

Semantic web

UZH BIO390

Ahmad Aghaebrahimian

Zurich University of Applied Sciences
agha@zhaw.ch

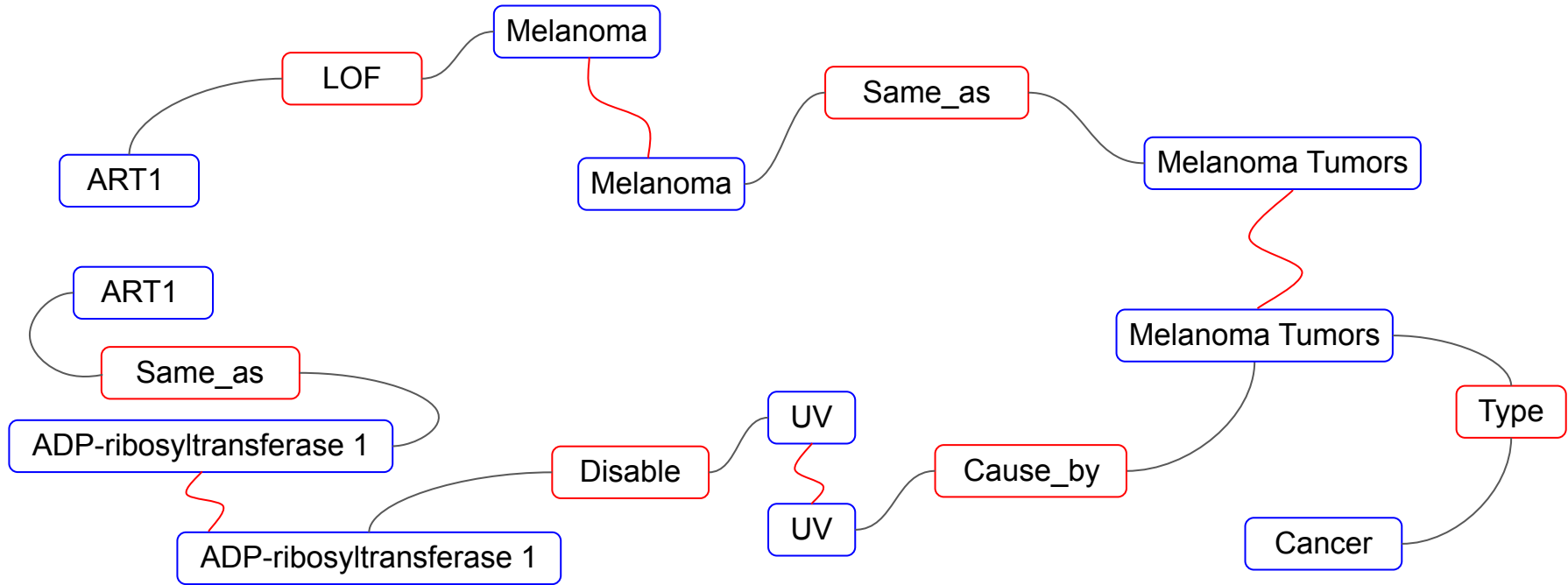
Introduction

- Senior researcher at ICLS-ZHAW
- Ph.D. Computer Sciences
- Area of interests:
 - Artificial Intelligence
 - Deep Learning
 - Semantic Web
 - Biomedical text analytics
 - Natural Language Processing
- Contact: agha@zhaw.ch

Content

1. Introduction
2. Stack of standards
 - a. URI
 - b. XML
 - c. RDF
 - d. OWL
3. Ontology
4. Knowledge graphs
5. SPARQL: <https://bit.ly/3wjyHpf>

Motivation



1- Introduction

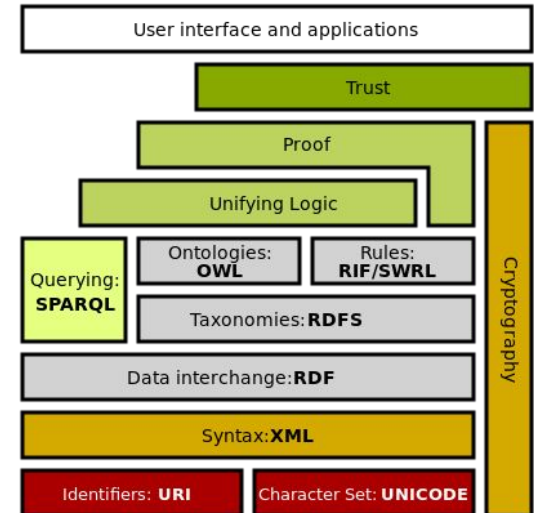
- *Web structure:*
 - Server vs. Client
- *Web Components:*
 - Uniform Resource Locator (URL): identify a document
 - Hypertext Markup Language (HTML): present a document
 - Hypertext Transfer Protocol (HTTP): transfer document
- *Moving from pages to resources*
 - Interactive web, Web 2.0 or semantic web

1- Introduction

- *Semantic Web:*

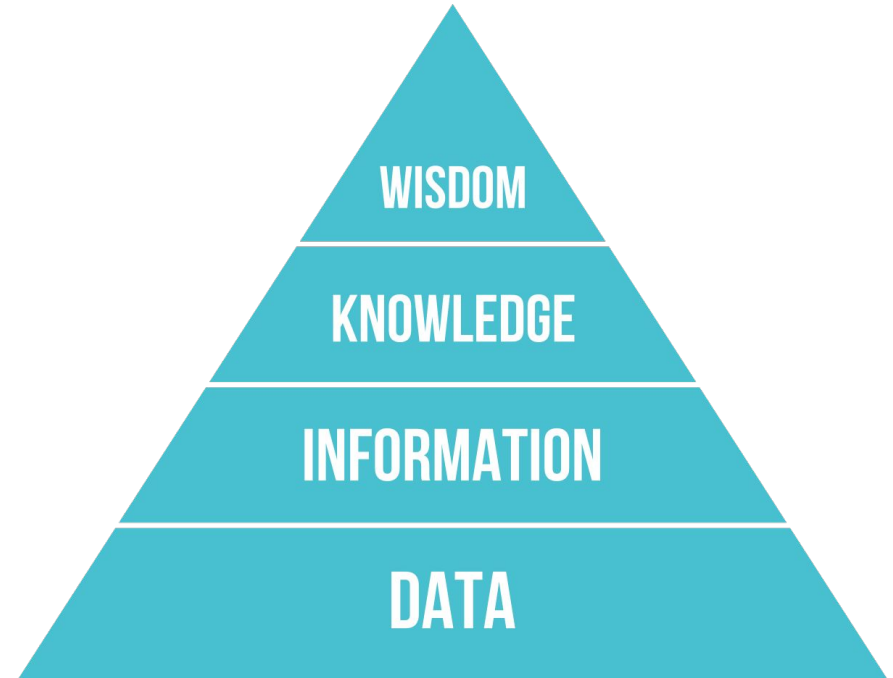
an extension of the W3 that uses the **Resource Description Framework (RDF)** and

Web Ontology Language (OWL), among other standards, to make the Internet **machine-readable**.



1- Introduction

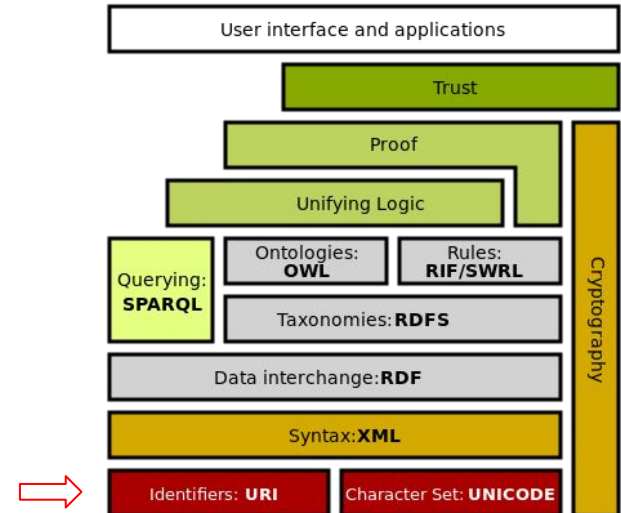
- *Benefits:*
 - Presenting knowledge about data
 - Allowing data integration from data silos
 - Introduce intelligence to systems



2- Standards: URI

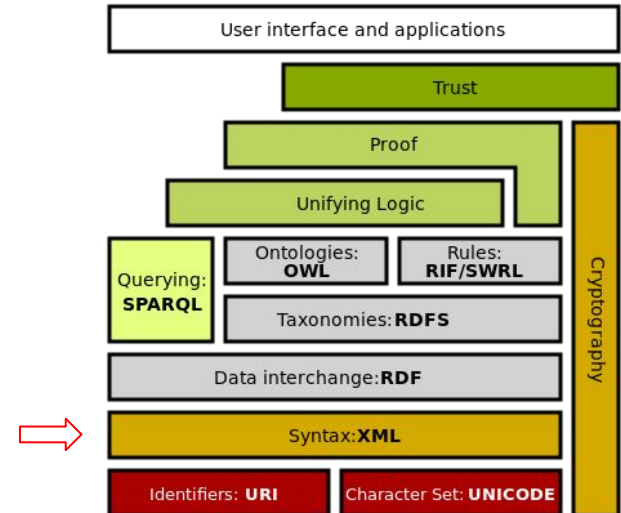
- *Identifying Resources:*

URL \Rightarrow URI \Rightarrow IRI



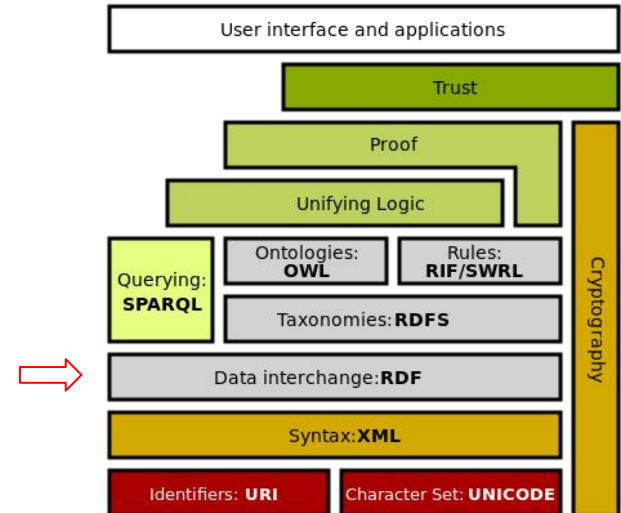
2- Standards: XML

- **Open** family of languages that represents **structured** data using **tags** and in **textual** format
- Principles
 - Only one root <root> </root>
 - Opening with closing <Gene></Gene>
 - No tag begin with number or xml
 - Case sensitive <Gene> != <gene>
 - Order matters <Gene> <nucl> </nucl></Gene>
 - Tags may have attributes <Gene inherited='true' />



2- Standards: RDF

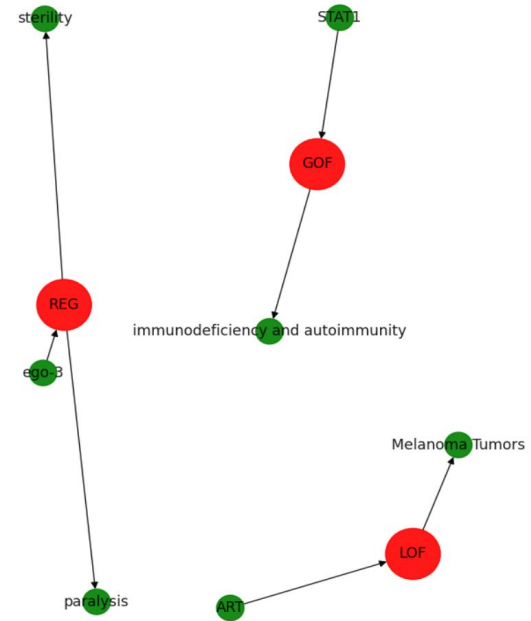
- A **graph data model** and a set of syntax that allows us to describe web resources.
- Principles
 - Triple structure: (subject, predicate, object)
 - Subject: URI entity **resource**
 - Predicate: URI relationship **resource**
 - Object: URI **resource** or **literal**
 - Predicates are **labeled** and **directed**
- Serialization
 - XML, N-triple, Turtle, TriG, JSON-LD



2- Standards: RDF

- **Graph data model**
- Example: Active Gene Annotation Corpus (AGAC)
 - Triple Expanded and Interlinked
 - ego-3 , REG , sterility
 - ego-3 , REG , paralysis
 - Triples can be queried by SPARQL

```
SELECT ?gene ?relation
WHERE {
    ?gene ?relation Melanoma_Tumors .
}
```



2- Standards: RDF

- **RDF data generation**

- Named Entity Recognition
- Named Entity Disambiguation
 - Aghaebrahimian, A., Cieliebak, M.(2020), Named Entity Disambiguation at Scale, ANNPR, Winterthur, Switzerland
- Relation Extraction
 - Aghaebrahimian, A. and Jurcicek, F., (2016), Open-domain Factoid Question Answering via Knowledge Graph Search, NAACL, San Diego, USA
 - Aghaebrahimian, A., Anisimova, M., & Gil M., Ontology-Aware Biomedical Relation Extraction, TSD, Springer - LNAI

Atorvastatin lowers LDL and triglycerides and raises HDL in the blood.

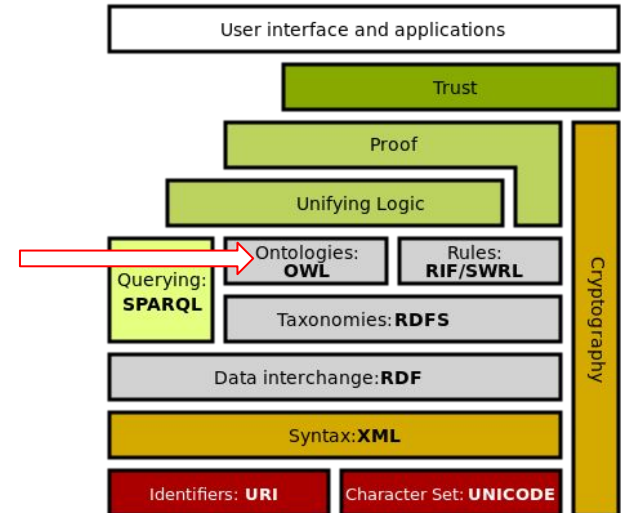
CHEBI:39548 lowers CHEBI:47774 and IUPAC:46823 and raises CHEBI:47775 in the blood.

CHEBI:39548 lowers CHEBI:47774 and IUPAC:46823 and raises CHEBI:47775 in the blood.



2- Standards: OWL

- Provides a rich **vocabulary** to add **semantics** and context
- Allows reasoning and inference



3- Ontology

- Ontology is
 - A model of a domain
 - A vocabulary consisting of classes and properties
 - Machine-readable knowledge representation
- Ontology (mostly) consists of
 - Definitions
 - Concepts
 - Relationships
- Benefits
 - Precise meaning
 - Data integration
 - Language processing

3- Ontology

- OBO: ChEBI

[Term]

id: CHEBI:60871

name: selenium(2+)

def: "The selenium ion with two positive charges."

synonym: "Se(2+)" RELATED [UniProt:]

synonym: "[Se++]" RELATED SMILES [ChEBI:]

synonym: "InChI=1S/Se/q+2" RELATED InChI [ChEBI:]

synonym: "InChIKey=MFSBVGSNNPNWMD-UHFFFAOYSA-N"
RELATED InChIKey [ChEBI:]

is_a: CHEBI:60250

is_a: CHEBI:30412

[Term]

id: CHEBI:60250

name: selenium ion

def: "A selenium atom having a net electric charge."

is_a: CHEBI:36904

is_a: CHEBI:36914

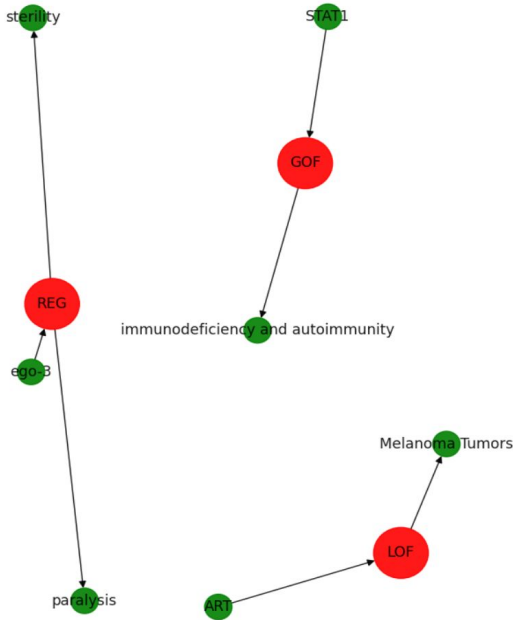
3- Ontology

- UMLS
 - Metathesaurus
 - Semantic network
 - Specialized Lexicon



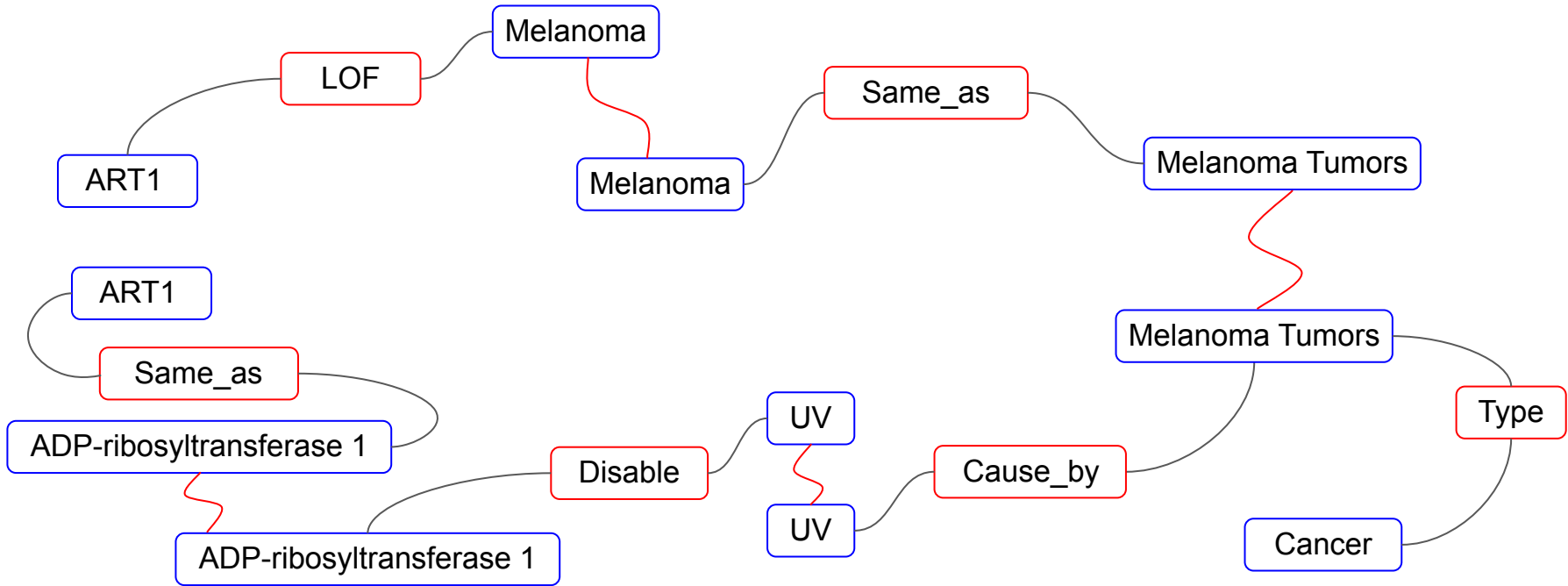
4- Knowledge graph

- Billions of triplets modeled in the RDF model in a graph

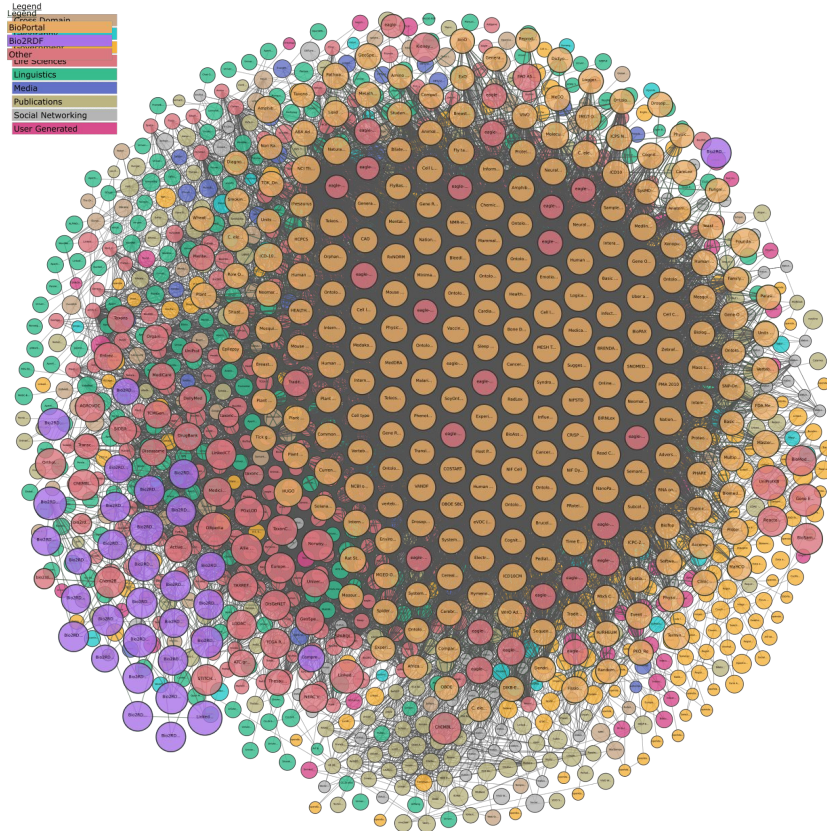


- Which genes are related to paralysis?
- How STAT1 impacts the immune system?
- Literature Based Discovery ([mock up](#))

Linked Open Data



Linked Open Data



5- SPARQL

- SPARQL Protocol And RDF Query Language
- Lab work: <https://bit.ly/3wjyHpf>

