Wrocław University
of Science and Technology

Topic: Semi - structured, RDF, XML Databases

228884,
Mateusz Guściora

HR EXCELLENCE IN RESEARCH

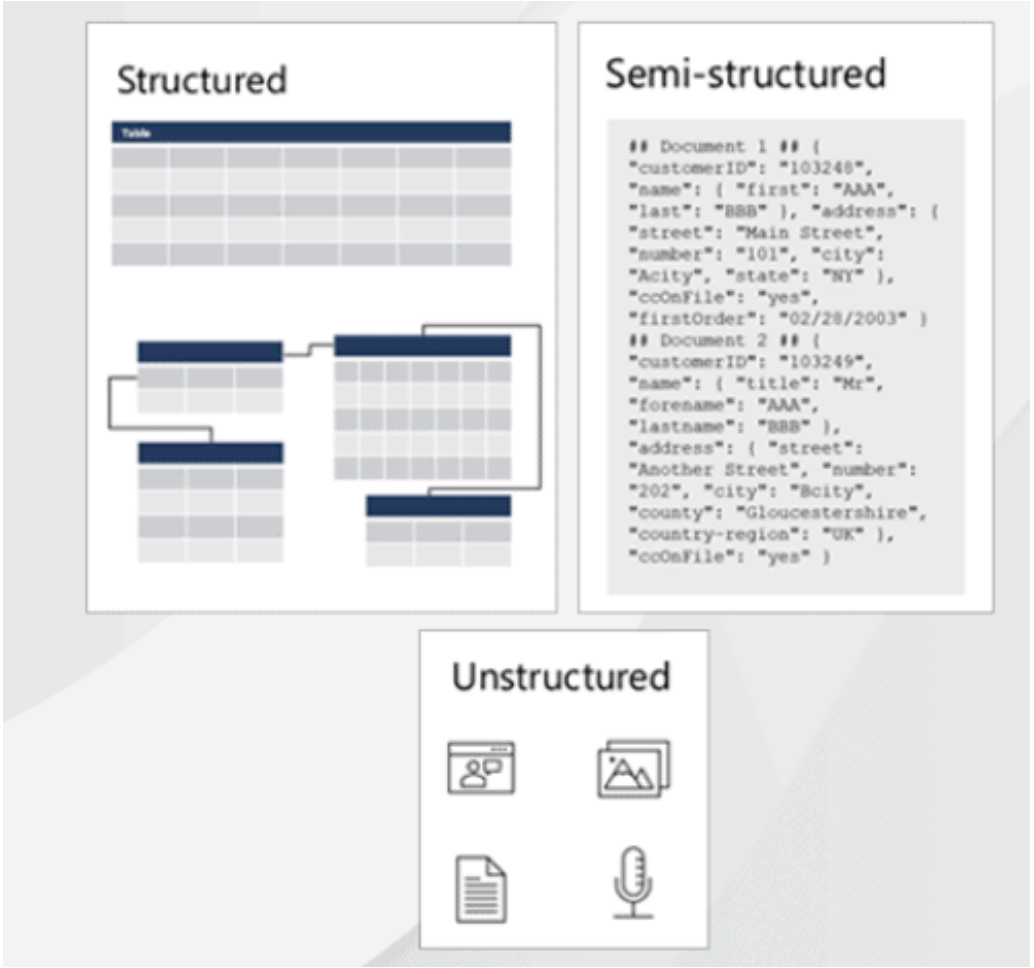
Agenda

1. Semi-structured data
2. RDF
3. XML databases
4. Questions/Discussion

Semi-structured data

- Semi-structured data is a mix of both structured and unstructured data.

not obey the tabular structure of data models but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies



Characteristics of Semi-structured data

- Data does not conforms to a data model but has some structure (lack of a well defined structure)
- Semi-structured data contains tags or other markers and elements
- Entities in the same group may or may not have the same attributes or properties
- Does not contains sufficient metadata which makes automation and management of data difficult

Examples (sources of semi-structured data): XML and other markup languages, JSON, emails, call center log, web pages, Integration of data from different sources.

Problems and Solutions

Disadvantage:

- Lack of fixed, rigid schema make it difficult in storage of the data
- Interpreting the relationship between data is difficult as there is no separation of the schema and the data.
- Queries are less efficient

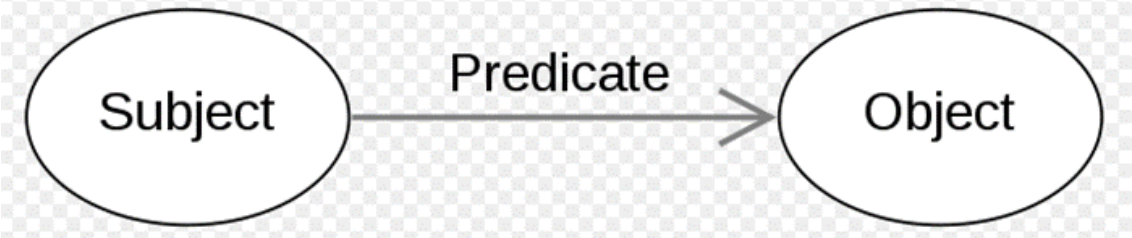
Possible solutions:

- XML is widely used to store and exchange semi-structured data.
- Object Exchange Model (OEM) can be used to store and exchange semi-structured data. OEM structures data in form of graph.
- DBMS specially designed to store semi-structured data or RDBMS can be used to store the data by mapping the data to relational schema and then mapping it to a table

RDF – Resource Description Framework

- method for describing Web resources / represent data, with an XML-based syntax. Specifically design to store meaning allongside the data

- RDF triples



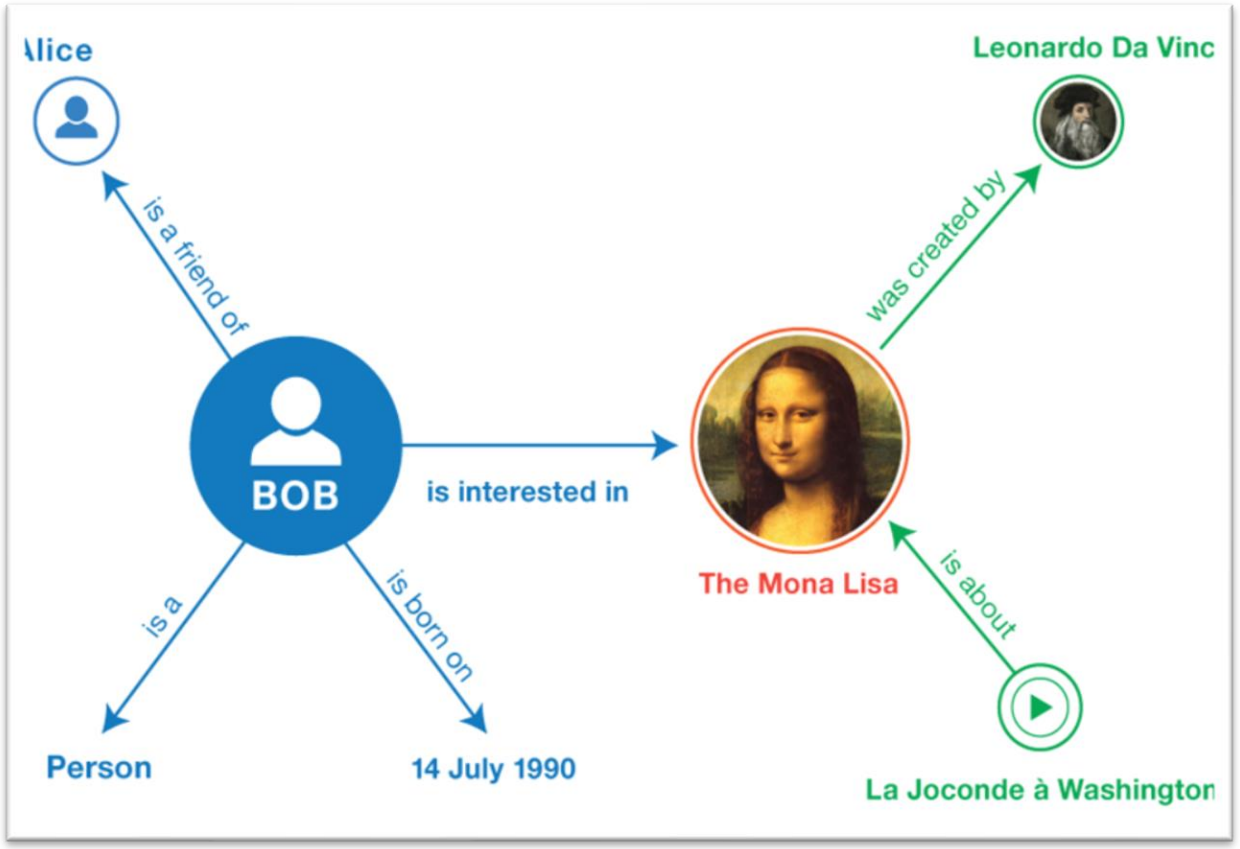
RDF graph

<Bob> <is a> <person>

<Bob> <is a friend of> <Alice>

<Bob> <is interested in> <Mona Lisa>

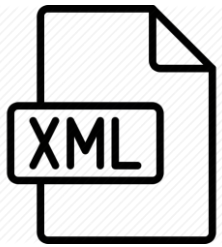
…



XML - eXtensible Markup Language

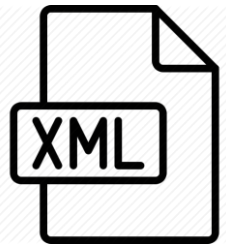
- data format/data model for sharing data between systems. Its form is hierarchy and one form of representation is textual !1

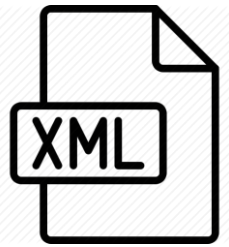

```
<?xml version="1.0" encoding="UTF-8" ?>
<Recipe TimeToPrepare="5" CookMethod="Grill" Difficulty="Easy" Serves="1"
Category="Entrees">
 <Title>Hot Dog with Onions</Title>
 <Ingredients>
```



XML databases

- „An XML database is a data persistence software system that allows data to be specified, and sometimes stored, in XML format. This data can be queried, transformed, exported and returned to a calling system” (wikipedia)
- XML databases types:
  - RDBMS
  - Native




XML

Questions/Discussions/Conclusions

Sources

- Practical RDF: Solving Problems with the Resource Description Framework: Shelley Powers, 2003
- Principles of Distributed Database Systems, M. Tamer Özsu • Patrick Valduriez, III edition, 2011
- Physical Database Design, Sam Lightstone, Toby Teorey, Tom Nadeau, 2007

- https://en.wikipedia.org/wiki/XML_database
- https://www.w3.org/
- https://onlineappsdba.com/index.php/2020/10/21/structured-data-vs-unstructured-data-vs-semi-structured-data/
- https://www.geeksforgeeks.org/what-is-semi-structured-data/
- https://www.scibite.com/news/scibite-rdf-resource-description-framework-a-natural-semantic-fit/

thank you