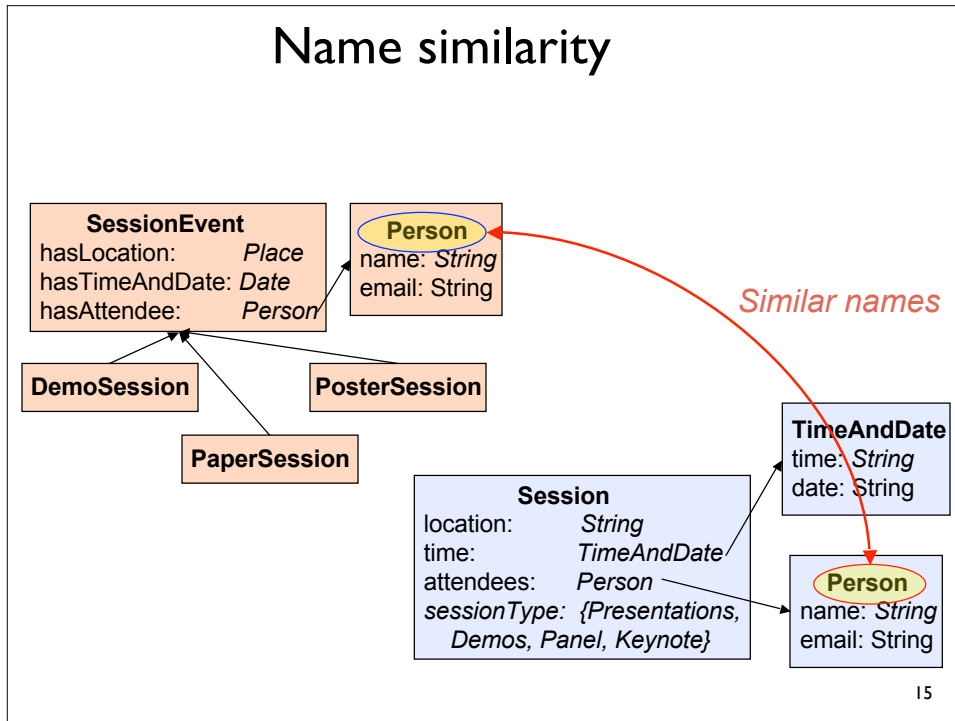
## Why is semantic interoperability difficult?



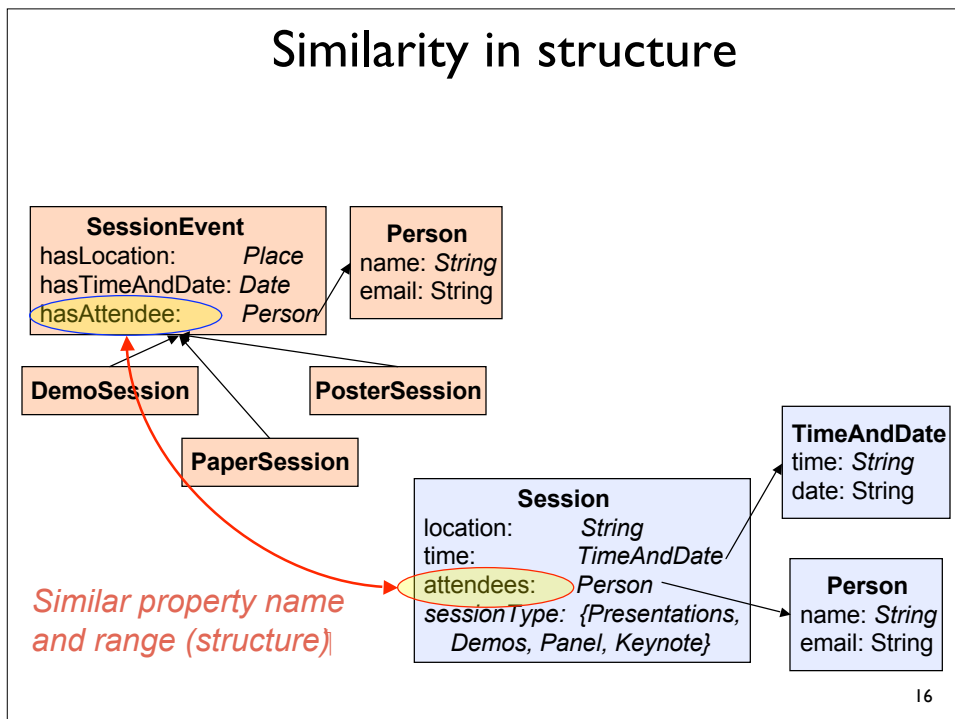
## How can we address the problem?

- Names of entities
  - Comments, alternate names, names of related entities
- Structure
  - Internal structure: constraints on relations, types
  - External structure: relations between entities
- Extensions
  - Instances themselves
  - Related resources: annotated documents, exchanged message or queries
- Semantics (models)
- Background knowledge
  - The Web
  - Ontologies
  - Thesauri, e.g. WordNet

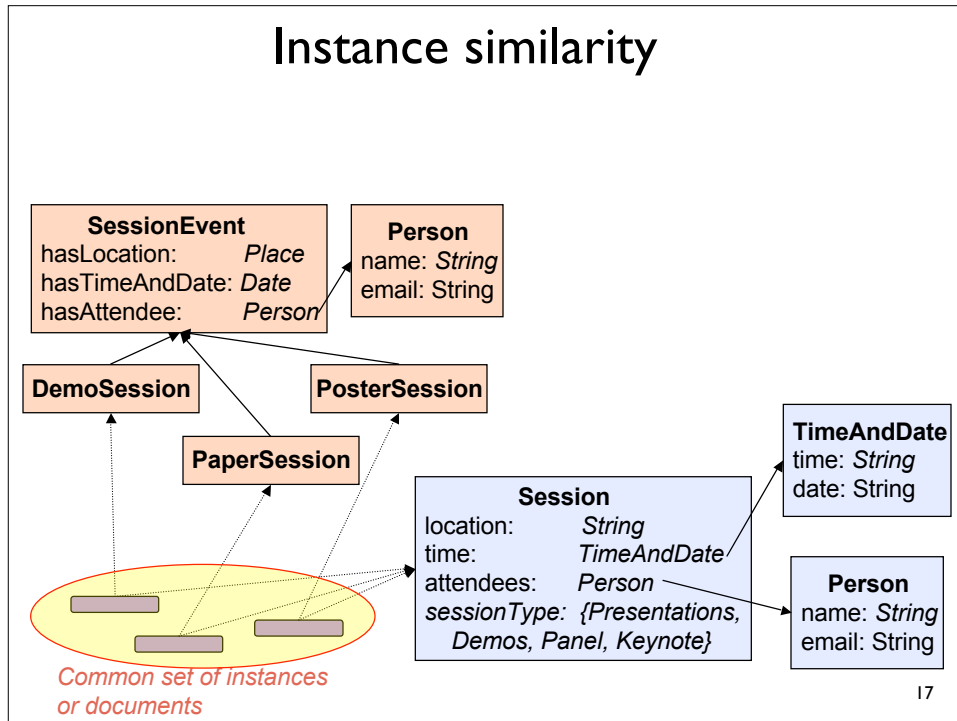
## Name similarity



## Similarity in structure



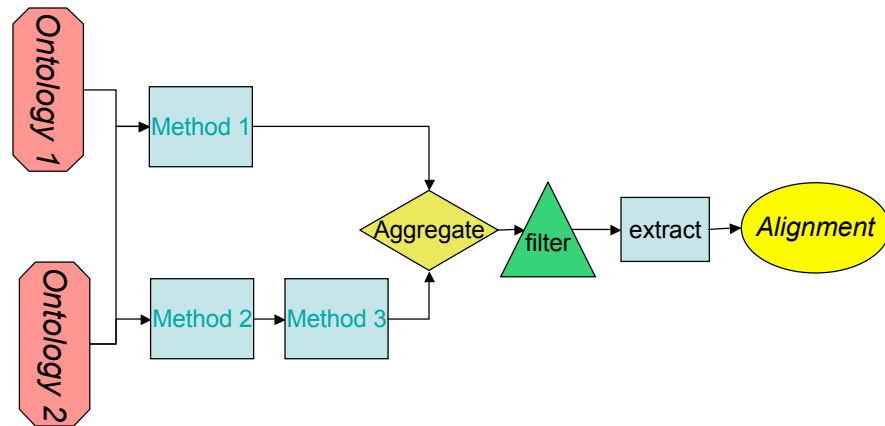
## Instance similarity



## External sources

- A common reference ontology
- User input
- Lexicons, thesauri, etc.
- Prior matches
- Background knowledge (other ontologies, documents, etc.)

## Combining different techniques



19

## Combining different techniques

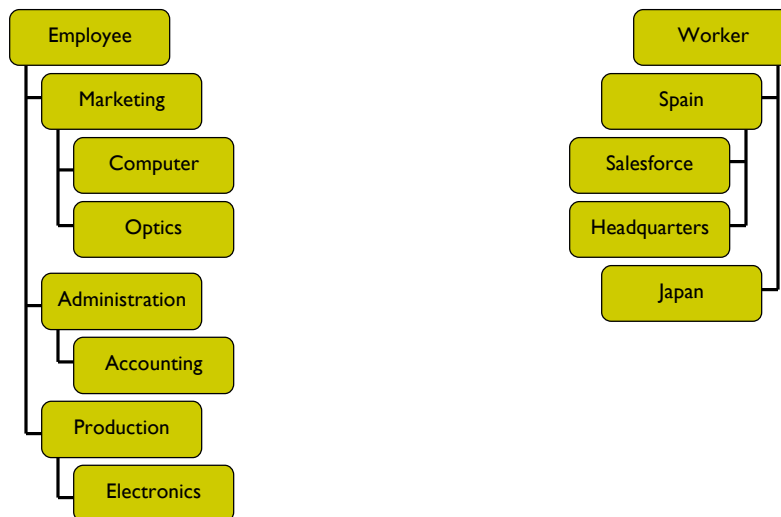
- Using several matchers in sequence (composing)
- Using several matchers in parallel (combining)
- Aggregating matcher results
  - aggregating specialised matcher results
  - aggregating competing matcher results
- Filtering results (trimming)
- Extracting alignment (optimizing)
- Iterating
- Learning

20

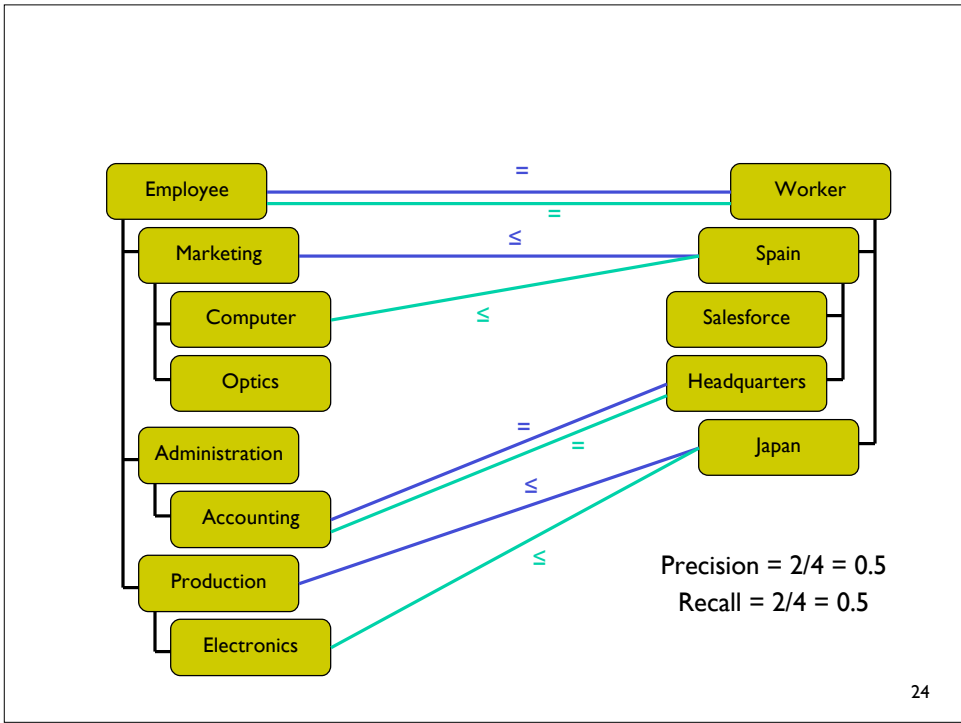
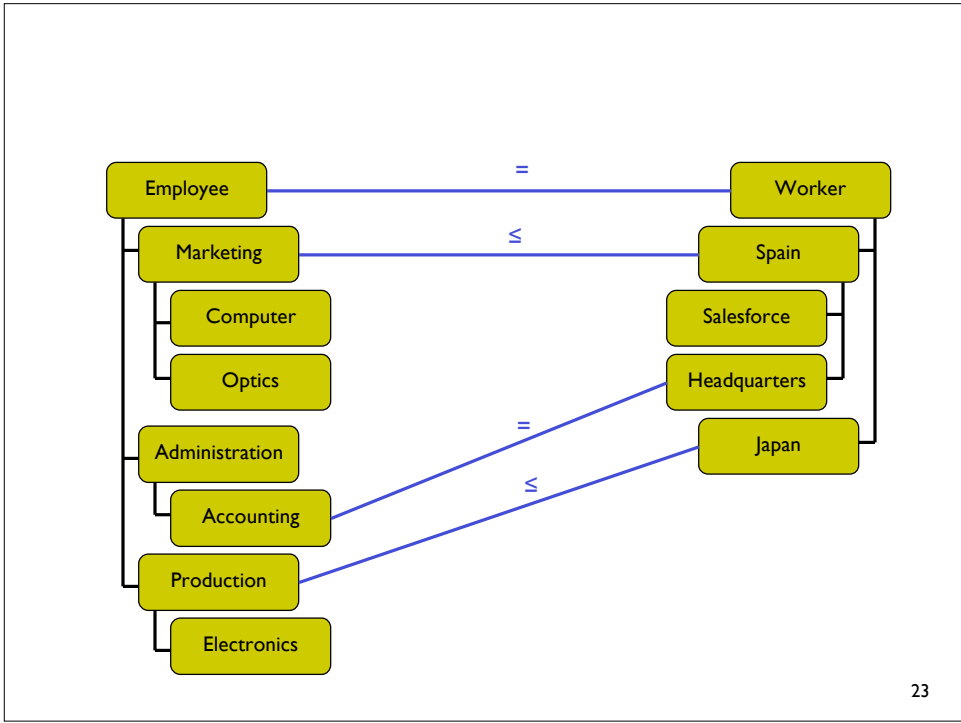
# Evaluation Metrics

- Precision and Recall
  - Classical measures for Information Retrieval systems
  - Provide a measure of how well a system performs
- Usually couched in terms of some “Gold Standard” or ideal result set
- Precision
  - How many of the answers are “right”?
  - Precision =  $|A \cap GS| / |A|$
- Recall
  - How many of the possible “right” answers does the system provide?
  - Recall =  $|A \cap GS| / |GS|$
- Usually trade off between precision and recall

21



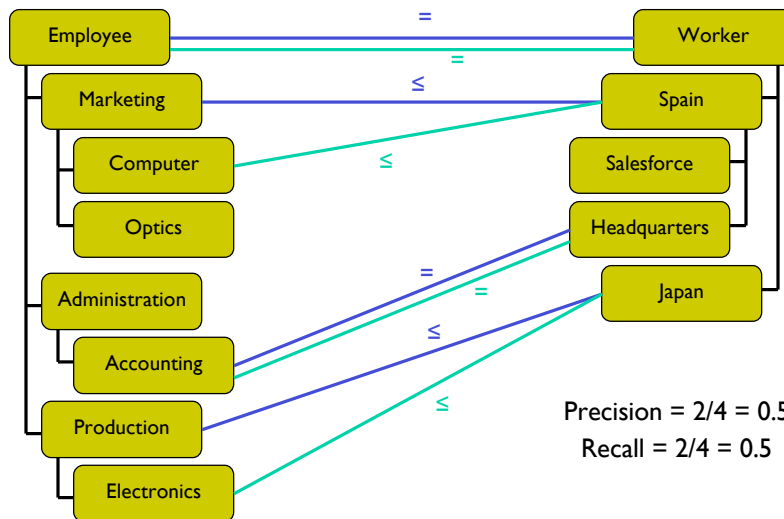
22



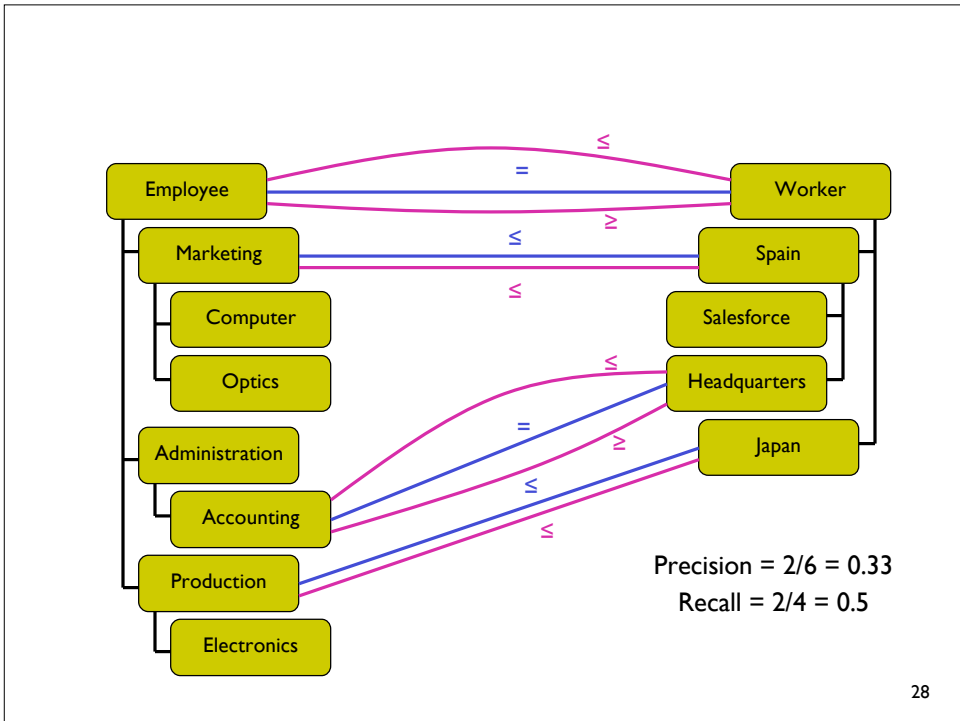
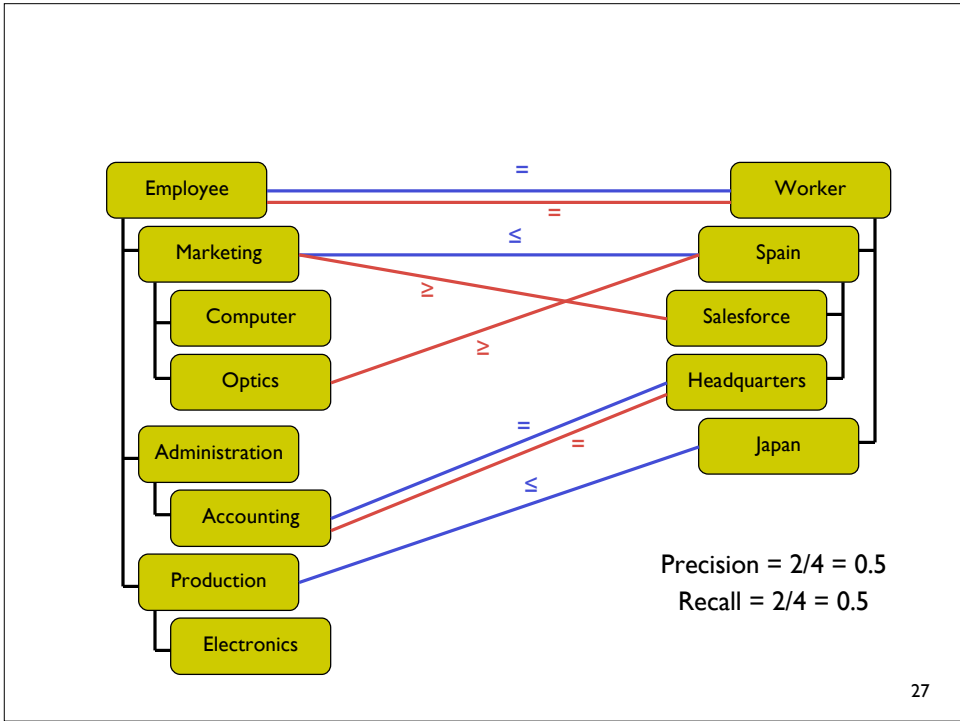
# Evaluation Metrics

- Precision and Recall are rather course grained measures for alignments
- They don't distinguish between a bad alignment and one that's almost right
  - Intolerant
- They don't take into account the semantics.
  - Equivalences

25



26



## Evaluation Metrics

- Precision and Recall are rather course grained measures for alignments
- They don't distinguish between a bad alignment and one that's almost right
  - Intolerant
- They don't take into account the semantics.
  - Equivalences
- Metrics need to be developed that take into account the semantics
  - Interesting research topic....

29

## How well do these approaches work?

- Ontology Alignment Evaluation Initiative
- Formal comparative evaluation of different ontology-matching tools
- Run every year
- Variety of test cases (in size, in formalism, in content)
- Results very dependent on the tasks and the data (from under 50% of precision and recall to well over 80% if ontologies are relatively similar)
- Results consistent across test cases
- Progress every year!



30

## Matching and Alignment Tools

- Frameworks
  - PROMPT (a Protégé plug-in): includes a user interface and a plug-in architecture
  - Alignment API: used by many tools in OAEI provides an exchange format and evaluation tools
  - COMA++: oriented toward database integration (many basic algorithms implemented).
- Matching systems
  - OAEI best performers (Falcon, RiMOM, etc.)
  - Available systems (FOAM, OLA, Rondo, etc.)
  - ...

31

## Current challenges

- How do we help users perform the alignments interactively?
- How do we explain the alignments that the tools create?
- How do we have system working across all cases? Do we need to?
- Can we use imperfect or inconsistent alignments?
- How do we maintain the alignments when ontologies evolve?

32

## Current challenges (cont'd)

- Design space of alignment approaches
  - Can we create a “toolbox for designing alignment approaches that fit a given problem?”
  - We have identified some components, but how can we bring them together?
- Have we discovered a “ceiling” in automatic discovery of alignments?
  - Will it be “lots of work for little gain” from now on?
  - Are there serious untapped resources?

33

## Further reading

- “Ontology Matching” by Euzenat and Shvaiko
- Proceedings of ISWC, ASWC, ESWC, WWW conferences, etc.
- Journal of web semantics, Journal on data semantics, etc.
- <http://www.ontologymatching.org>



34