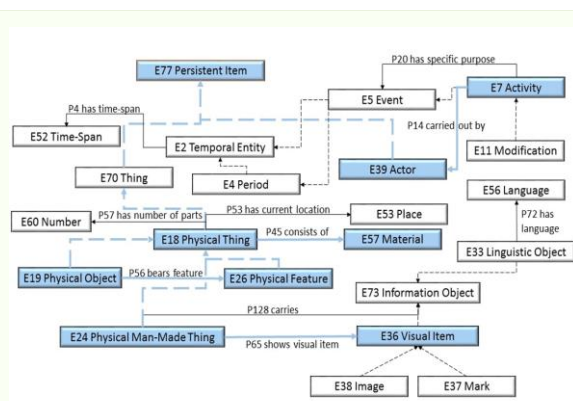# *RDF Digest* (Effective Summarization of RDF/S KBs)

Overview

Ontology summarization aspires to produce an abridged version of the original ontology that highlights its most representative concepts. RDF Digest is a novel platform that automatically produces and visualizes summaries of RDF/S Knowledge Bases (KBs). A summary is a valid RDFS document/graph that includes the most representative concepts of the schema, adapted to the corresponding instances.

To construct this graph the corresponding algorithm exploits the semantics and the structure of the schema and the distribution of the corresponding data/instances. Initially to identify the most important nodes the system is using the notion of *relevance*. Then the system selects the edges connecting these nodes by trying to maximize either locally or globally the importance of the selected edges.

A novel feature of the platform is in addition, that it allows summary exploration through extensible summaries.

Selecting the most important nodes from CIDOC-CRM
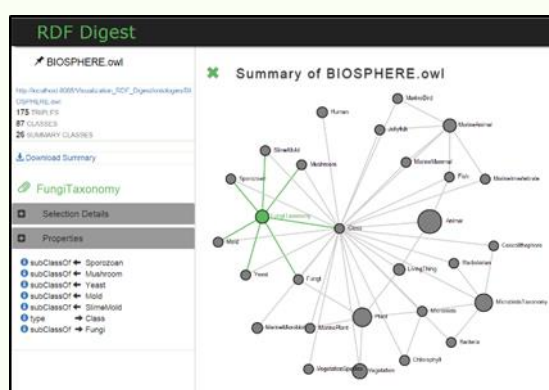
Summarizing BIOSHPERE Ontology using RDF Digest

Target Domains

- ❖ Ontology Engineering
- ❖ Quick Understanding and exploration of RDF/S KBs

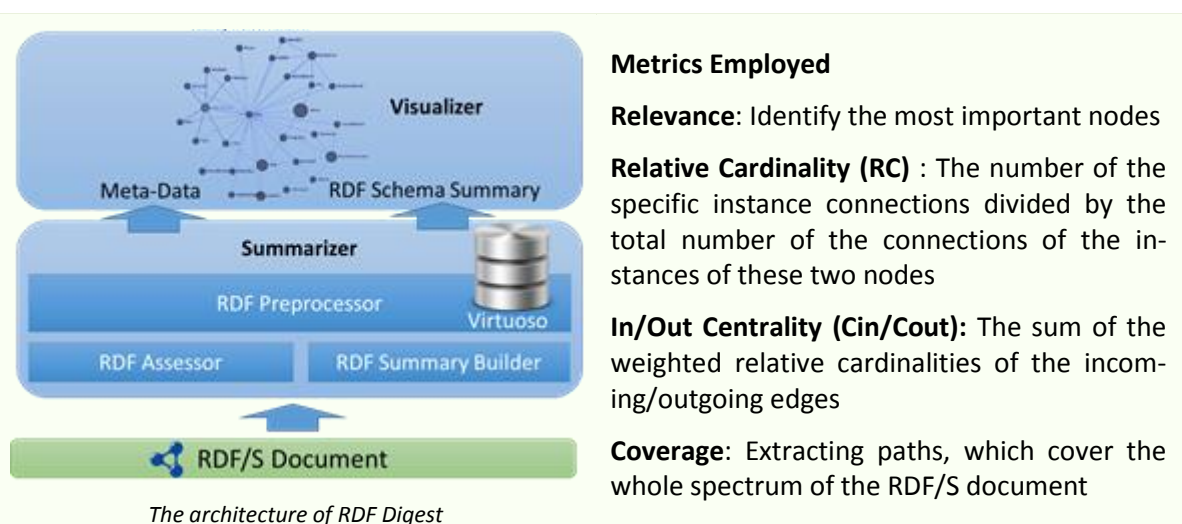Pilot Installations in the following EU research projects

- ❖ iManageCancer - Empowering patients and strengthening self-management in cancer diseases (H2020).
- ❖ MyHealhAvatar - A Demonstration of 4D Digital Avatar Infrastructure for Access of Complete Patient Information (FP7-ICT-2011-9)
- ❖ eHealthMonitor - Intelligent Knowledge Platform for Personal Health Monitoring Services (FP7-ICT-2011-7)

Description

The RDF Digest is composed of two major components, the **Summarizer** and the **Visualizer**.

Using the interface, a user can select or give the URL of an online RDF/S document, she would like to be summarized and is optionally able to define the expected length of the summary. The Summarizer gets the input RDF/S document and pre-processes it (using the RDF Preprocessor module) by computing the corresponding RDF/S KB. The result is stored in a Virtuoso instance to enable efficient data access. Then, the RDF Accessor module calculates the relevance of each node. The RDF Summary Builder generates the final summary of the schema, based on the rankings produced by the RDF Assessor and the requested size of the summary. The result and additional meta-data are returned to the Visualizer which enables effective visualization of the summary and exploration of the data source.

By selecting and clicking on a node, **additional meta-data** (its relevance and centrality, the number of instances, the connected properties, instances etc.) are provided to enhance ontology understanding. Besides meta-data, further exploration of the data source is allowed by clicking further on the details (on the left) of the selected class and the properties. When clicked, its instances and connections appear in a pop-up window. Moreover, further exploration of the data source is allowed by double-clicking on a node to **extend the summary on that specific node**. Finally the user is able to download the summary as a valid RDFS document.

The architecture of RDF Digest

Metrics Employed

Relevance: Identify the most important nodes

Relative Cardinality (RC) : The number of the specific instance connections divided by the total number of the connections of the instances of these two nodes

In/Out Centrality (Cin/Cout): The sum of the weighted relative cardinalities of the incoming/outgoing edges

Coverage: Extracting paths, which cover the whole spectrum of the RDF/S document

Additional Information

Scientific Publications

❖ Ontology Understanding Without Tears: The Summarization Approach, Semantic Web Journal 2016

❖ RDF Digest: Efficient Summarization of RDF/S KBs, ESWC 2015

❖ RDF Digest: Ontology Exploration Using Summaries, ISWC 2015

RDF Digest online

Contact details:	Georgia Troullinou	Haridimos Kondylakis	Dimitris Plexousakis
	troulin@ics.forth.gr	kondylak@ics.forth.gr	dp@ics.forth.gr
	www.ics.forth.gr/isl		