

This tool is a research effort for annotating text and tabular information in a pdf document via crowdsourcing. It also facilitates a user to annotate tabular information using DataCube RDF vocabulary. This tool transforms selected pdf table info into data cube RDF.

ANNOTATION WINDOW

The screenshot shows a web browser interface for PDF annotation. On the left, a PDF document titled "Hello, world!" program is displayed. On the right, an annotation panel is visible with buttons for "Show Simple Annotate Panel", "Confirm annotation", and "Reset Selection". Below these buttons, a table of extracted information from the PDF is shown, with a callout indicating "Extracted table information from pdf".

DATA CUBE RDF SEGMENT OF TRANSFORMED TABLE INFO

http://example.org/ns/table1R3C2	http://eaisAnnotation/#value	"Birth information for the United States from 1969 to 2008"
http://example.org/ns/table1R5C2	http://eaisAnnotation/#value	"Revision information for Wikipedia articles"
http://example.org/ns/table1R1C2	http://eaisAnnotation/#value	"Samples from US weather stations since 1929"
http://example.org/ns/table1R4C2	http://eaisAnnotation/#value	"Word index for works of Shakespeare"
http://example.org/ns/table1R2C2	http://eaisAnnotation/#value	"Measurement data of broadband connection performance"
http://example.org/ns/table1R4C3	http://eaisAnnotation/#value	"164 Kilo"
http://example.org/ns/table1R2C3	http://eaisAnnotation/#value	"240 Billion"
http://example.org/ns/table1R5C3	http://eaisAnnotation/#value	"314 Million"
http://example.org/ns/table1R3C3	http://eaisAnnotation/#value	"68 Million"
http://example.org/ns/table1R1C1	http://eaisAnnotation/#value	"Gsoo sipa"
http://example.org/ns/table1R2C1	http://eaisAnnotation/#value	"Mlab abc"
http://example.org/ns/table1R3C1	http://eaisAnnotation/#value	"Natality sky"
http://example.org/ns/table1R4C1	http://eaisAnnotation/#value	"Shakes peare"
http://example.org/ns/table1R5C1	http://eaisAnnotation/#value	"Wiki pedia"
http://example.org/ns/table1R1C3	http://eaisAnnotation/#value	"115 Million"

REPRESENTATION OF CELL IN THE DATA CUBE

For example, in data cube the cell - "Birth information for the United States from 1969 to 2008", has been represented in the following way.

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://purl.org/linked-data/cube#Observation
http://eaisAnnotation/#value	"Birth information for the United States from 1969 to 2008"
http://purl.org/linked-data/cube#dataSet	http://example.org/ns/table1
http://example.org/ns/table1Row	3
http://example.org/ns/table1Column	2

TECHNOLOGY USED

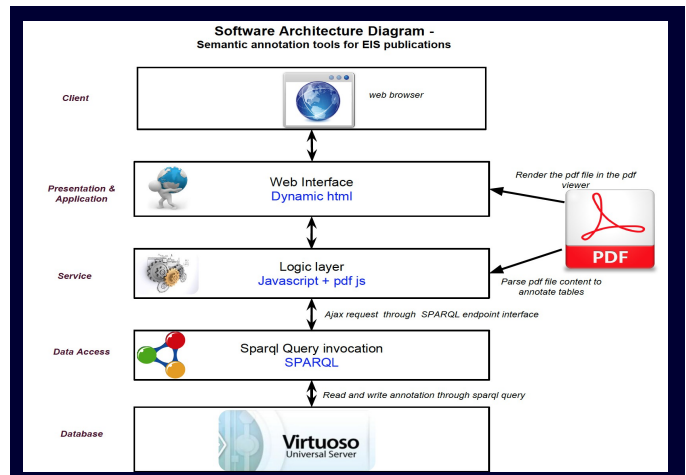
- Virtuoso Universal Server
- SPARQL as a query
- JavaScript + Html
- Pdf.js as rendering
- Qunit as JavaScript test environment
- DataCube (<http://www.w3.org/TR/vocab-data-cube/>)

PHYSICAL WORK STATION



Hostname: EIS-02

Description: Physical PC at Uni Bonn
 OS: Windows 7 Enterprise SP1 64bit
 RAM: 16 GB (14 GB Free)
 CPU: Intel Core i7-4770 @3.4GHz
 HDD: (C:) 240 GB & (L:) 930 GB
 IP Address: 131.220.153.88



SETUP / INSTALLATION PROCESS

- Install virtuoso universal server
- Download the code from git repo
- In the downloaded directory open the index.html in any browser.
- For details set up please follow the git hub link section in the bottom.

VIRTUOSO SETTING

- Make sure that virtuoso endpoint is listening at <http://localhost:8890/>
- Login as admin (user: dba, pass: ***) in the backend (<http://localhost:8890/conductor/>)
- Update write permission settings for user : sparql
- Enable CORS for the Data Access of JS client

VM SETTINGS

VM Name : Ubuntu-sciPubAnn
 OS: Ubuntu 12.04 LTS (32 bit)
 RAM: 3.9 GB
 CPU: Intel® Core™ i7-4770 CPUs @3.40GHz
 HDD: 4.2 GB (possible increase)

PROJECT'S GIT HUB LINKS

- Code link: <https://github.com/saifulnipo/eis-semantic-annotation>
- Documentation link : <https://github.com/saifulnipo/eis-semantic-annotation/wiki>
- Release note : <https://github.com/saifulnipo/eis-semantic-annotation/releases>