KDI : Knowledge and Data Integration

A World of Data

Phase: 1. Introduction & Representation Diversity

W1.L12.M1.T211

Contents

Contents

Unstructured Data

Data without any metadata or structure.

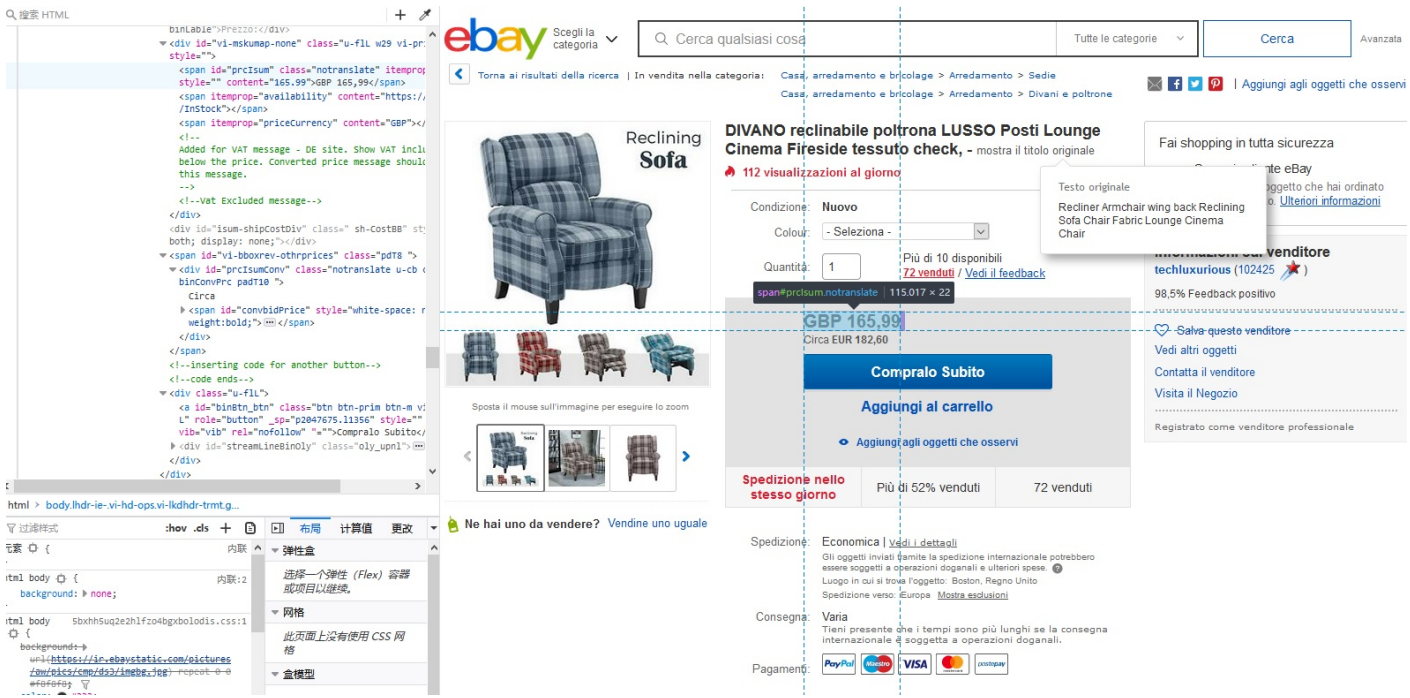
- Text
- Figure (without metadata)
- Audio (without metadata)
- Video (without metadata)

Unstructured data example: figures

Semi-structured Data

Data with some metadata (such as tags but without formal semantics).

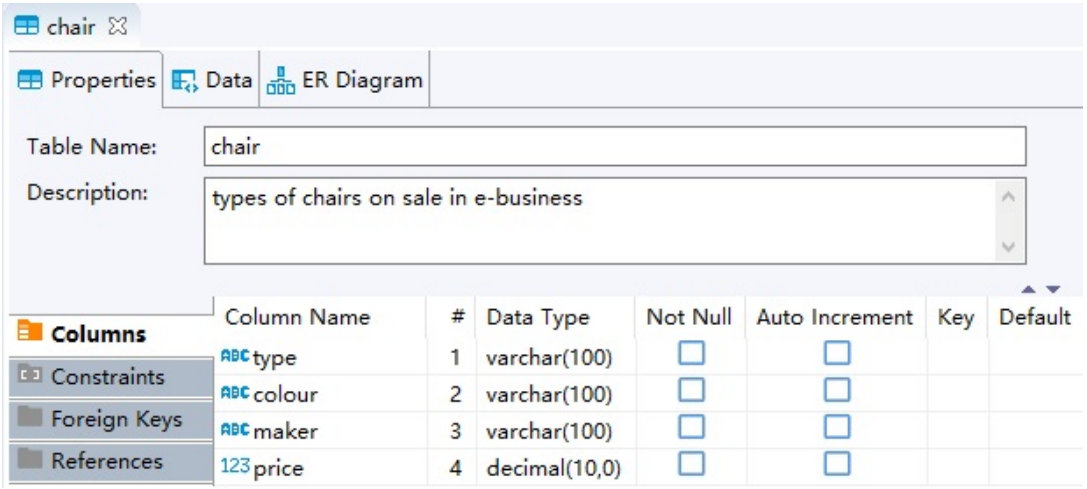
- HTML
- XML
- RDF

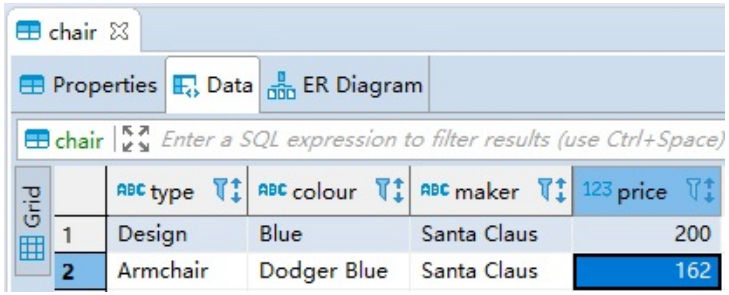


Semi-structured data example: html file

Structured Data

Data in formal structures (such as in related database tables).



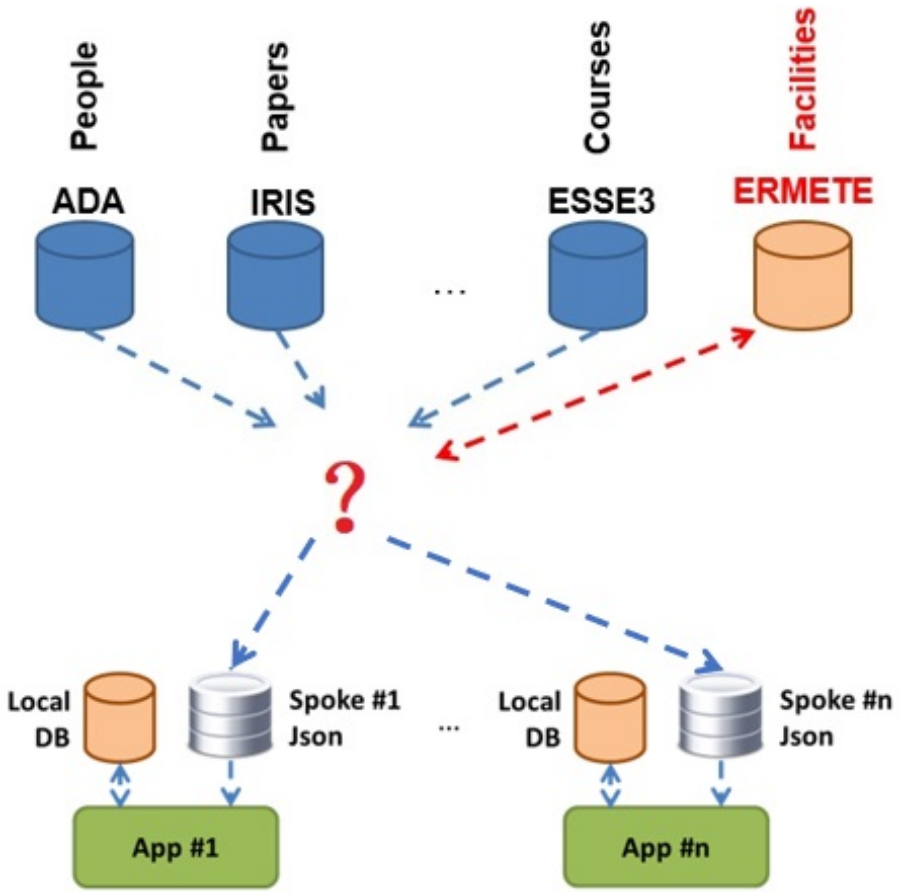


Structured data example: schemas and records in database tables

Contents

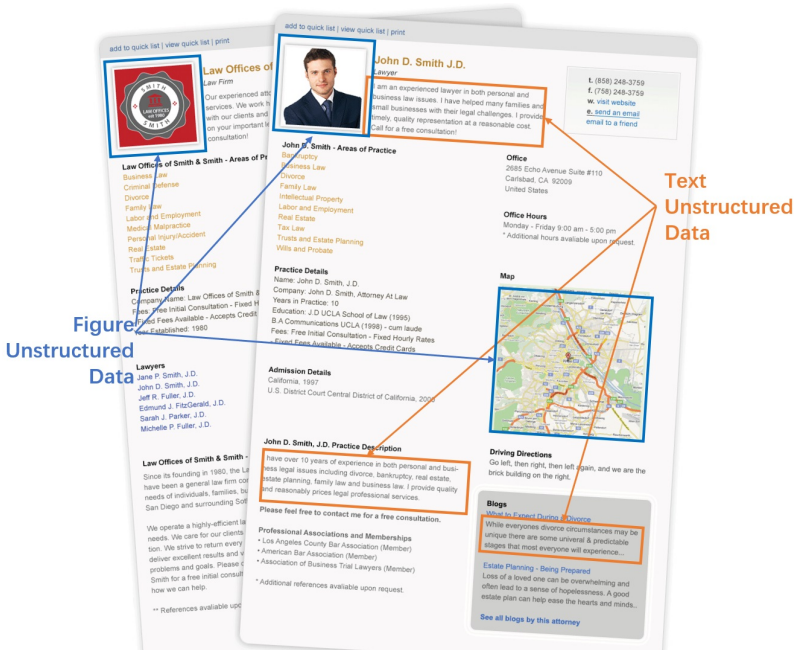
Case Study

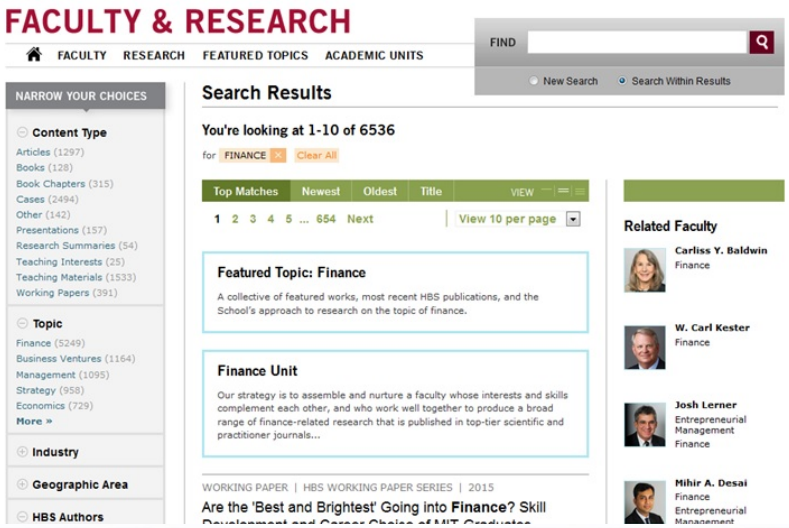
Digital University (DU) is an application running at the University of Trento which provides **extensive query answering services** across existing resources, such as ADA for people, IRIS for papers, ESSE3 for courses, ERMETE for faculties, etc.



Extensive Query Services

Examples of the extensive queries across multiple legacy systems: *university profile generation, research activity explore, student performance assessment, etc.*





Contents

Diversity Everywhere

The data on the web or within the corporate organization are represented in unstructured, semi-structured and/or structured fashions.

For the same entity in the real world, it can be represented ...

- in different languages.
- with different tags.
- with different predefined attributes in databases.

For the different entities in the real world, they can be represented ...

- with the same language symbols.
- with the same tags and tag values.
- with the same record values.

In conclusion, in the world of data, diversity is everywhere.

KDI : Knowledge and Data Integration

W1.L12.M1.T211

A World of Data

Phase: 1. Introduction & Representation Diversity