KDI
Introduction

Fausto Giunchiglia and Mattia Fumagallli

University of Trento

Course objectives and contents

The goal of this course is to provide motivations, definitions, theorems and techinques for an effective modeling of knowledge and data, and for their integration.

The course will have succeeded if it provides to the students the basic techniques knowledge and data modeling, and stimulates their interest to continue their career with higher interest into data and knowledge representation in their own field of expertise, and to produce computer-processable solutions of relevant problems.

Method

Theory. Introduction to the most important notions and resources, (semantic-heterogeneity, representation languages, ontologies, lexical semantics, WordNet, ...).

Tools. Introduction to the basic tools for knowledge and data modelling.

Exercises. The students are asked to apply the basic notions and techniques and to prove that they are able to address a simple knowledge-data modeling and integration task.

Syllabus

Due to exceptional situations, the Dates could be modified

INTRODUCTION
Introduction to the Course "Knowledge and Data Integration"
Semantic Heterogeneity Problem (OpenData + GeoData)
SOA Solution: Language resources and ontologies
METHODOLOGY
Data Integration Methodology (4 knowldge levels + Informal modelling)
Data Integration Methodology (Formal modelling + Importing)
INFORMAL MODELLING
Informal Modelling: Competency Queries + Data Selection
Informal Modelling. EER
Practical class (queries + data selection + yEd)
FORMAL MODELLING
Formal Modelling: Knowledge - OWL
Formal Modelling: Knowledge - RDF
Formal Modelling: Knowledge - Protégé
Practical class (Ontology Design with Protégé)
Practical class (Ontology Design with Protégé)
Formal Modelling: Language - DERA
Formal Modelling: Language - WordNet
Practical class (DERA + WordNet + Protégé)
IMPORTING AND QUERIES
Importing (schema macthing and entity matching)
Formal importing (sparql queries + formal validation)
Formal importing (sparql queries + formal validation) - hands on in protege
Informal importing
Informal importing - hands on in KARMA
FINAL EXAM

Course Website

http://disi.unitn.it/~ldkr/ldkr2017/index.html

People

- Fausto Giunchiglia

  website: http://disi.unitn.it/~fausto/

  email: fausto@dit.unitn.it

- Mattia Fumagalli

  email: mattia.fumagalli@unitn.it

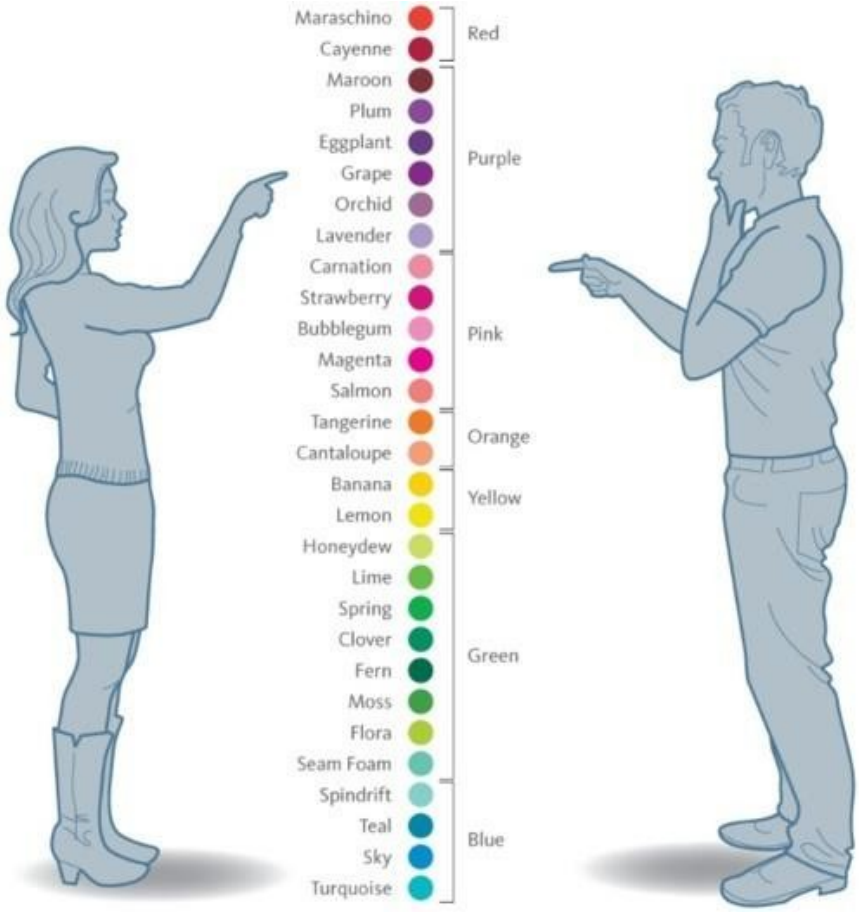
- Subhashis Das

  email: subhashis.das@unitn.it

Exam

- Report submission (9-10 pages document) [9 Marks]
  - Date of submission “to be defined”
- Report submission (9-10 pages document) [9 Marks]
  - Date of submission “to be defined”
- PPT presentation [9 Marks]
  - Dates of presentation “to be defined”
- Q&A and Evaluation [6 Marks]

The Semantic Heterogeneity Problem



The difficulty of establishing a certain level of connectivity between people, software agents or IT systems [Uschold & Gruninger, 2004] at the purpose of enabling each of the parties to appropriately understand the echanged information [Pollock, 2002]

The Semantic Heterogeneity Problem (Examples)

Data about P/E ratio for a company obtained from four different financial information sources [Madnick Stuart, 2003]

"Which source is correct?"

Source	P/E Ratio
ABC	11.6
Bloomberg	5.57
DBC	19.19
MarketGuide	7.46

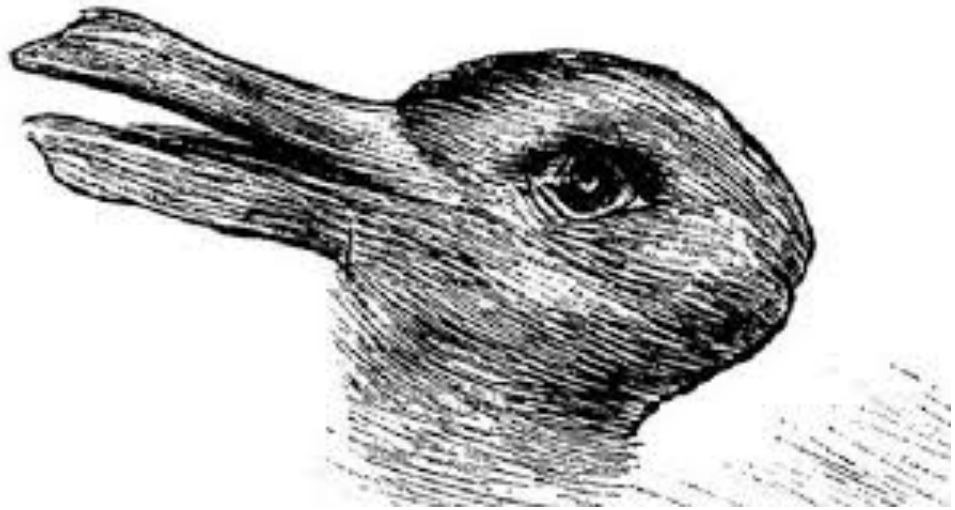
they are all correct!

The Semantic Heterogeneity Problem (Examples)

Is it a picture of a duck or a rabbit?

The point here is that some will see it one way, some will see it the other way, and most be able to see both images – but only one at a time.

This is the situation that we often face in real life. There is often no "right" answer and different people will continue to see things in different ways. Merely saying that everyone should see it the same way does not change the reality that multiple different legitimate, and often essential, views exist.



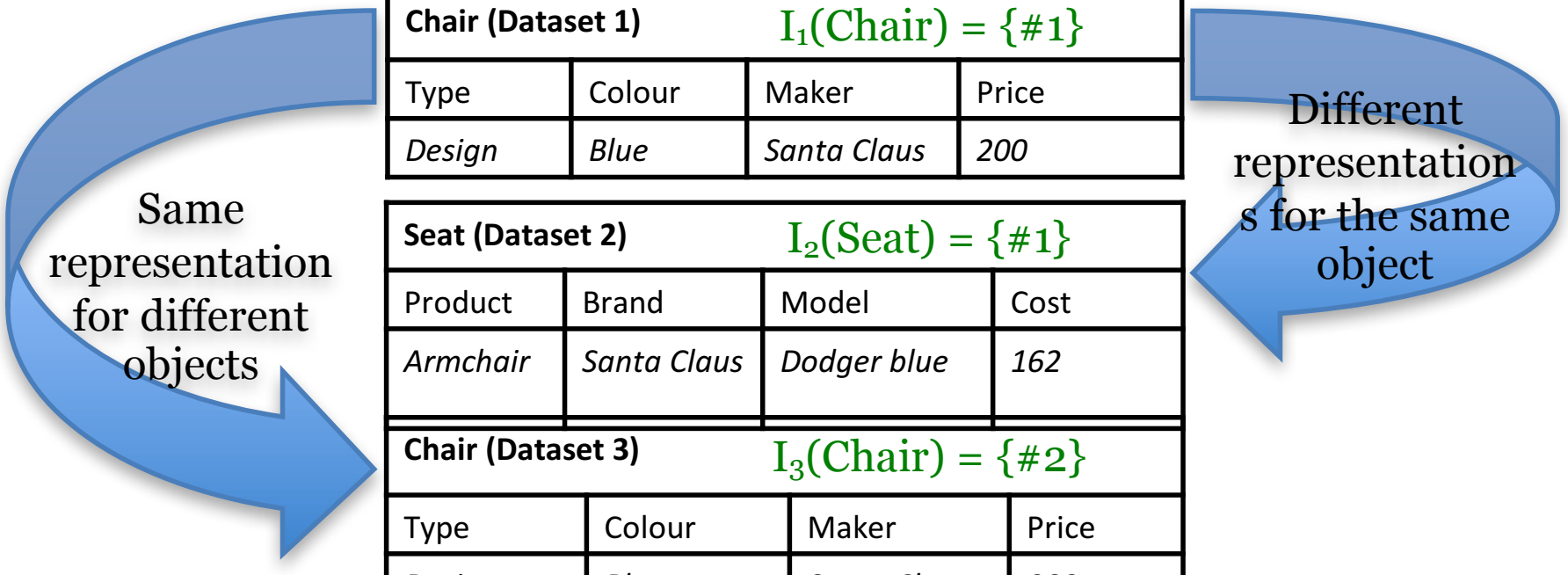
The Semantic Heterogeneity Problem (Examples)

"What makes #1 different from #2?"

"What makes #1 #1 and #2 #2?"

The Semantic Heterogeneity Problem (Examples)

$\Delta = \{\#1, \#2, \ldots\}$



Chair (Dataset 1) $I_1(\text{Chair}) = \{\#1\}$

Type	Colour	Maker	Price
Design	*Blue*	*Santa Claus*	*200*

Seat (Dataset 2) $I_2(\text{Seat}) = \{\#1\}$

Product	Brand	Model	Cost
Armchair	*Santa Claus*	*Dodger blue*	*162*

Chair (Dataset 3) $I_3(\text{Chair}) = \{\#2\}$

Type	Colour	Maker	Price
Design	*Blue*	*Santa Claus*	*200*

Kinds of Heterogeneity

Hardware and Operating Systems
Data Management Software
Data Models, Schemas and Semantic
Middle-ware
User Interfaces
Business Rules and Integrity Constraints
...

Knowledge

Language

Reasons for integration

- facilitate information access and reuse through a single information access point
- data from different complementing information systems can be combined to gain a more comprehensive basis to satisfy a need
- ...

Application Examples

Integrated information can be used for querying and reporting for:

- Statistical Analysis
- OLAP
- Data Mining
- …

In order to enable

- Forecasting
- Decision Making
- Enterprise-wide Planning
- …

Types of integrations

Manual integration. Users directly interact with all relevant information systems and manually integrate selected data

User interface. The user is supplied with a common user interface (e.g. a web browser) that provides a uniform look and feel

Application. Integration of applications that access various data sources and return integrated results to the user

Types of integrations

Middle-ware. Middleware provides functionality used to solve aspects of the integration problem

Uniform access. A logical integration of data is accomplished at the data access level

Common storage. Physical data integration is performed by transferring data to a new data storage

#MediatedQuerySystems

- Represent a uniform data access solution by providing a single point for read-only querying access to various data sources

- Uses a mediator that contains a global query processor to send sub-queries to local data sources; returned local query results are then combined

#Portals

- Another form of uniform data access are personalized doorways to the internet or intranet
- Each user is provided with information tailored to his information needs
- Web mining is applied to determine user-profiles by click-stream analysis

Examples of integration

#DataWarehouses

- Realize a common data storage approach
- Data from several operational sources(OLTP) are extracted, transformed, and loaded (ETL) into a data warehouse
- Analysis, such as OLAP, can be performed on cubes of integrated and aggregated data

Examples of integration

#FederatedDatabaseSystems(FDBMS)

- Achieve a uniform data access (meta) solution by logically integrating data from underlying local DBMS
- Implement their own data model, support global queries, global transactions, and global access control

#WorkflowManagementSystems(WFMS)

- Represent an integration-by-application approach
- Allow to implement business processes where each single step is executed by a different application or user
- Support modeling, execution, and maintenance of processes that are comprised of interactions between applications and human users

#IntegrationbyWebServices

- Performs integration through software components (web services) that support machine-to-machine interaction by XML-based messages conveyed by internet protocols

- Depending on offered integration functionality either represent: i) a uniform data access approach, or ii) a common data access for later manual or application-based integration

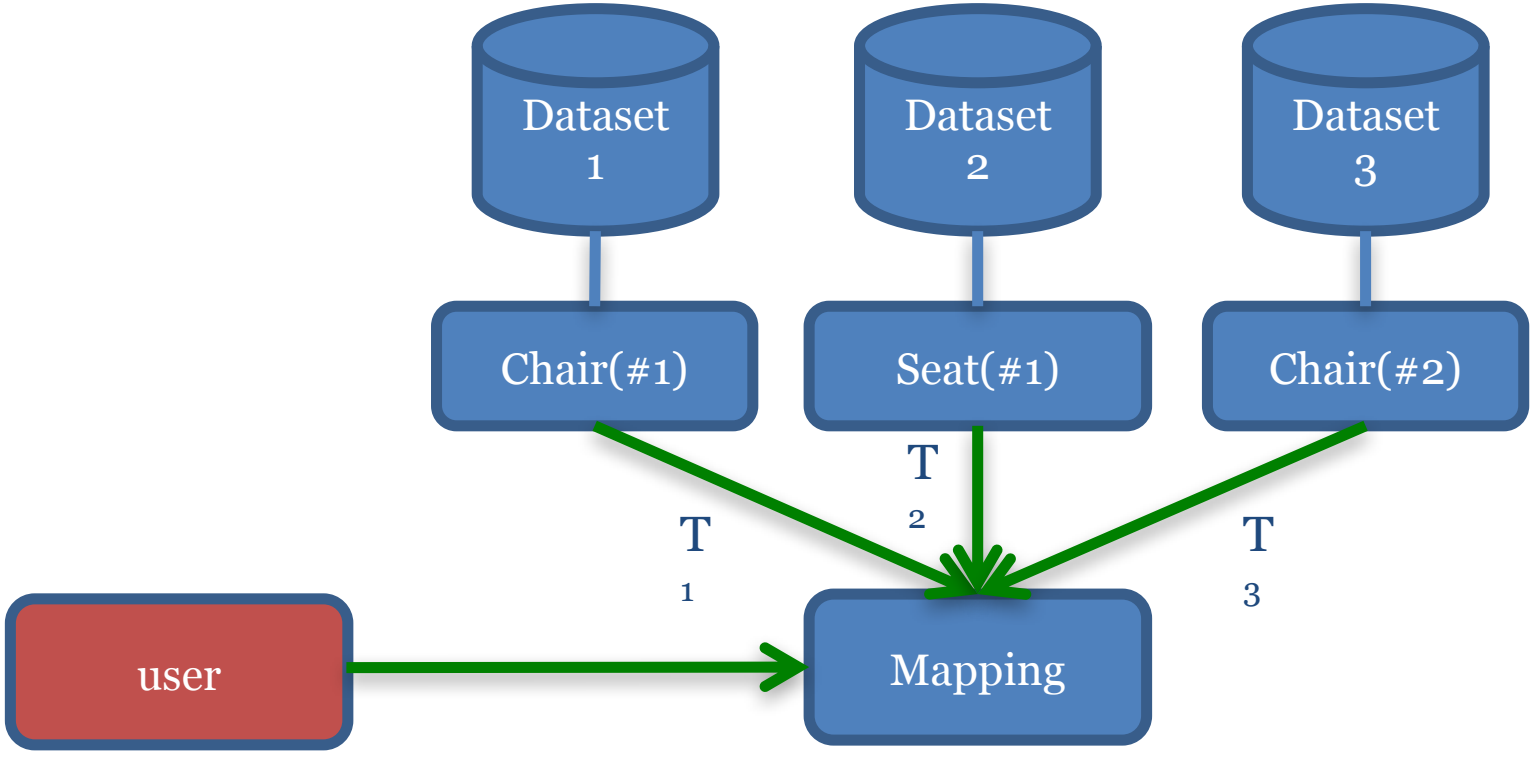
#Peer-to-Peer(P2P)Integration

- A decentralized approach to integration between distributed peers where data can be mutually shared and integrated

- Depending on offered integration functionality either represent: i) a uniform data access approach, or ii) a common data access for later manual or application-based integration

Examples of integration

#Semantic integration...



Exercises

1. Search on the Web information about how many languages are spoken in Europe and in the whole world.
2. What is the most widely spoken language in the world?
3. Provide an example of concept which is heavily cultural dependant.
4. What are the top level entity types (up to 10) that to you are necessary to codify the whole world knowledge?