

JIST2018 Tutorial

Practical Use of a Knowledge Graph with Case Studies using Semantic Web Publishing Tools

2018.11.26

Fujitsu Laboratories LTD.
Kenji Kobayashi

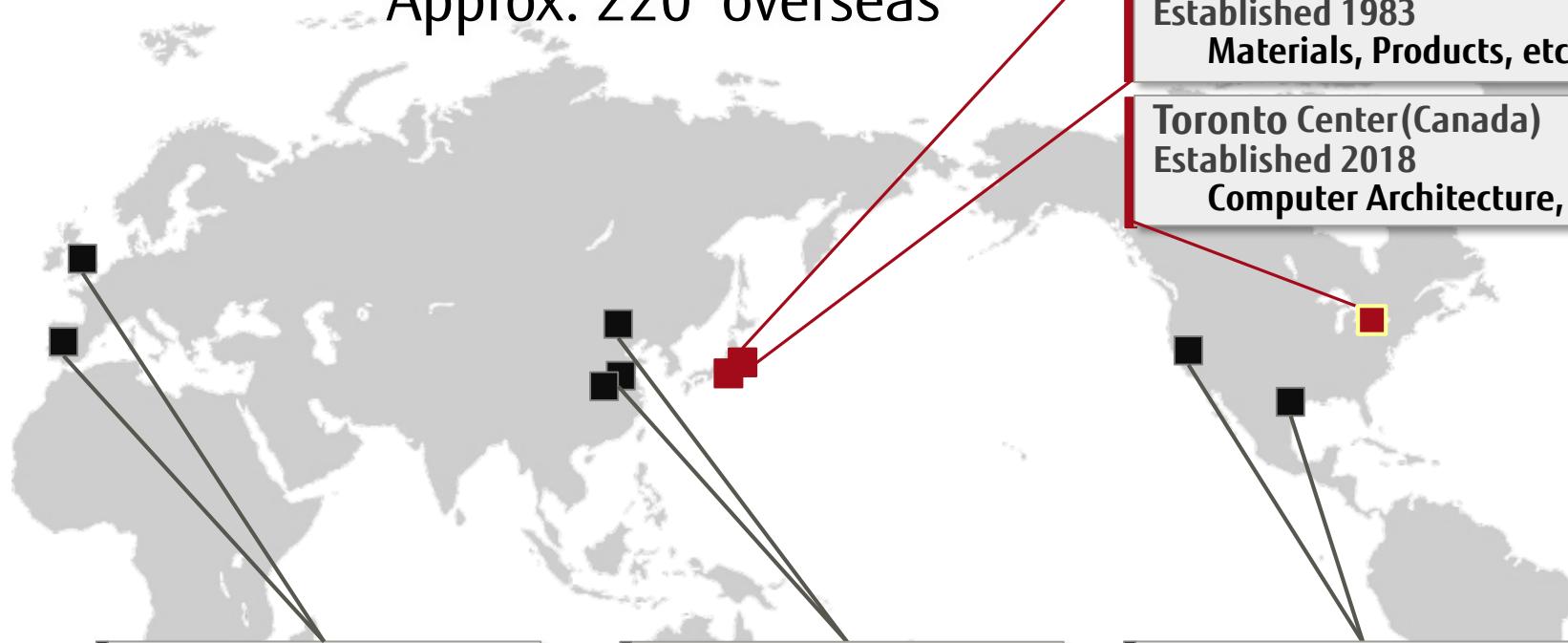
Introduction : Schedule

Morning	Introduction to Linked Data tools	9:00-10:00 (60 mins)	Presentation
		10:00-10:10 (10 mins)	Break
		10:10-10:50 (40 mins)	Presentation
		10:50-11:10 (20 mins)	Use Case Studies
Afternoon	Hands-on	13:00-13:10 (10 mins)	Introduction
		13:10-14:55 (95 mins)	Trial 1: Let's visualize corporate data Trial 2: Let's develop a financial analysis application
		14:55-15:00 (5 mins)	Closing

Introduction: Fujitsu Laboratories LTD.

FUJITSU

- CEO: Shigeru Sasaki
- R&D Budget: Approx. US\$ 280 million
- Employees: Approx. 1200 in Japan
Approx. 220 overseas



**Fujitsu Laboratories
of Europe Ltd. (Europe)**
Established 2001
Networking, Standardization, Security,
High Performance Computing

**Fujitsu R & D Center
Co., Ltd. (China)**
Established 1998
Image and Character Recognition
AI, Networking

**Fujitsu Laboratories Ltd.
Kawasaki Laboratory (Japan)**
Established 1968
Computer, AI, Cloud system,
Network, IoT, Software,
Security, User Interface, etc

Atsugi Laboratory (Japan)
Established 1983
Materials, Products, etc

Toronto Center(Canada)
Established 2018
Computer Architecture, AI, etc

**Fujitsu Laboratories
of America, Inc. (U.S.)**
Established 1993
Software, AI, Networking
High Performance Computing

Introduction : Tutors

FUJITSU



Kenji Kobayashi
(Presenter)



Hiroaki Morikawa
(Presenter)



Takanori Ugai



Fumihiro Nishino



Yusuke Koyanagi



Bin Piao



Kyomoto Matsushita



Seiji Okajima

- Department: AI Laboratory
- Research Focus: Semantic Web, Linked Data (LD), Open Data

Introduction: LOD Search Engine



FUJITSU

- Crawl LOD from around the world (several tens of billion data items), and provide a high-speed search service

Solution1

Use data without being aware of what site it is on



Use

Solution2

Applications need not to implement complex processes!

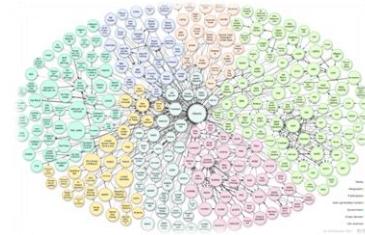


Application

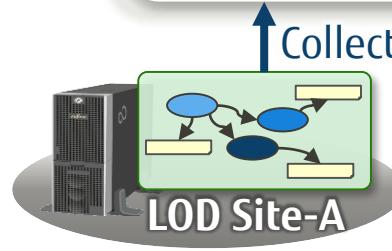
Search UI

Standard API

LOD4ALL

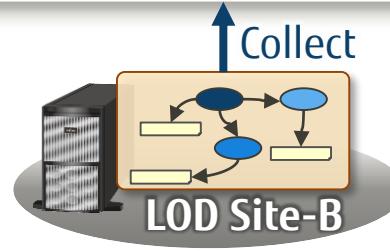


<http://lod4all.net/>



Collect

LOD Site-A



Collect

LOD Site-B



Collect

Solution3
Can even search sites without search functions!

Introduction : Our Research

■ Publications

- Manuel Peña Muñoz, Alejandro Llaves and Terunobu Kume. Populating the FLE Financial Knowledge Graph Proceedings of the 17th International Semantic Web Conference
- Hiroaki Morikawa,H., Nishino,F: Design and Implementation of a Standpoint-based Linked Data Visualization Approach n Proc. of the JIST 2018 Poster and Demonstrations Track. (2018).
- Villazon Terrazas, Boris & Garcia-Santa, Nuria & San Miguel, Beatriz & del Rey-Mejías, Angel & Muria Tarazon, Juan Carlos & Seara, Germán & Reneses, Blanca & de la Torre, Victor. (2018). Fujitsu HIKARI, a Healthcare Decision Support System based on Biomedical Knowledge. International Journal of Privacy and Health Information Management. 6. 26-49. 10.4018/IJPHIM.2018070103.
- Abe, S., Mitsuishi, Y., Tago, S., Igata, N., Okajima, S., Morikawa, H., Nishino, F.: Linked Corporations Data in Japan. In Proc. of the ISWC 2016 Poster and Demonstrations Track. (2016).
- Naseer, A., Kume, T., Izu, T., Igata, N.: LOD for All: Unlocking infinite opportunities. In: The Semantic Web Challenge 2014, The 13th International Semantic Web Conference (2014)
- Kobashi, H., Carvalho, N., Hu B., Saeki, T.: Cerise: an RDF store with adaptive data reallocation. In Proc. of the 13th Workshop on Adaptive and Reflective Middleware Article No. 1. (2014)
- Carlos Buil-Aranda, Aidan Hogan, Jurgen Umbrich, and Pierre-Yves Vandenbussche. SPARQL Web-Querying Infrastructure: Ready for Action? Proceedings of the 12th International Semantic Web Conference, (Oct. 2013). *Best Paper Award

Introduction : Our Research (cont.)



■ Publications (cont.)

- Shohei Yamane, Takanori Ugai, Conversion of Physical Quantity and its Application, Proceedings of the 17th International Semantic Web Conference Poster and Demo,
- Lu Fang, Qingliang Miao and Yao Meng. DBpedia Entity Type Inference Using Categories
- Bo Hu, Aisha Naseer, Eduarda M. Rodrigues, Pierre-Yves Vandenbussche, and Masatomo Goto (2013). Linked Data for Financial Reporting. Fourth International Workshop on Consuming Linked Data, collocated with the 12th International Semantic Web Conference, (Oct. 2013).
- Pierre-Yves Vandenbussche. Empower Flexibility via Big Data: Unleashing Constraints in Financial Analytics. Proceedings of the 26th XBRL conference, (Apr. 2013).
- Pierre-Yves Vandenbussche, Carlos Buil-Aranda, Aidan Hogan, and Jurgen Umbrich. Monitoring SPARQL Endpoint Status. Proceedings of the 12th International Semantic Web Conference Poster and Demo, (Oct. 2013).

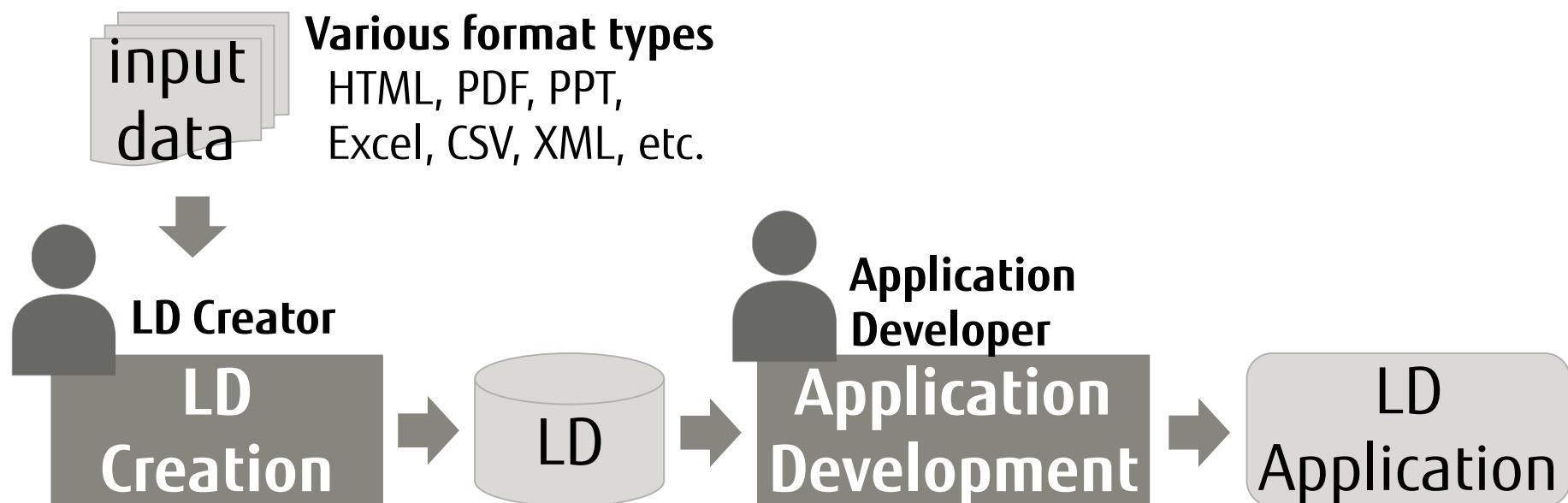
■ Award

- **Won Corporation Information Award** in RESAS Application Contest Organized by METI(Ministry of Economy, Trade and Industry)

Introduction : Contents

This tutorial covers useful tools*, from Linked Data creation to application development using Linked Data.

*used by us, and license free



Introduction : Sample Input Data

■ Accepted Paper page on the JIST2018 website

- HTML format
- Each paper has {id, author, title, category}

The screenshot shows the JIST2018 website homepage. At the top, there's a banner for the conference: "JIST2018: The 8th Joint International Semantic Technology Conference" held from Nov 26 - 28, 2018, in Awaji City, Hyogo, Japan. Below the banner is a navigation bar with links: Home, Calls, Important Dates, Program, Attending, Organization, Sponsorship, and Contact. A dropdown menu for "Program" is open, showing "Accepted Papers" as the selected option. The main content area is titled "Accepted Papers" and is further divided into "Research Track". It lists several papers, each with a title, authors, and a brief abstract. At the bottom right of this section, there's a logo for "OntoEconomy, LLC" with the text "Sponsors Gold Sponsors".

JIST2018: The 8th Joint International Semantic Technology Conference
Nov 26 - 28, 2018, Awaji City, Hyogo, Japan

Home Calls Important Dates Program Attending Organization Sponsorship Contact

Accepted Papers

Research Track

9. Elham Andaroodi and Frederic Andres.
Ontology-based Semantic Representation of Spatial Configuration: Shape Grammar for Caravanserai Architectural Heritage

DeFind: A Protege Plugin for Computing Concept Definitions in EL Ontologies

18. Ricardo Soares, Jeff Z. Pan, Elspeth Edelstein and Adam Wyner.
Knowledge Driven Intelligent Survey Systems for Linguists

19. Kemas Wihaja, Jeff Z. Pan, Martin Kollingbaum and Yu Deng.
More is Better: Sequential combinations of knowledge graph embedding approaches

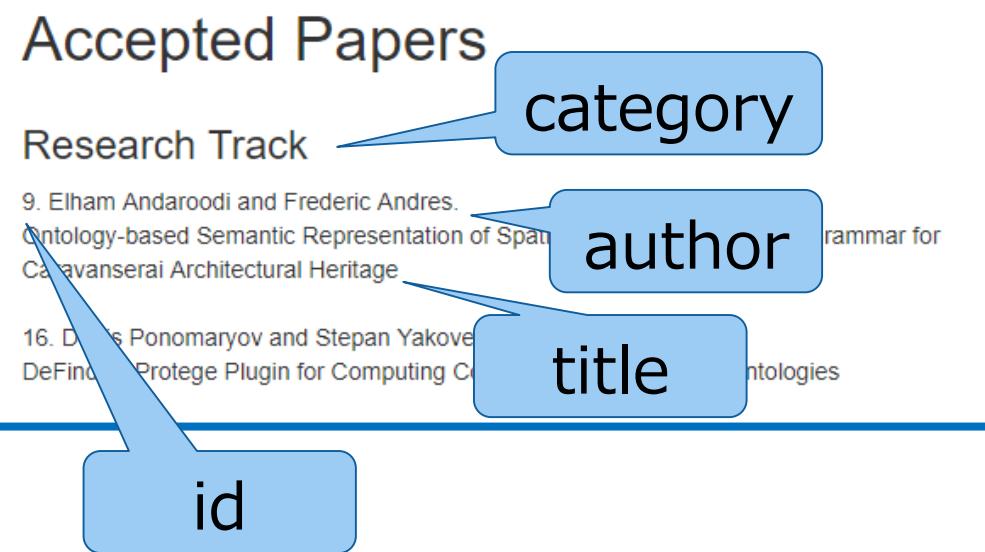
20. Daniel Mercier, Hyunmin Cheong and Chaitanya Tapaswi.
Unified Access to Data Sources using an Ontology

24. Rubén Francisco Manrique, Christian Grévisse, Olga Marino and Steffen Rothkugel.
Knowledge-Cross-based Cross Concept Identification in various Resources

Sponsors
Gold Sponsors

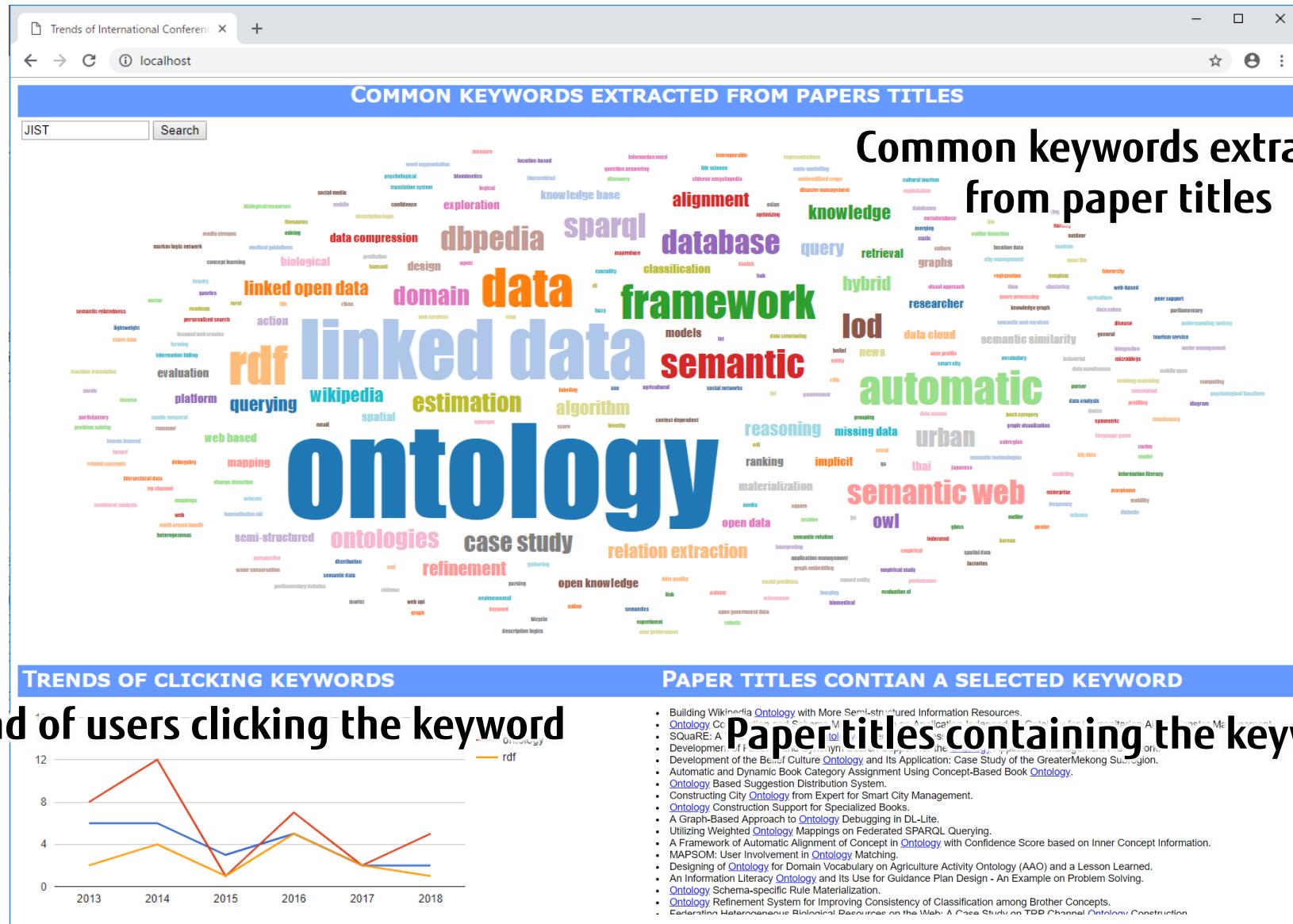
OntoEconomy, LLC

JIST2018 website



Introduction : Sample LD Application

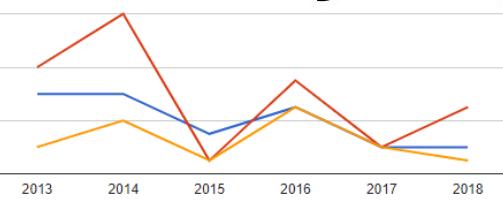
FUJITSU



Common keywords extracted from paper titles

TRENDS OF CLICKING KEYWORDS

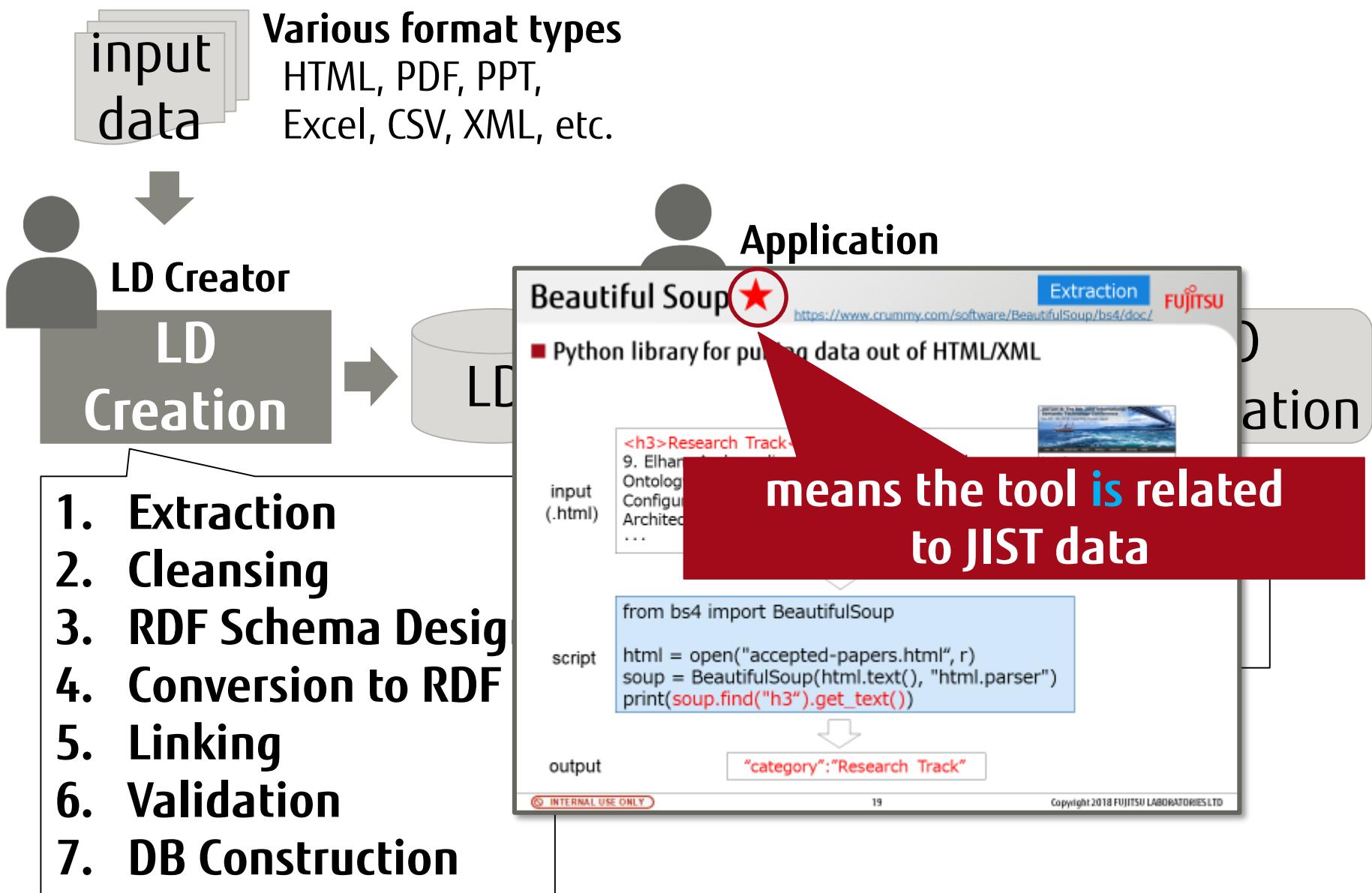
Trend of users clicking the keyword



PAPER TITLES CONTAIN A SELECTED KEYWORD

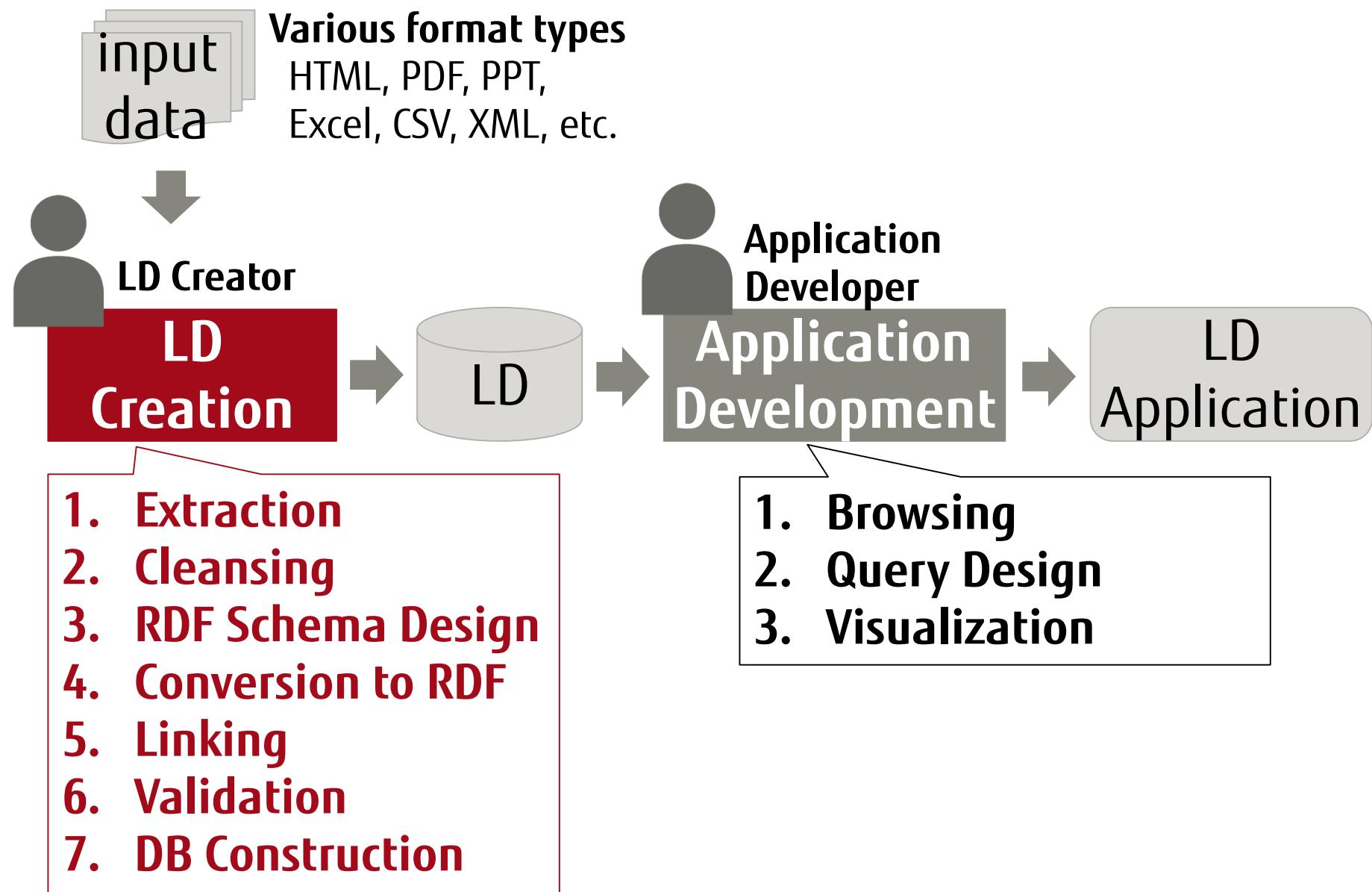
Outline

FUJITSU

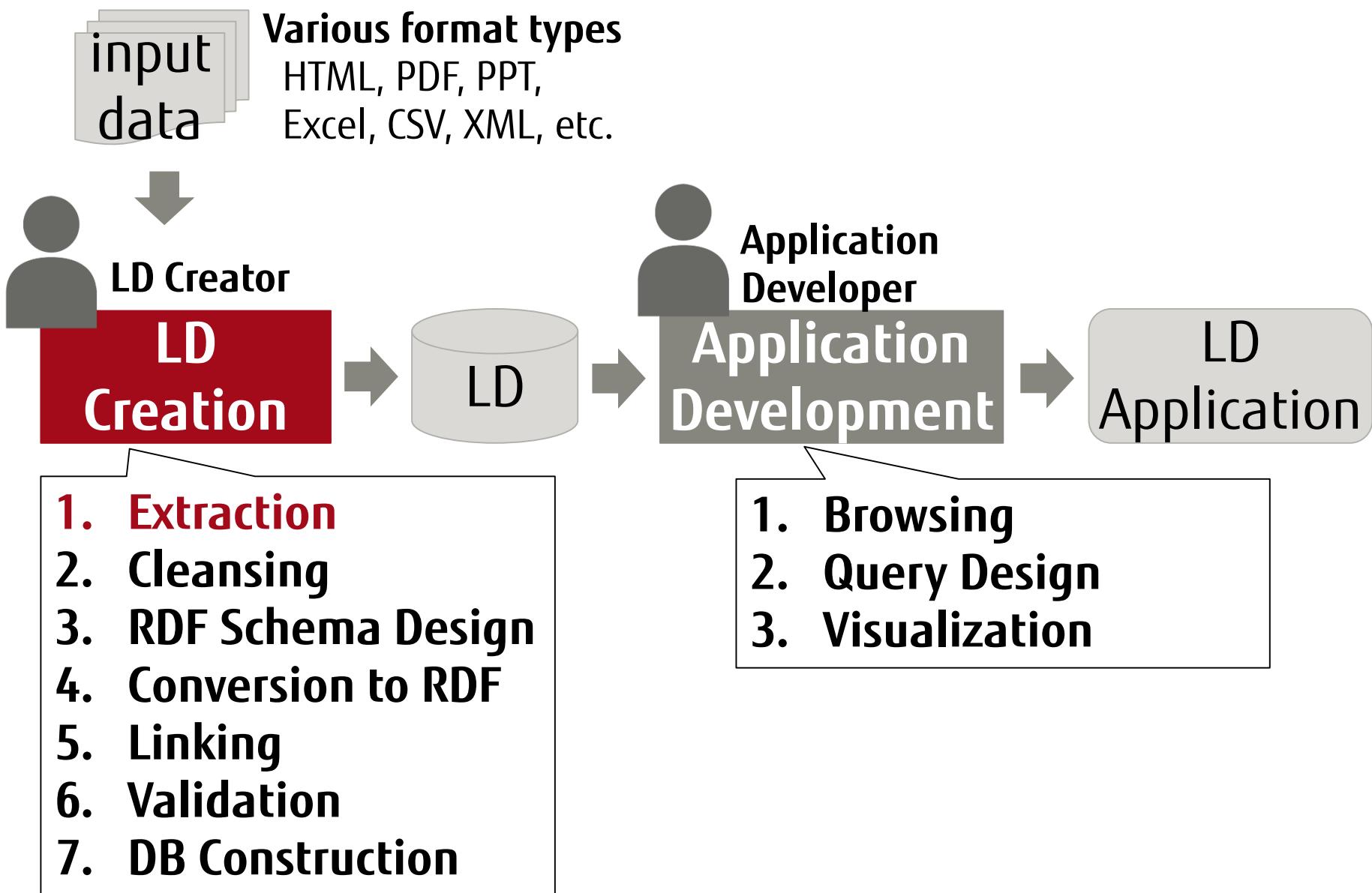


1. LD Creation

Outline

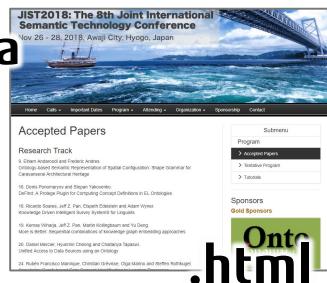


Outline



Extract useful data from various data formats

Ex. JIST data



Extraction



```
{"paperId": "id_x",  
 "author": "author_x",  
 "title": "title_x"}
```

.json

■ From HTML documents

- [Beautiful Soup](#)★

■ From PDF documents

- [pdfminer](#), [tabula](#), [pdfx](#)

■ From PowerPoint documents

- [python-pptx](#)

Beautiful Soup ★

Extraction(1/6)

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

FUJITSU

■ Python library for pulling data out of HTML/XML

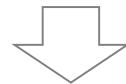
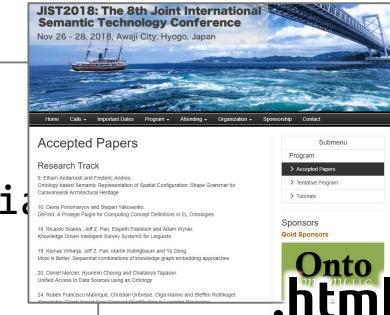
input
(.html)

<h3>Research Track</h3>

9. Elham Andaroodi and Frederic Andres.

Ontology-based Semantic Representation of Spatial
Configuration: Shape Grammar for Caravanserai
Architectural Heritage

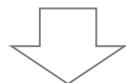
...



```
from bs4 import BeautifulSoup
```

script

```
html = open("accepted-papers.html", r)
soup = BeautifulSoup(html.text(), "html.parser")
print(soup.find("h3").get_text())
```



“category”：“Research Track”

output

- Command-line tool for extracting information from PDF documents

```
$ pdf2txt.py -t html p177-constantin.pdf
<html><head>
<meta http-equiv="Content-Type" content="text/html;
charset=utf-8">
</head><body>
<span style="position:absolute; border: gray 1px solid;
left:0px; top:50px; width:612px; height:792px;"></span>
<div style="position:absolute; top:50px;"><a
name="1">Page 1</a></div>
<div style="position:absolute; border: textbox 1px
solid; writing-mode:lr-tb; left:90px; top:116px;
width:429px; height:22px;">
...
...
```

■ Converts tables in PDF files to CSV

- It generates XML with coordinates as intermediate data, and recognizes words, cell borders, and tables
- WebUI

The 8th Joint International Semantic Technology Conference (JIST 2018)
November 26 to 28, 2018
in the Awaji Yumebutai International Conference Center, Awaji City, Hyogo,
Japan. Asian neighbors and from Europe and the United States
<http://jist2018.knowledge-graph.jp/>

The Joint International Semantic Technology Conference (JIST) is a regional federation of Semantic Technology related conferences. The mission of JIST is to bring together researchers in Semantic Technology research community and other areas of semantic related technologies to present their innovative research results or novel applications of semantic technologies.

JIST2018 got 75 submissions from 25 countries, and we expect around 100 participants from Asia and Western countries.
The conference proceeding will be published by Springer as a volume of the famous LNCs.

The organizers are pleased to announce three levels of sponsorship: Platinum, Gold, and Silver with several patterns of benefits.

Levels of sponsorship:

	Platinum	Gold	Silver
Sponsorship fee (USS)	USD 2,500	USD 1,500	USD 1,000
Sponsorship fee (Japanese Yen)	JPY 275,000	JPY 165,000	JPY 110,000
(1) Logo on Web and Proceedings	✓	✓	✓
(2) Ads on printed programs	✓	✓	✓
(3) Leaflet for all participants	✓	✓	✓
(4) Free registration	2	1	0
(5) Exhibition at a booth	✓	✓	
(6) Commercial poster and/or demo	✓	✓	

Automatic recognition tables in PDF

	Platinum	Gold	Silver
Sponsorship fee (USS)	USD 2,500	USD 1,500	USD 1,000
Sponsorship fee (Japanese Yen)	JPY 275,000	JPY 165,000	JPY 110,000
(1) Logo on Web and Proceedings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(2) Ads on printed programs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(3) Leaflet for all participants	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
(4) Free registration	2	1	0
(5) Exhibition at a booth	<input type="checkbox"/>	<input type="checkbox"/>	
(6) Commercial poster and/or demo	2	1	

Extraction result

■ Python wrapper of Tabula

- It recognizes tables in PDF automatically as well as Tabula (Default)
 - manual setting is also possible

In the case of setting a fixed area

Set a fixed area in PDF



Ex. Using the Mac
Preview function

http://docs.whirlpool.eu/_doc/Cookbook_Mwo_only_501912000448EN.pdf

Code

```
from tabula import read_pdf
df = read_pdf("recipes2.pdf", guess=False,
               area=(50.0, 200.0, 82.0, 548.0),
               pages='all')
```

set

Extraction result

Cheese and Mushroom Toast
Asparagus and Cheese Soup
Tomato Soup
Mushroom Soup
Mushroom and Saffron Risotto
Pork with Onions and Peppers
Jacket Potatoes
Vegetable Ratatouille
Tomato Sauce

■ Converts scholarly papers (in English) to HTML/XML

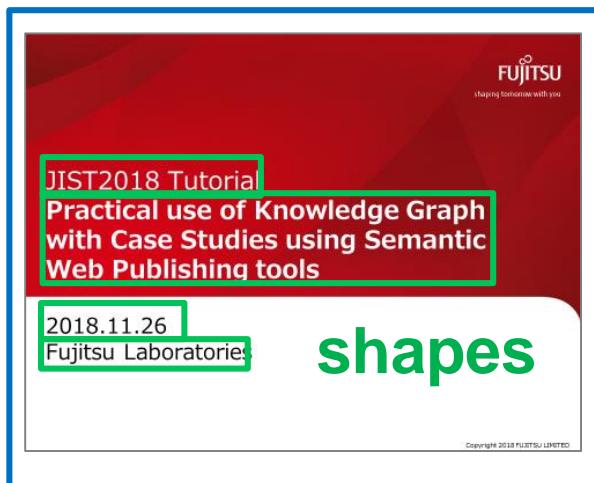
■ Web Service

PDF

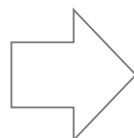


```
<!DOCTYPE html PUBLIC "-//W3C//DTD  
XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-  
transitional.dtd">  
<html  
xmlns="http://www.w3.org/1999/xhtml"><h  
ead><meta http-equiv="Content-Type"  
content="text/html; charset=UTF-8">  
      <title>pdfx -  
HTML for "The Identity Problem ..."</title>
```

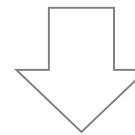
- Python library for extracting information from PowerPoint (.pptx) documents
 - It extracts not only text boxes, but also graphs, tables, and so on, with page numbers and coordinates



slide



slides[0].shape.text_frame.text
slides[0].shape.left
slides[0].shape.top

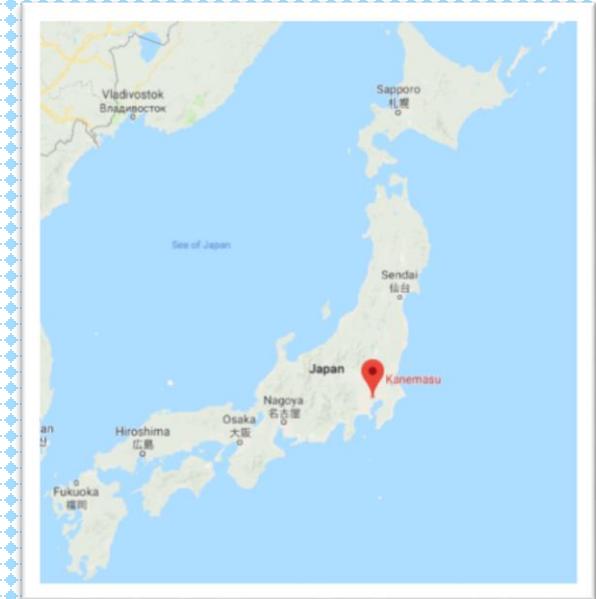


text	coordinates
JIST2018 Tutorial	100,2000
Practical use of Knowledge...	100,2350
2018.11.26	100,3970
Fujitsu Laboratories	100,4320

My Favorite Things

FUJITSU

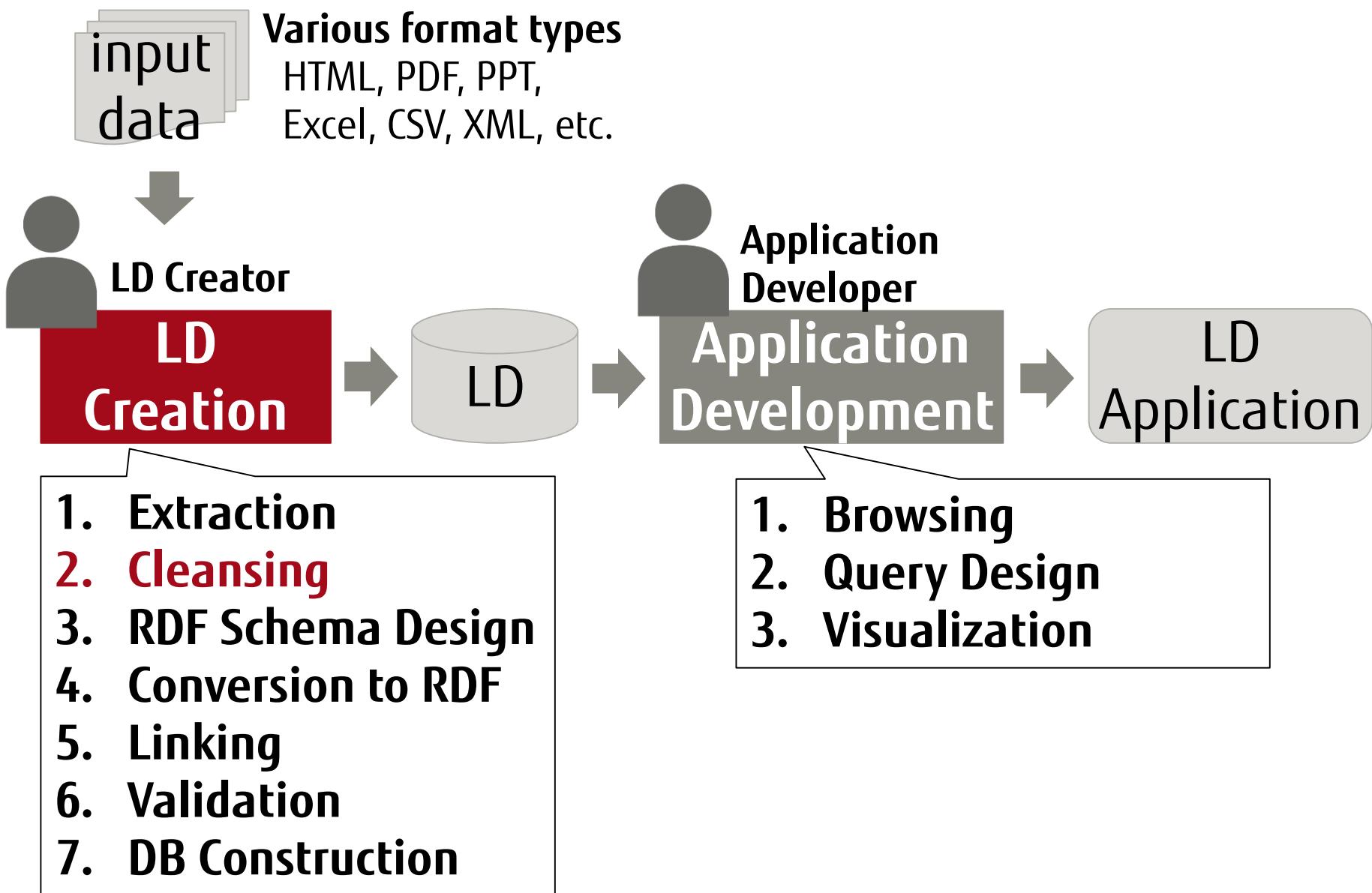
Sea Urchin and Beef Sashimi Roll



Kanemasu Bar / かねます
Tokyo, Japan

<https://tabelog.com/tokyo/A1313/A131302/13002243/>

Outline



Cleansing is format conversion to improve the quality and convenience of the extracted data

Ex. JIST data

```
{"paperId":"id_x",  
 "author":"author_x" ,  
 "title":"title_x"}
```

.json



paperId, author, title
id_x, author_x, title_x

.CSV

```
<paperId>id_x</paperId>  
<author>author_x</author>  
<title>title_x</title>
```

.XML

■ Cleansing for various data formats

■ OpenRefine

■ Cleansing for CSV, XML, JSON

107 rows

Show as: rows records Show: 5 10 25 50 rows

<input type="checkbox"/> All	<input type="checkbox"/> _ - id	<input type="checkbox"/> _ - title	<input type="checkbox"/> _ - category	<input type="checkbox"/> _ - authors - autl
1.	9	Ontology-based Semantic Representation of Spatial Configuration: Shape Grammar for Caravanserai Architectural Heritage	Research Track	Elham Andarodi
2.	9			Frederic Andres
3.	16	DeFind: A Protege Plugin for Computing Concept Definitions in EL Ontologies	Research Track	Denis Ponomaryov
4.	16			Stepan Yakovenko
5.	18	Knowledge Driven Intelligent Survey SystemS for Linguists	Research Track	Ricardo Soares
6.	18			Jeff Z. Pan
7.	18			Elspeth Edelstein
8.	18			Adam Wyner
9.	19	More is Better: Sequential combinations of knowledge graph embedding approaches	Research Track	Kemas Wiharja
10.	19			Jeff Z. Pan

Facet / Filter Undo / Redo 22 / 22

Refresh Reset All Remove All

_ - id change reset

_ - category change

3 choices Sort by: name count Cluster

- Research Track 20
- Special Session on Government
- Open Data 5
- Special Session on Semantic Web for Life Sciences 4
- (blank) 78
- Facet by choice counts

Facet

107 rows

Show as: rows records Show: 5 10 25 50 rows

<input type="checkbox"/> All	<input type="checkbox"/> _ - id	<input type="checkbox"/> _ - title
1.	Facet	ed Semantic Representation of Spatial Configuration: Shape Grammar for Caravanserai Arc
2.	Text filter	
3.	Edit cells	Transform...
4.	Edit column	Common transforms
5.	Transpose	Trim leading and trailing whitespace
6.		Collapse consecutive whitespace
7.	Sort...	Fill down
8.		Blank down
9.	View	Unescape HTML entities
10.	Reconcile	To titlecase
		To uppercase
		To lowercase
		To number
		To date
		To text
		To null
		To empty string

Edit cells

■ Cleansing for CSV, XML, JSON

107 rows

Show as: rows records

Facet / Filter Undo / Redo 22 / 22

All _ - id Refresh Reset All Remove All

_ - id change reset

9.00 — 92.00

_ - category change

3 choices Sort by: name count Cluster

- Research Track 20
- Special Session on Government
- Open Data 5
- Special Session on Semantic Web for Life Sciences 4
- (blank) 78
- Facet by choice counts

Facet

	<input type="checkbox"/> _ - category	<input type="checkbox"/> _ - authors - autl
anserai Architectural Heritage	Research Track	Elham Andarodi
		Frederic Andres
	Research Track	Denis Ponomaryov
		Stepan Yakovenko
	Research Track	Ricardo Soares
		Jeff Z. Pan
		Elspeth Edelstein
		Adam Wyner
	Research Track	Kemas Wiharja
		Jeff Z. Pan

107 rows

Show as: rows records Show: 5 10 25 50 rows

All _ - id _ - title

1.	Facet	→ ed Semantic Representation of Spatial Configuration: Shape Grammar for Caravanserai Arc
2.	Text filter	→
3.	Edit cells	→ Transform...
4.	Edit column	→ Common transforms → Trim leading and trailing whitespace
5.	Transpose	→ Collapse consecutive whitespace
6.	Sort...	→ Fill down
7.	View	→ Blank down
8.		→ Unescape HTML entities
9.		→ To titlecase
10.	Reconcile	→ To uppercase
		→ To lowercase
		→ To number
		→ To date
		→ To text
		→ To null
		→ To empty string

Edit cells

■ Cleansing for CSV, XML, JSON

The screenshot shows the OpenRefine interface with two main panels. The left panel, titled 'Facet', displays 107 rows of data with columns '_id' and '_title'. The right panel, titled 'Edit cells', shows the same 107 rows and a detailed view of the '_title' column's data. A context menu is open over the first few rows, specifically over the entry 'Edit cells'. This menu lists various transformation options:

- Transform...
- Common transforms
 - Trim leading and trailing whitespace
 - Collapse consecutive whitespace
 - Unescape HTML entities
 - To titlecase
 - To uppercase
 - To lowercase
 - To number
 - To date
 - To text
- Fill down
- Blank down
- Split multi-valued cells...
- Join multi-valued cells...
- Cluster and edit...

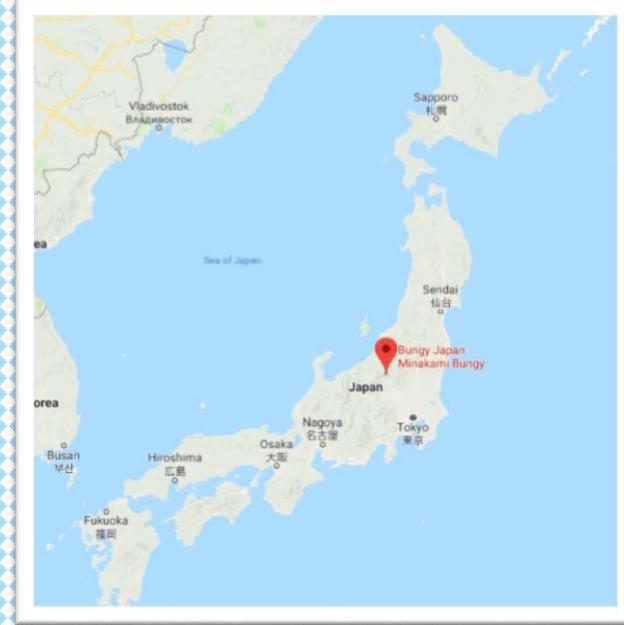
The 'Edit cells' menu is highlighted with a blue border.

Facet **Edit cells**

My Favorite Things

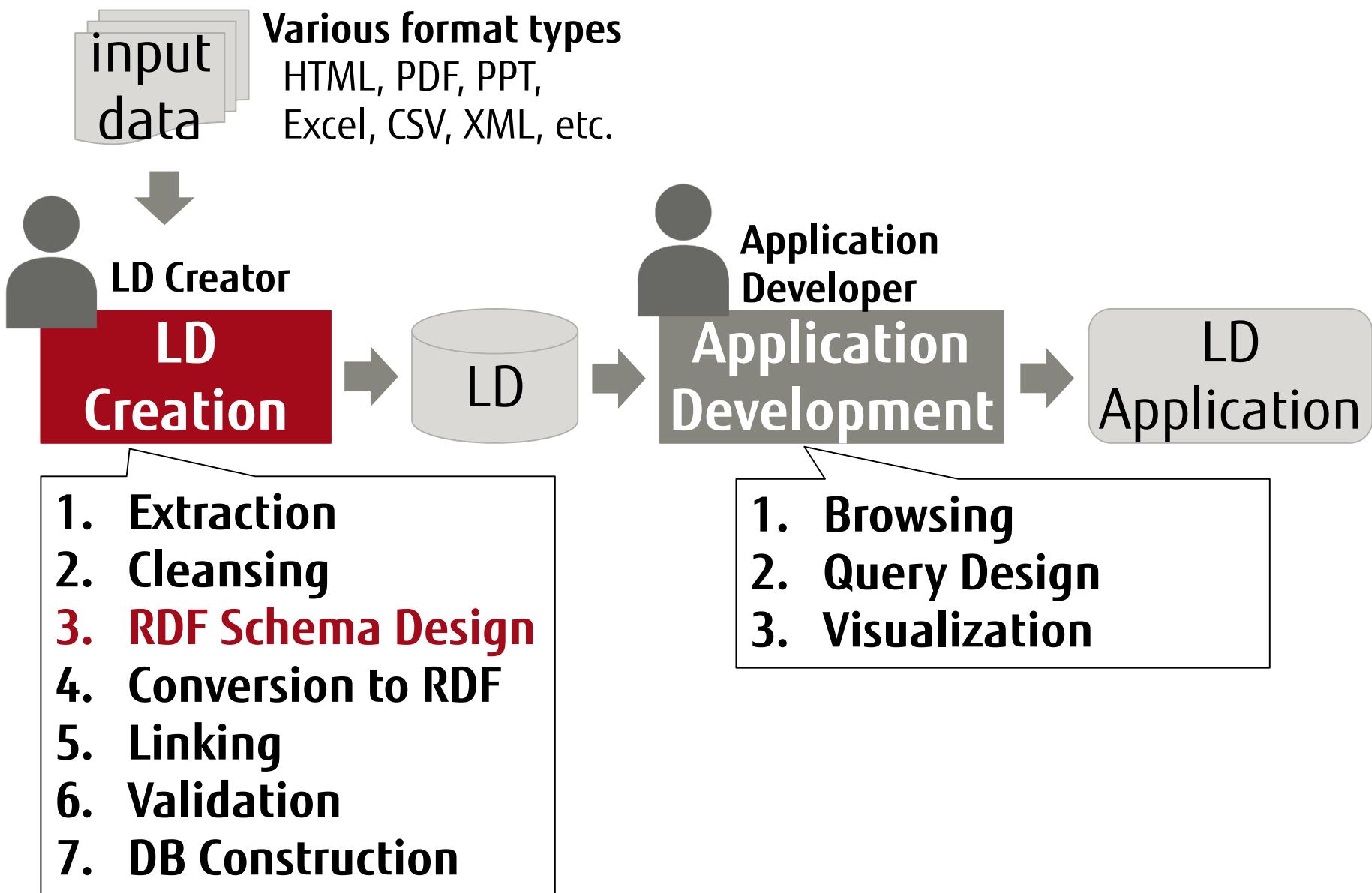
FUJITSU

Bungee Jump



Minakami Camping Area, Japan
/ みなかみキャンプ場

Outline



RDF Schema Design

FUJITSU

Design the schema for RDF conversion of the extracted / cleansed data, and construct an ontology as necessary

Ex. JIST data



■ Searching

- Vocabulary / Ontology: [Linked Open Vocabularies](#)★
- Dataset: [Google Dataset Search](#)★
- Instance: [LOD Laundromat](#), [LOD4ALL](#)

■ Browsing

- [Data Sniffer](#)★

■ Building

- [Protégé](#)

- A service to search for existing vocabulary
 - 651 vocabularies are included (*October 2018)

Example: "research article title"



The screenshot shows the search results for the query "research article title". The results are presented in a card-based interface.

TERMS → research article title

1869 results

dcterms:title (dcterms)
13,420,023 occurrences in 132 LOD datasets
<http://purl.org/dc/terms/title>
rdfs:label Title @en
localName title

dce:title (dce)
7,626,720 occurrences in 107 LOD datasets
<http://purl.org/dc/elements/1.1/title>
rdfs:label Title @en
localName title

dbpedia-owl:Article (dbpedia-owl)
n/a (use in LOD)
<http://dbpedia.org/ontology/Article>
rdfs:label article @en
rdfs:label article @fr
localName Article

gov:hasTitle (gov)
n/a (use in LOD)

Type

- vocabulary >
- property/class
- property (1065)
- class (804)

Tag

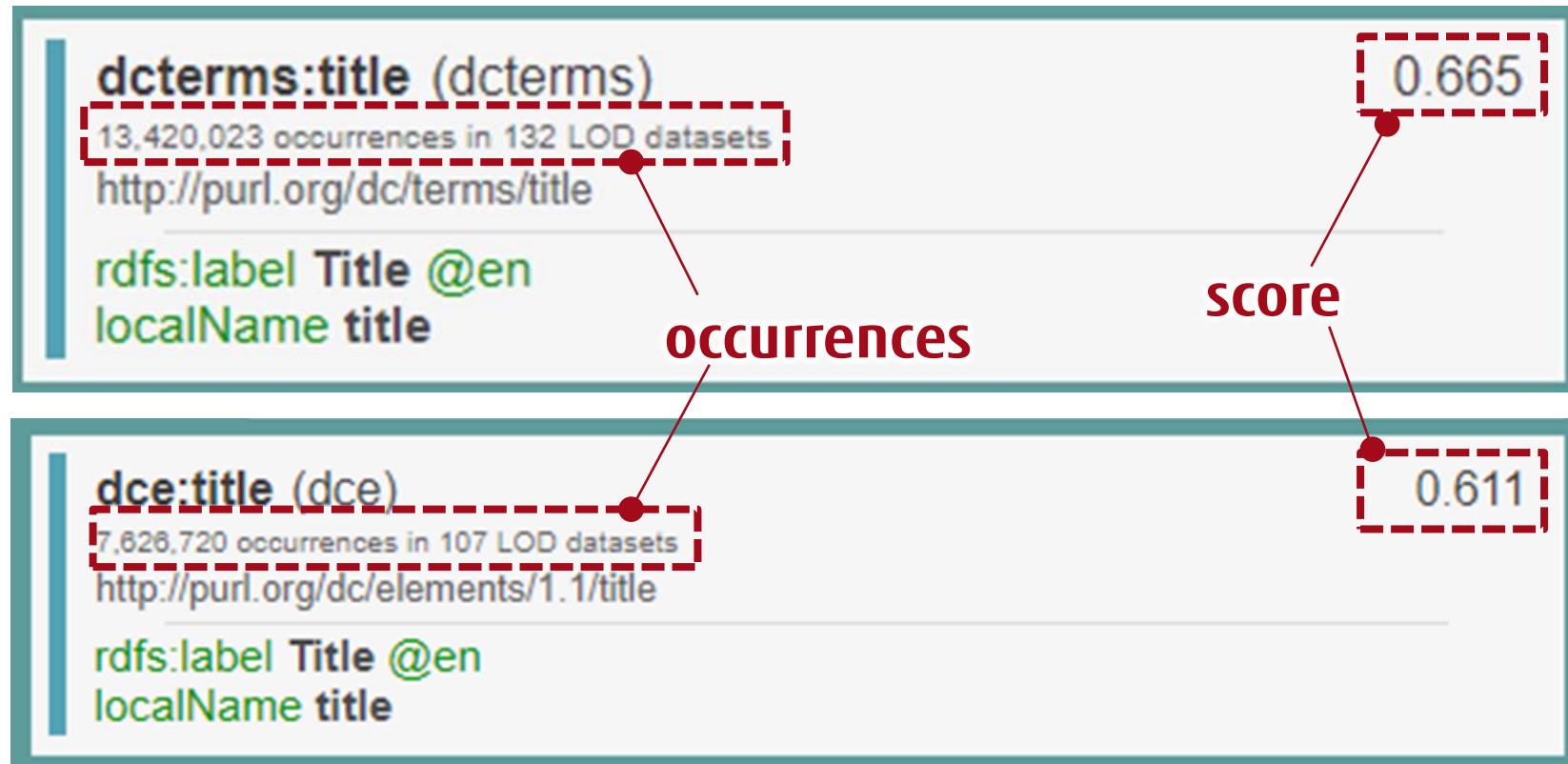
- Academy (643)
- General & Upper (167)
- Metadata (162)

search results

property

class

A red bracket on the right side groups the "Type" and "Tag" sections under the heading "search results". A blue dashed arrow points from the "property" label to the "property" section in the sidebar. An orange dashed arrow points from the "class" label to the "class" section in the sidebar.

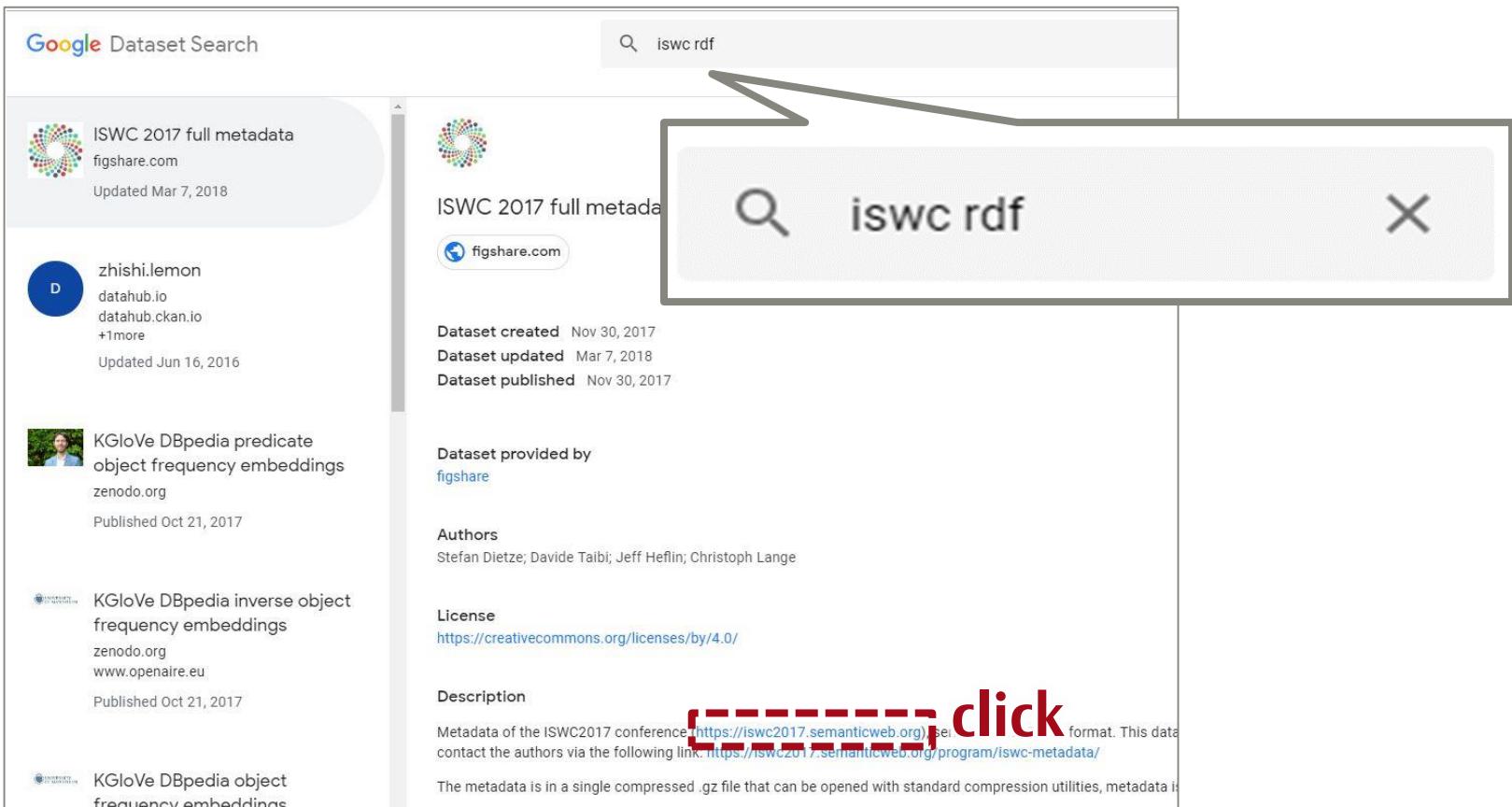


dcterms:title / dce:title are most widely used

*`dcterms:title` is a sub property of `dce:title`.

- Dataset search service by Google (launched September, 2018)
 - Enables searching of datasets published on the web.

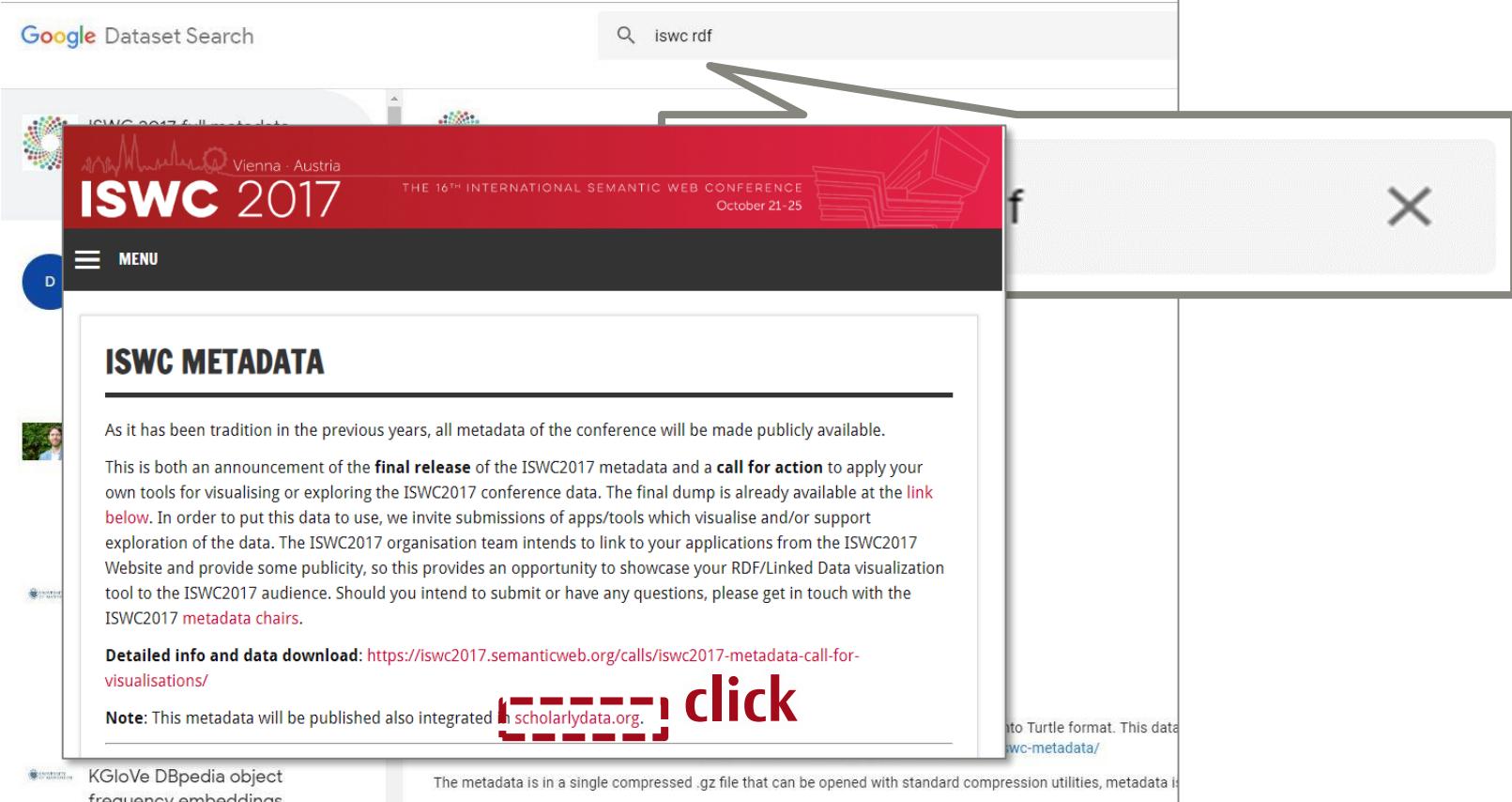
Example: ISWC RDF data



The screenshot shows the Google Dataset Search interface. A search bar at the top contains the query "iswc rdf". Below the search bar, a list of datasets is displayed. The first dataset in the list is "ISWC 2017 full metadata" from figshare.com, which is highlighted with a large red rectangular callout box. This dataset was created on Nov 30, 2017, updated on Mar 7, 2018, and published on Nov 30, 2017. It is provided by figshare and has authors Stefan Dietze, Davide Taibi, Jeff Heflin, and Christoph Lange. The license is CC-BY 4.0. The description notes that it is metadata in RDF format, available at https://iswc2017.semanticweb.org/repositories/iswc2017_rdf.ttl. The dataset is also associated with the ISWC 2017 conference and is available on zenodo.org. Other datasets listed include KGloVe DBpedia predicate object frequency embeddings, KGloVe DBpedia inverse object frequency embeddings, and KGloVe DBpedia object frequency embeddings.

- Dataset search service by Google (launched September, 2018)
 - Enables searching of datasets published on the web.

Example: ISWC RDF data



The screenshot shows the Google Dataset Search interface. In the search bar at the top, the query "iswc rdf" is entered. Below the search bar, a red banner for the "ISWC 2017" conference in Vienna, Austria, is displayed. The banner includes the text "THE 16TH INTERNATIONAL SEMANTIC WEB CONFERENCE October 21-25". Below the banner, the search results are shown under the heading "ISWC METADATA". The results page contains text about the metadata release, a "click" button, and links for detailed info and data download. A note mentions integration with scholarlydata.org. At the bottom, there is information about KGloVe DBpedia object frequency embeddings.

Google Dataset Search

iswc rdf

ISWC 2017 Full Conference

Vienna - Austria

ISWC 2017

THE 16TH INTERNATIONAL SEMANTIC WEB CONFERENCE

October 21-25

MENU

D

ISWC METADATA

As it has been tradition in the previous years, all metadata of the conference will be made publicly available.

This is both an announcement of the **final release** of the ISWC2017 metadata and a **call for action** to apply your own tools for visualising or exploring the ISWC2017 conference data. The final dump is already available at the [link below](#). In order to put this data to use, we invite submissions of apps/tools which visualise and/or support exploration of the data. The ISWC2017 organisation team intends to link to your applications from the ISWC2017 Website and provide some publicity, so this provides an opportunity to showcase your RDF/Linked Data visualization tool to the ISWC2017 audience. Should you intend to submit or have any questions, please get in touch with the ISWC2017 [metadata chairs](#).

Detailed info and data download: <https://iswc2017.semanticweb.org/calls/iswc2017-metadata-call-for-visualisations/>

Note: This metadata will be published also integrated in scholarlydata.org.

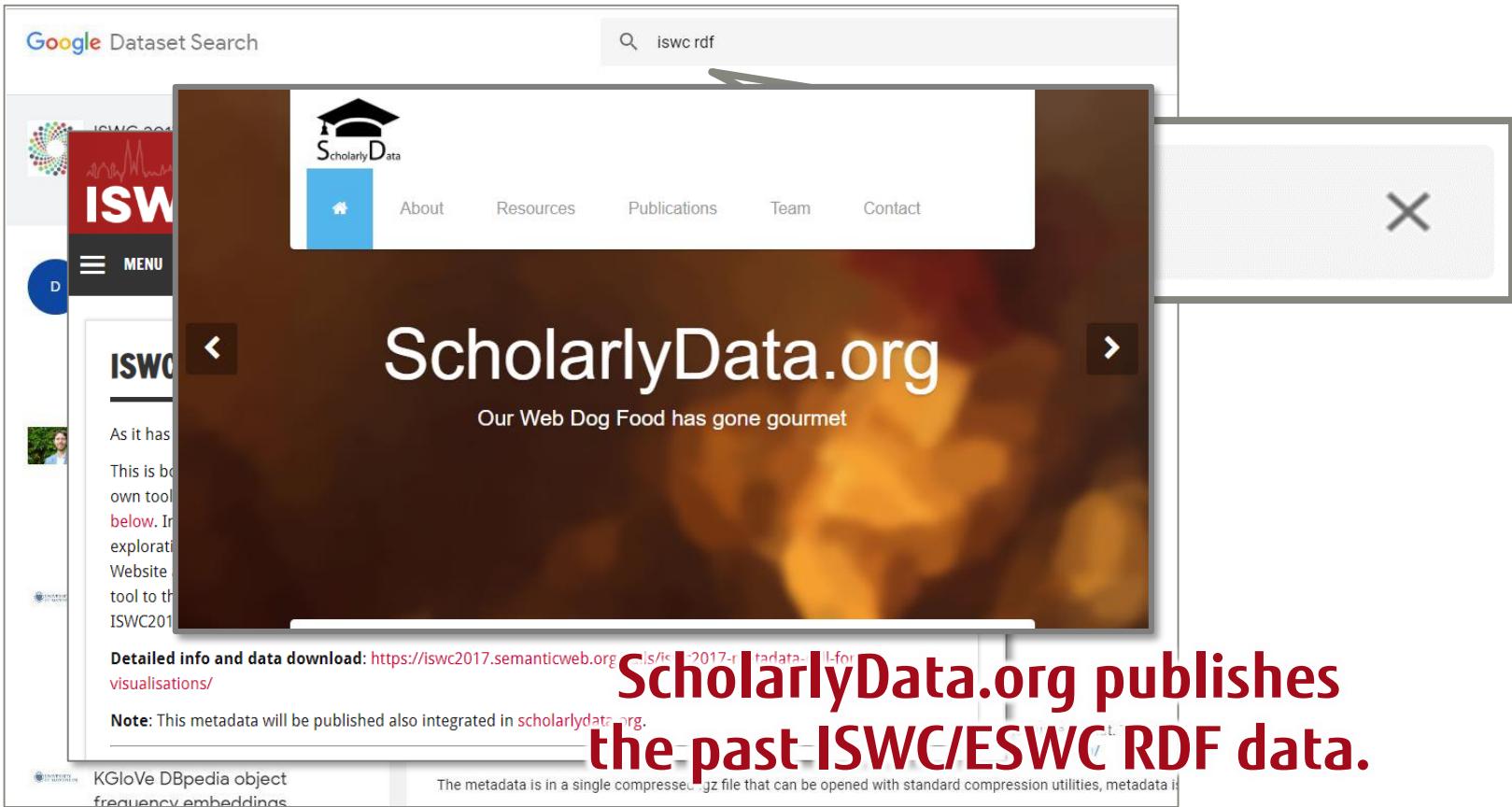
 click

KGloVe DBpedia object frequency embeddings

The metadata is in a single compressed .gz file that can be opened with standard compression utilities, metadata is in Turtle format. This data is available at <https://iswc2017.semanticweb.org/calls/iswc2017-metadata/>

- Dataset search service by Google (launched September, 2018)
 - Enables searching of datasets published on the web.

Example: ISWC RDF data



The screenshot shows the Google Dataset Search interface. In the search bar, the query "iswc rdf" is entered. Below the search bar, a card for the "ScholarlyData.org" dataset is displayed. The card features the "ScholarlyData" logo (a graduation cap icon) and the text "About", "Resources", "Publications", "Team", and "Contact". The main content area has a dark background with the text "ScholarlyData.org" in large white letters and "Our Web Dog Food has gone gourmet" below it. A navigation bar with arrows is visible on the left side of the main content area. At the bottom of the card, there is a note: "Detailed info and data download: <https://iswc2017.semanticweb.org/isic-2017-1.tardata-1-for-visualisations/>". Another note says: "Note: This metadata will be published also integrated in scholarlydata.org". The footer of the card includes links to "KGlove DBpedia object frequency embeddings" and a note about compressed metadata files.

ScholarlyData.org publishes the past ISWC/ESWC RDF data.

- Crawling from the LOD Cloud and converting it to beautiful data
 - Removes duplicates, syntax errors, and empty nodes
- SPARQL endpoint and entity search function (LOTUS) also provided

The screenshot shows a search interface for 'Hideaki Takeda'. At the top left is a blue circular logo with a white lotus flower icon. To its right is a search bar containing the name 'Hideaki Takeda' with a magnifying glass icon to its right. Further to the right is a grey button labeled 'LOD Search'. Below the search bar, the text 'About 86 results (0.19 seconds)' is displayed. The results are listed in a vertical stack, each entry consisting of the name 'Hideaki Takeda', a blue link to the URI resolver, and two blue links for 'SOURCE DATA' and 'METADATA'.

Hideaki Takeda
http://tw.rpi.edu/wiki/Special:URIResolver/Hideaki_Takeda
[SOURCE DATA](#) | [METADATA](#)

Hideaki Takeda
http://wiki.ontoworld.org/index.php/_Hideaki_Takeda
[SOURCE DATA](#) | [METADATA](#)

Hideaki Takeda
<http://data.semanticweb.org/person/hideaki-takeda>
[SOURCE DATA](#) | [METADATA](#)

Hideaki Takeda

- A service delivering a single-stop entry point for LOD utilization
 - Browsing, searching, and accessing capability for publicly available Linked Data.
- LOD4ALL also provides a development platform for applications using LOD

Instance Search

Search for instances

You can search an instance in the LOD4ALL. You can use two methods as follows.

- Keyword: Search triples from the keyword include the object value. Target predicates are "http://www.w3.org/2000/01/rdf-schema#label", "http://purl.org/dc/elements/1.1/title", "http://purl.org/dc/terms/title", "http://xmlns.com/foaf/0.1/name", "http://www.w3.org/2004/02/skos/core#prefLabel", "http://www.w3.org/2004/02/skos/core#altLabel", "http://schema.org/name" and "http://schema.org/alternateName".
- URI: Search triples from URI(subject).

Keyword URI

[Download](#)

Count:67

If you click a header, you can view the instance details.

Q2330082	>
<ul style="list-style-type: none">■ g:http://wikidata.dbpedia.org/■ s:http://wikidata.dbpedia.org/resource/Q2330082■ p:http://www.w3.org/2000/01/rdf-schema#label■ o:Hideaki Takeda	>

Q2330082	>
<ul style="list-style-type: none">■ g:http://wikidata.dbpedia.org/■ s:http://wikidata.dbpedia.org/resource/Q2330082■ p:http://www.w3.org/2000/01/rdf-schema#label■ o:Hideaki Takeda	>

Q2330082	>
<ul style="list-style-type: none">■ g:http://wikidata.dbpedia.org/■ s:http://wikidata.dbpedia.org/resource/Q2330082■ p:http://www.w3.org/2000/01/rdf-schema#label■ o:Hideaki Takeda	>

■ Enable to confirm RDF and metadata in HTML

■ Chrome extension

DBLP website

[–] Publication search results

found 36 matches

2017

Franz Baader, Daniel Borchmann, Adrian Nuradiansyah:
Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding.
 g. JIST 2017: 102-117

Turtle

Statement Collection #1	
Entity	https://dblp.uni-trier.de/rec/conf/jist/BaaderBN17
Attributes	https://dblp.org/rec/conf/jist/BaaderBN17

Statement Collection #2	
Entity	https://dblp.org/rec/conf/jist/BaaderBN17
Attributes	https://dblp.uni-trier.de/rdf/schema-2017-04-18#Publication https://doi.org/10.1007/978-3-319-70682-5_7
	The Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding. http://data.bibbase.org/ontology/#Inproceedings https://dblp.uni-trier.de/rdf/schema-2017-04-18#Inproceedings https://dblp.org/pers/b/Baader:Franz https://dblp.org/pers/b/Borchmann:Daniel https://dblp.org/pers/n/Nuradiansyah:Adrian https://doi.org/10.1007/978-3-319-70682-5_7 https://dblp.org/db/conf/jist/jist2017
Entity	JIST 102-117 2017

Statement Collection #3	
Entity	https://dblp.org/rec/conf/jist/BaaderBN17.nt

RDF N-Triples

■ Chrome extension developed by OpenLink

- Enable to confirm RDF and metadata in HTML

DBLP website

[-] Publication search results

found 36 matches

2017

Franz Baader, Daniel Borchmann, Adrian Nuradiansyah:
[Identity Problem in Description Logic Ontologies and Its Application to View-Based Information](#)
g. JIST 2017: 102-117

RDF N-Triples

Turtle

Statement Collection #1
Entity <https://dblp.uni-trier.de/rec/conf/jist/BaaderBN17>
Attributes <https://dblp.org/rec/conf/jist/BaaderBN17>
<https://dblp.org/rec/conf/jist/BaaderBN17>

Statements of one publication of JIST 2017:

Entity	https://dblp.org/rec/conf/jist/BaaderBN17
Attributes	

DBLP's own vocabulary

https://dblp.uni-trier.de/rdf/schema-2017-04-18#title	https://dblp.uni-trier.de/rdf/schema-2017-04-18#Publication https://doi.org/10.1007/978-3-319-70682-5_7
https://dblp.uni-trier.de/rdf/schema-2017-04-18#authoredBy	The Identity Problem in Description Logic Ontologies and Its A
https://dblp.uni-trier.de/rdf/schema-2017-04-18#authoredBy	https://dblp.org/pers/b/Baader:Franz
https://dblp.uni-trier.de/rdf/schema-2017-04-18#authoredBy	https://dblp.org/pers/b/Borchmann:Daniel
https://dblp.uni-trier.de/rdf/schema-2017-04-18#authoredBy	https://dblp.org/pers/n/Nuradiansyah:Adrian

Entity <https://dblp.org/rec/conf/jist/BaaderBN17.nt>

■ An ontology editor developed by Stanford University

Example: "has first item" property

The screenshot shows the Protégé interface with the following details:

- Toolbar:** Standard window controls (Minimize, Maximize, Close).
- Title Bar:** Conference-ontology.owl : [C:\Users\koyanagi.yusu...]
- Tab Bar:** Active Ontology (highlighted), Entities, Individuals by class, DL Query.
- Object Property Hierarchy:** A tree view under the 'Object properties' tab showing 'has first item' as a child of 'has item', which is itself a child of 'owl:topObjectProperty'.
- Annotations Tab:** Shows annotations for 'has first item':
 - rdfs:label: [language: en] has first item
 - rdfs:comment: [language: en] has first item
- Description Tab:** Shows properties of 'has first item':
 - Functional (checkbox checked)
 - Inverse function (checkbox uncheckable)
 - Transitive (checkbox uncheckable)
 - Symmetric (checkbox uncheckable)
 - Asymmetric (checkbox uncheckable)
 - Reflexive (checkbox uncheckable)
 - Irreflexive (checkbox uncheckable)
- Relationships:** Equivalent To (linked to 'has item'), SubProperty Of (linked to 'has item'), Inverse Of (linked to 'is first item of'), Domains (intersection) (List), Ranges (intersection) (ListItem).
- Bottom Status Bar:** No Reasoner set. Select a reasoner from the Reasoner menu. Show Inferences.

■ An ontology editor developed by Stanford University

Example: "has first item" property

The screenshot shows the Protégé interface with the following details:

- Title Bar:** conference-ontology (https://w3id.org/scholarlydata/ontology/conference-ontology.owl) : [C:\Users\koyanagi.yusu...]
- Toolbar:** Standard window controls.
- Navigation:** Back, Forward, Home, Search...
- Tab Bar:** Active Ontology, Entities, Individuals by class, DL Query.
- Property Hierarchy:** Object property hierarchy: 'has first item'. A tree view shows:
 - owl:topObjectProperty
 - 'during'
 - 'has address'
 - 'has affiliation'
 - 'has author list'
 - 'has content'
 - 'has item'
 - 'has first item'
 - 'has last item'
 - 'has location'
 - 'has part'
 - 'has sub event'
 - 'is series of'
 - 'has site'
 - 'has topic'
 - 'hasProceedings'
 - 'holds role'
 - 'in affiliation during'
 - 'is address of'
 - 'is affiliation of'
 - 'is document related'
 - 'is held by'
 - 'is item of'
 - 'is location for'
 - 'is part of'
 - 'hasSeries'
 - 'is sub event of'
 - 'is site of'
 - 'is topic of'
 - 'isProceedingsOf'
- Annotations Tab:** Shows annotations for 'has first item'.
 - rdfs:label [language: en] has first item
 - rdfs:comment [language: en] has first item
- Annotations Panel:** A detailed view of the 'has first item' annotation, showing:
 - owl:topObjectProperty
 - 'during'
 - 'has address'
 - 'has affiliation'
 - 'has author list'
 - 'has content'
 - 'has item'
 - 'has first item'
 - 'has last item'
 - 'has location'
 - 'has part'
- Status Bar:** No Reasoner set. Select a reasoner from the Reasoner menu. Show Inferences.

■ An ontology editor developed by Stanford University

Example: "has first item" property

The screenshot shows the Protégé interface with the following details:

- Object property hierarchy:** A tree view on the left shows 'has first item' under 'owl:topObjectProperty'.
- Annotations:** A modal window titled 'Annotations: 'has first item'' displays two annotations:
 - rdfs:label [language: en] has first item
 - rdfs:comment [language: en] has first item
- Annotations sidebar:** On the right, a sidebar titled 'Annotations' lists the same two annotations.

■ An ontology editor developed by Stanford University

Example: "has first item"

The screenshot shows the Protégé interface with the following components:

- Annotations Dialog:** A modal window titled "Annotations: 'has first item'" containing:
 - An "Annotations" section with a plus sign icon.
 - A list of annotations:
 - rdfs:label [language: en] has first item
 - rdfs:comment [language: en] has first item
- Object Property Hierarchy:** A tree view on the left showing the hierarchy of owl:topObjectProperty, including properties like 'has address', 'has item', and 'has first item'.
- Characteristics Dialog:** A modal window titled "Characteristics" containing a list of property types:
 - Functional
 - Inverse function
 - Transitive
 - Symmetric
 - Asymmetric
 - Reflexive
 - Irreflexive

■ An ontology editor developed by Stanford University

Example: "has first item"

Characteristics:

- Functional
- Inverse function
- Transitive
- Symmetric
- Asymmetric
- Reflexive
- Irreflexive

Description: 'has first item'

Equivalent To +

SubProperty Of +

- has item

Inverse Of +

- is first item of

Domains (intersection) +

- List

Ranges (intersection) +

- ListItem

No Reasoner set. Select a reasoner from the Reasoner menu Show Inferences

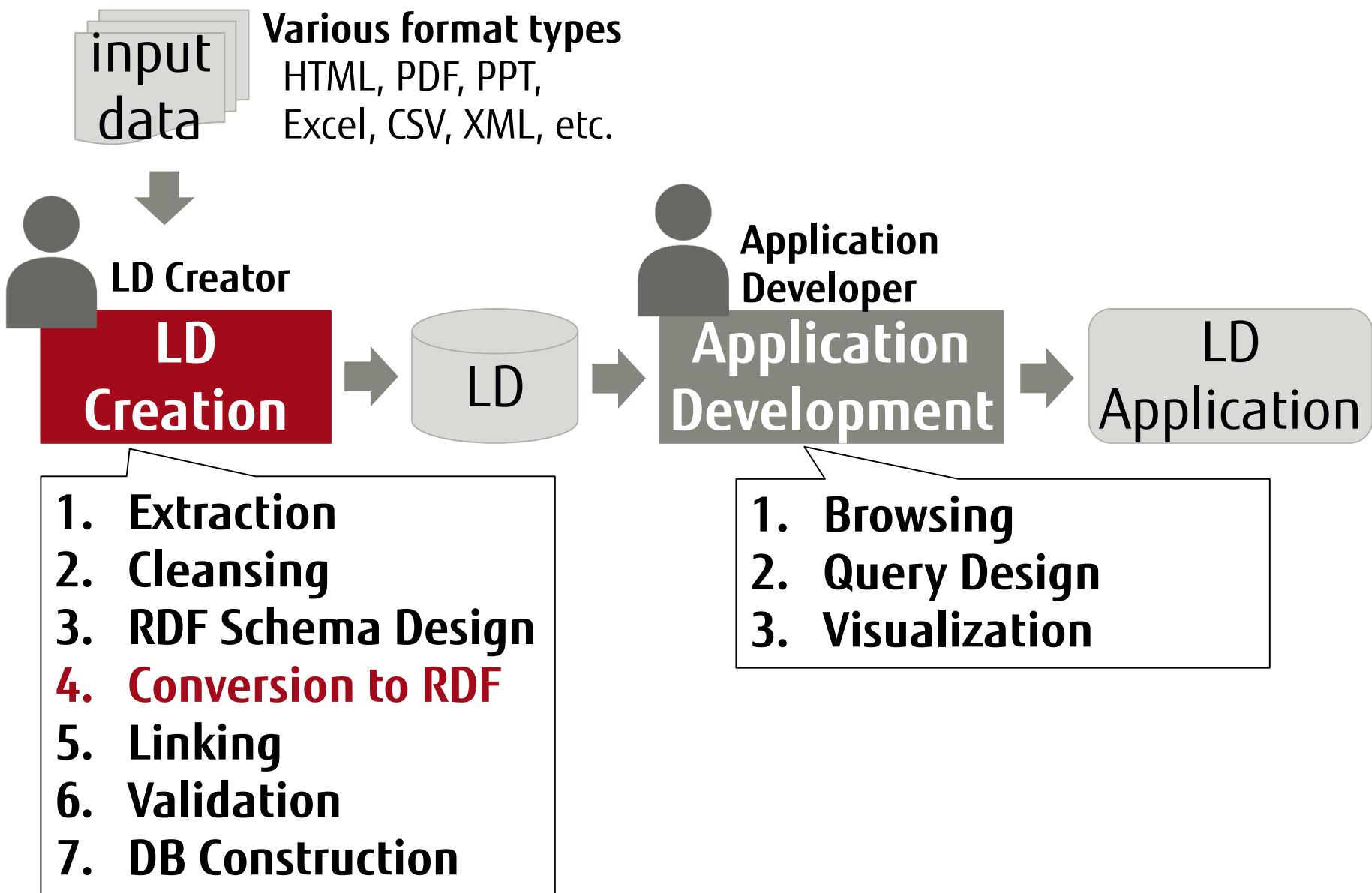
My Favorite Things

FUJITSU

Oilfish / Escalar Sashimi



Outline



Conversion to RDF

Convert an original dataset and an RDF schema to RDF format

Ex. JIST data



■ JSON/CSV/XML to RDF

- Open Refine + RDF Extension★

- YARRRML

■ CSV to RDF

- tarql

Open Refine + RDF Extension ★ Conversion to RDF (1/3) FUJITSU

<https://github.com/sparkica/rdf-extension>

- The skeleton specifies how the RDF data will be generated from your grid-shaped data.

Base URI: <https://lod4all.github.io/jist2018-tutorial/jist.ttl#edit>

RDF Skeleton RDF Preview

Available Prefixes: jist rdf rdfs dblp dc [+ add prefix](#) [manage prefixes](#)

_ - id URI
add rdf:type

_ - authors - authors URI
add rdf:type

Add another root node

The screenshot displays the 'RDF Skeleton' tab of the Open Refine RDF Extension. It shows a complex mapping rule for generating RDF triples from a grid. The rule uses various predicates such as dc:identifier, dc:subject, dc:title, rdf:type, and dblp:title to map cells from the grid to RDF concepts and URIs. The 'Available Prefixes' section lists jist, rdf, rdfs, dblp, and dc. The main area shows a hierarchical tree of mapping rules, with some nodes expanded to show their details. The 'Add another root node' button is visible at the bottom left.

Open Refine + RDF Extension ★ Conversion to RDF (1/3)

<https://github.com/sparkica/rdf-extension>

FUJITSU

Base URI: <https://lod4all.github.io/jist2018-tutorial/jist.ttl#edit>

RDF Skeleton

RDF Preview

Available Prefixes:

jist rdf rdfs dblp dc [+ add prefix](#) [manage prefixes](#)

_ - id URI

[add rdf:type](#)

[X ➔ dc:identifier](#) _ - id cell

[⊖ →](#) [X ➔ dc:subject](#) _ - category cell

[→](#) [X ➔ dblp:authoredBy](#) _ - authors -
authors URI

[→](#) [add rdf:type](#)

[X ➔ dc:title](#) → _ - title cell

[X ➔ rdf:type](#) → <https://dblp.org/rdf/schema-2017...> [+ add rdf:type](#)

[X ➔ dblp:title](#) → _ - title cell
[⊖ →](#)

[add property](#)

[X ➔ rdfs:label](#) →

[X ➔ rdf:type](#) →

_ - authors - authors cell

<https://dblp.org/rdf/schema-2017...> [+ add rdf:type](#)

_ - authors - authors URI

[add rdf:type](#)

Add another root node

■ Human-readable text-based representation for declarative Linked Data generation rules.

- A subset of YAML, a widely used data serialization language designed to be human-friendly.
- Can already be used to represent R2RML and RML rules.

```
prefixes:  
jist: https://lod4all.github.io/jist2018-tutorial/jist.ttl#  
dblp: https://dblp.org/rdf/schema-2017-04-18#  
mappings:  
  accepted_papers:  
    sources:  
      - ['jist.json~jsonpath', '$.accepted_papers[*]']  
    s: https://lod4all.github.io/jist2018-tutorial/jist.ttl#${(id)}  
    po:  
      - [a, dblp:Publication]  
      - [dblp:authoredBy, jist:${(authors)}~iri]  
      - [dblp:title, ${title}]
```

■ Command-line tool for converting CSV files to RDF

- Uses SPARQL 1.1 syntax
- Written in Java and based on Apache ARQ.

■ Usage

```
CONSTRUCT {  
    ?URI a dblp:Publication;  
          dblp:authoredBy ?AUTHOR;  
          dc:title ?title ;  
          dc:category ?category ;  
}  
FROM <file:jist.csv>  
WHERE {  
    BIND (URI(?id) AS ?URI)  
    BIND (URI(?authors)  
AS ?AUTHOR)  
}
```

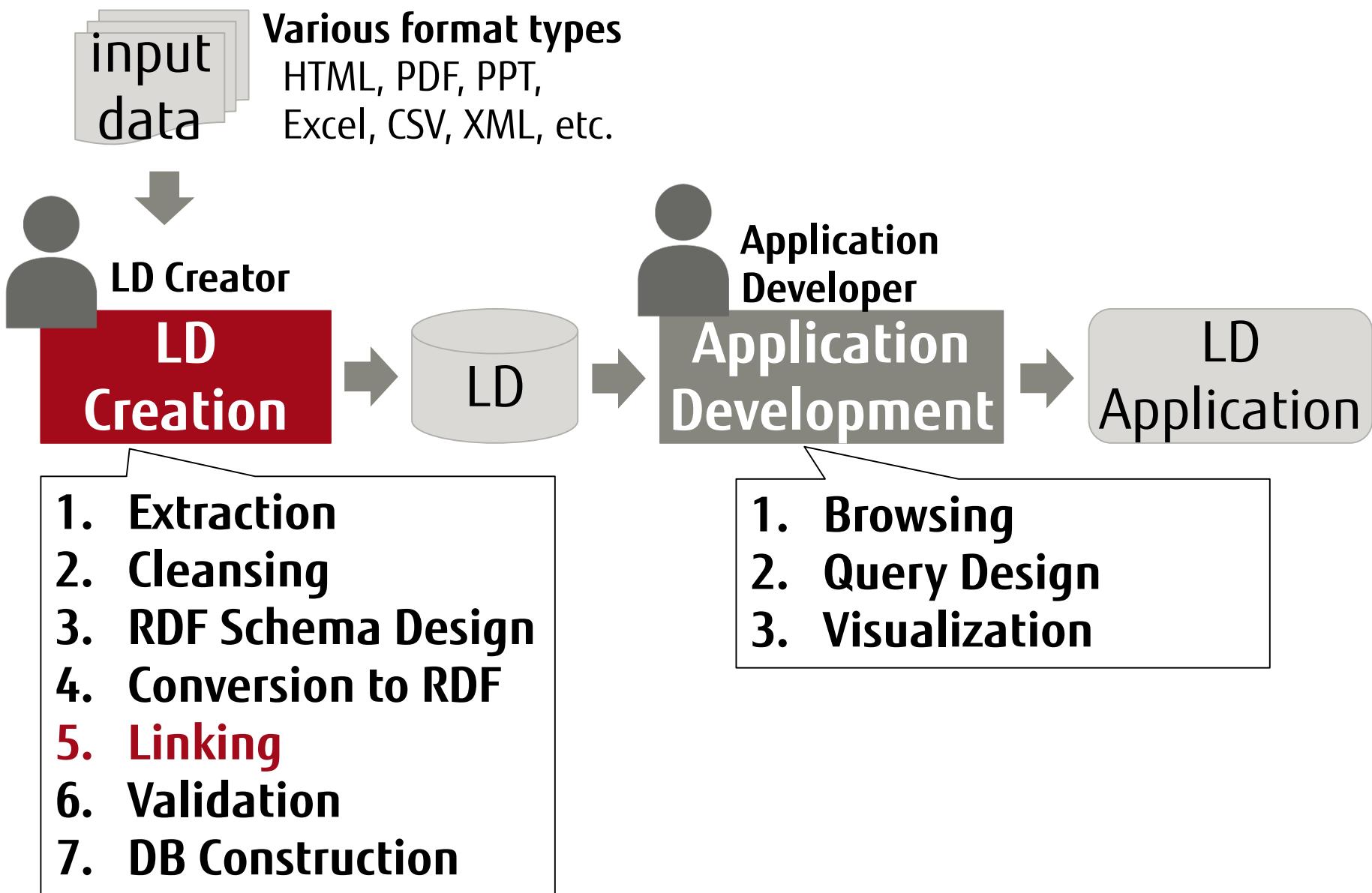
My Favorite Things

FUJITSU

Sky Diving(Hawaii)

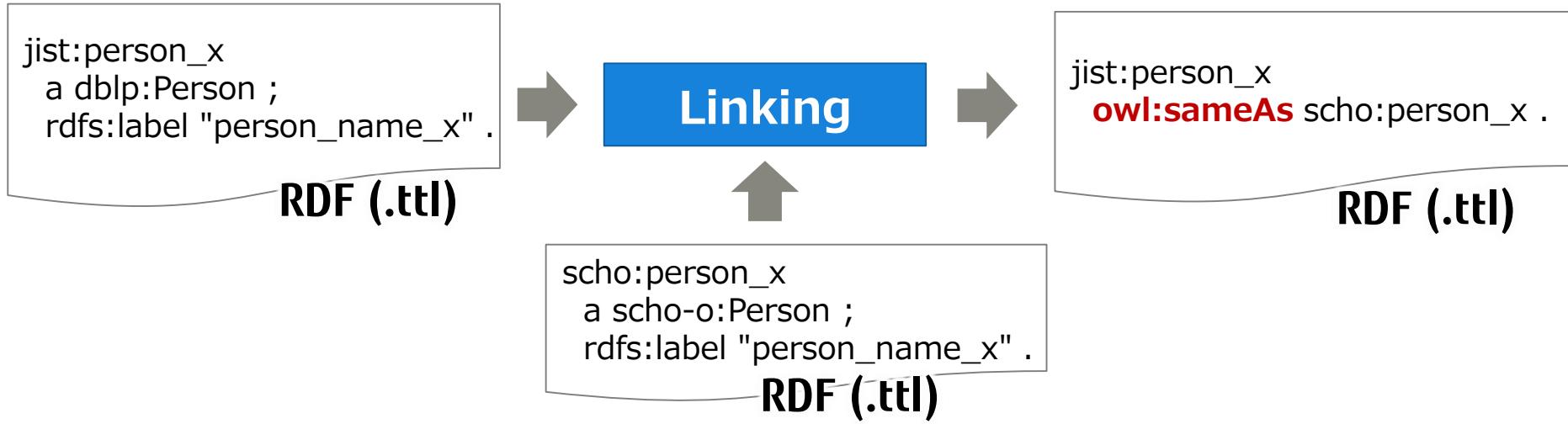


Outline



Link RDF data with another dataset
in order to enable cross-sectional search by SPARQL

Ex. JIST data

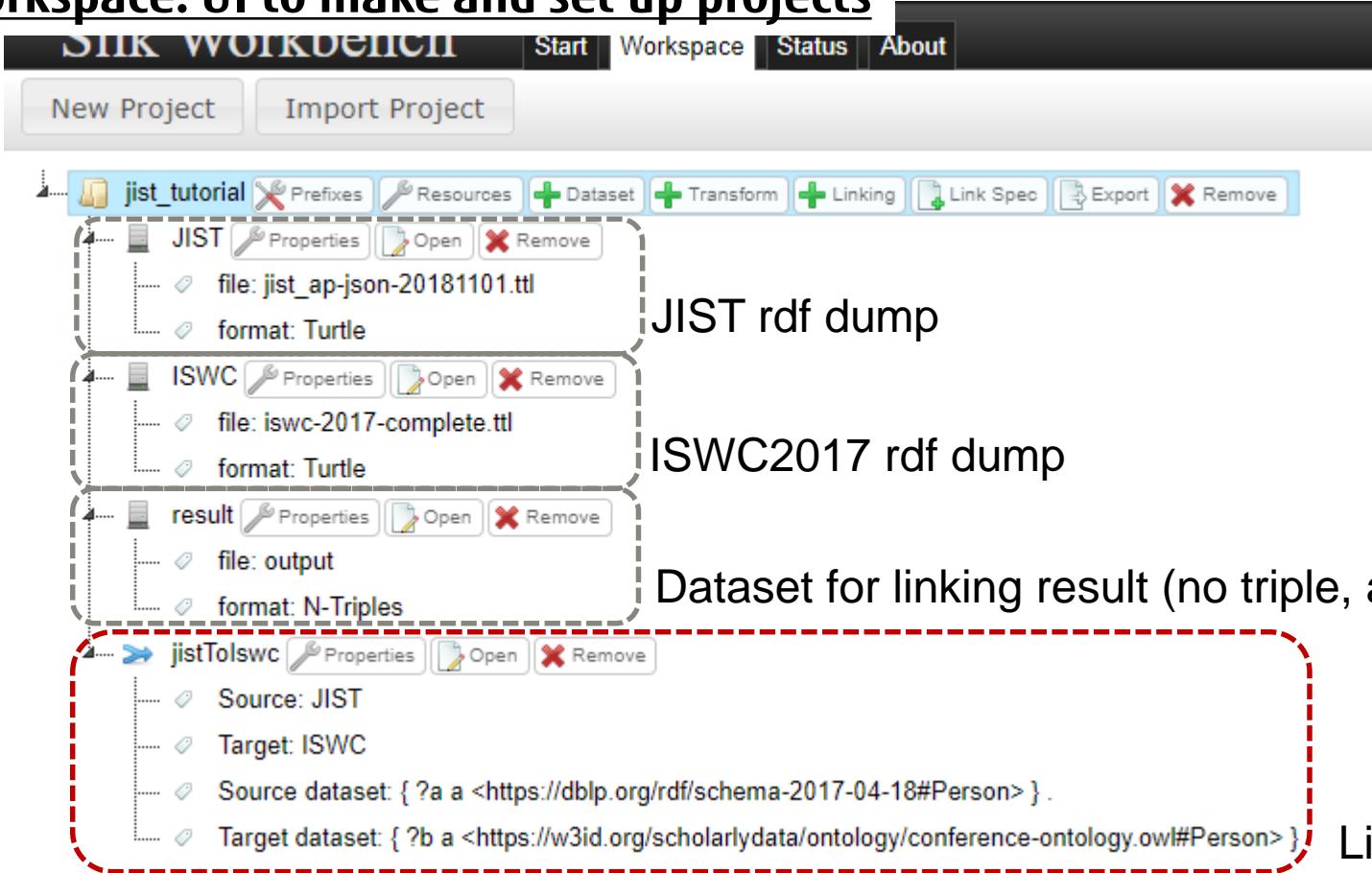


■ Linking tool

■ Silk★

■ Integration tool for heterogeneous data sources

Workspace: UI to make and set up projects



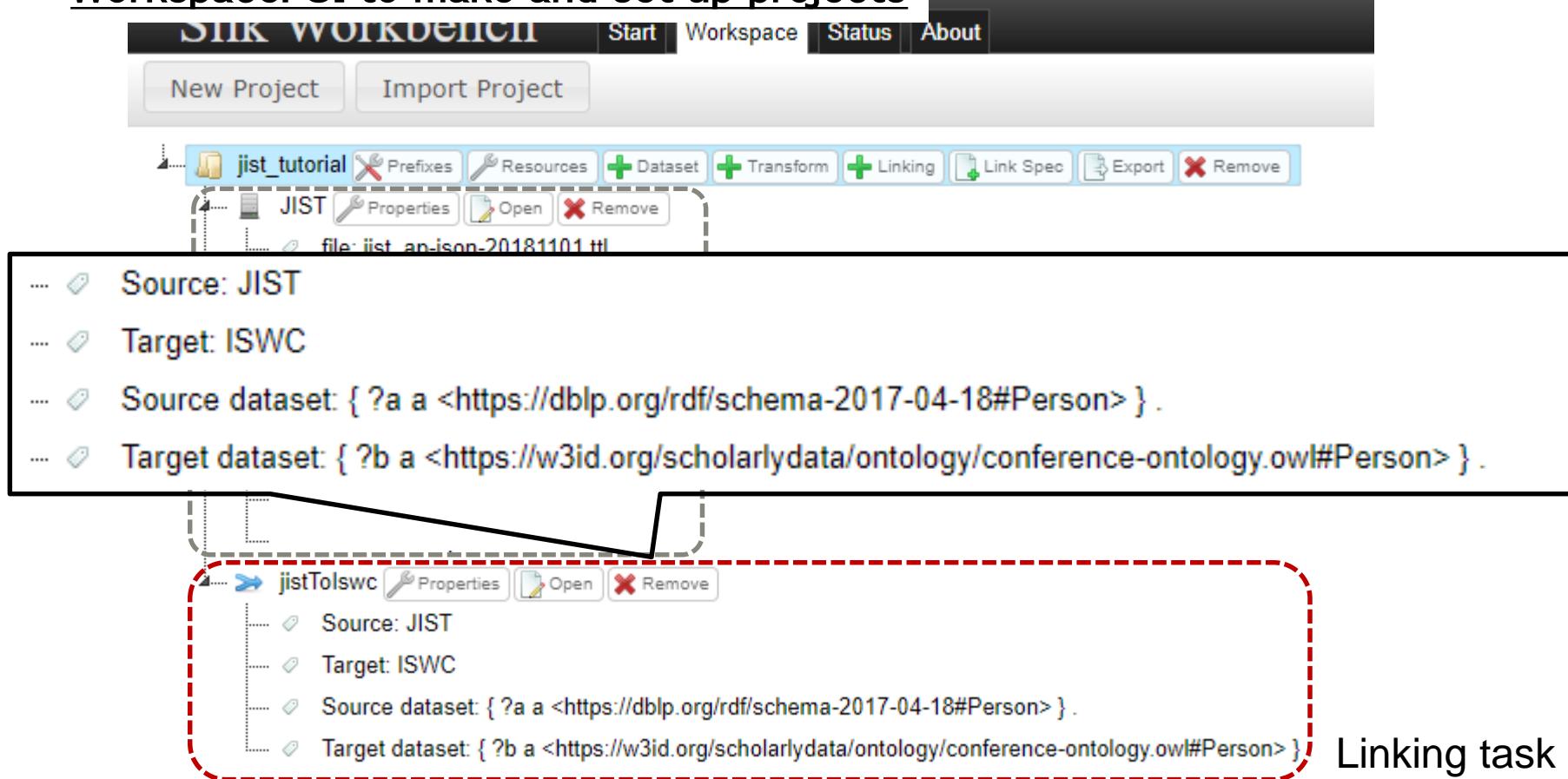
The screenshot shows the Silk Workspace interface with the following components:

- Toolbar:** Includes "New Project", "Import Project", "Start", "Workspace" (selected), "Status", and "About".
- Project Structure:** A tree view under "jist_tutorial" containing:
 - JIST:** Properties, Open, Remove. Contains "file: jist_ap-json-20181101.ttl" and "format: Turtle".
 - ISWC:** Properties, Open, Remove. Contains "file: iswc-2017-complete.ttl" and "format: Turtle".
 - result:** Properties, Open, Remove. Contains "file: output" and "format: N-Triples".
 - jistTolswc:** Properties, Open, Remove. Contains "Source: JIST", "Target: ISWC", "Source dataset: { ?a a <https://dblp.org/rdf/schema-2017-04-18#Person> } .", and "Target dataset: { ?b a <https://w3id.org/scholarlydata/ontology/conference-ontology.owl#Person> } .". This item is highlighted with a red dashed circle.
- Labels:** "JIST rdf dump", "ISWC2017 rdf dump", "Dataset for linking result (no triple, at first)", and "Linking task".

You can register datasets such as RDF dumps and SPARQL endpoints.

■ Integration tool for heterogeneous data sources

Workspace: UI to make and set up projects



The screenshot shows the Silk Workspace interface. At the top, there is a navigation bar with tabs: SILK WORKBENCH, Start, Workspace (which is selected), Status, and About. Below the navigation bar, there are two buttons: New Project and Import Project. The main workspace displays a project structure under the 'jist_tutorial' folder:

- Source: JIST
- Target: ISWC
- Source dataset: { ?a a <https://dblp.org/rdf/schema-2017-04-18#Person> } .
- Target dataset: { ?b a <https://w3id.org/scholarlydata/ontology/conference-ontology.owl#Person> } .

Below this, a 'jistTolswc' task is listed, which contains the same four items. A red dashed box encloses the 'jistTolswc' task, and the text 'Linking task' is written next to it.

You can register datasets such as RDF dumps and SPARQL endpoints.

Editor: UI for defining the flow of the Linking task

Source

Target

Transformations

Comparators

Aggregators



Editor:

Source:

- (custom path)
- ?a/<https://dblp.org/rdf/schema-2017-01-01.ttl#
- ?a/rdf:type
- ?a/rdfs:label
- ?a/<https://dblp.org/rdf/schema-2017-01-01.ttl#
- ?a/<https://dblp.org/rdf/schema-2017-01-01.ttl#

Link Target:

- (custom path)
- ?b/<https://w3id.org/scholarlydata/
- ?b/<http://xmlns.com/foaf/0.1/mbo#
- ?b/rdfs:label
- ?b/<https://w3id.org/scholarlydata/
- ?b/rdf:type

Transformations Recommended ▾

- Constant
- Lower case
- Tokenize

Comparators Recommended ▾

- Equality
- Jaccard
- Levenshtein distance

Aggregators Recommended ▾

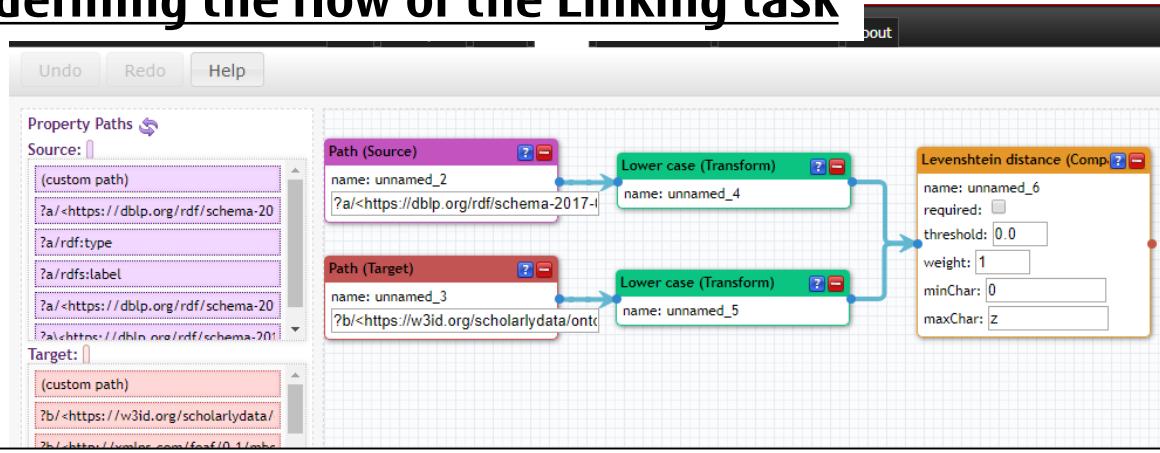
- Average
- Maximum
- Minimum

Maximum
Minimum

Editor: UI for defining the flow of the Linking task

Source

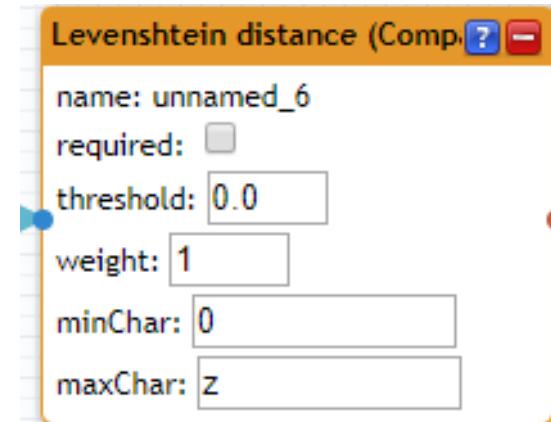
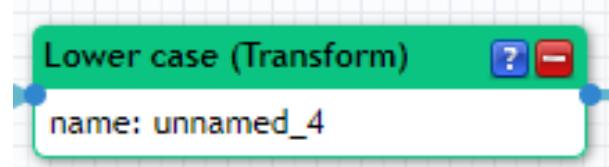
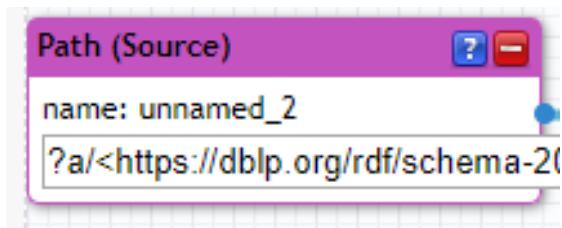
Target



1. Retrieve entity's label
(person's name)

2. Transform labels to lowercase

3. Calculate score with
Levenshtein distance



Generate Links: UI to run the linking task and check the results

Silk Workbench Start | Workspace | Status | Editor | Generate Links | Reference Links | About

Start | Stop Generate Links finished in 578ms

Expand All | Collapse All | Prev | 1 | Next | Filter:

Source:	Target:	Score	Correct
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Arild+Waaler	https://w3id.org/scholarlydata/person/arild-waaler	100.0%	
Comparison:levenshteinDistance (unnamed_6) 100.0%			
Transform:lowerCase (unnamed_4) arild waaler			
Input:?a<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Arild Waaler			
Transform:lowerCase (unnamed_5) arild waaler			
Input:?b<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Arild Waaler			
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Evgeny+Kharlamov	https://w3id.org/scholarlydata/person/evgeny-kharlamov	100.0%	
Comparison:levenshteinDistance (unnamed_6) 100.0%			
Transform:lowerCase (unnamed_4) evgeny kharlamov			
Input:?a<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Evgeny Kharlamov			
Transform:lowerCase (unnamed_5) evgeny kharlamov			
Input:?b<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Evgeny Kharlamov			
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Thepchai+Supnithi	https://w3id.org/scholarlydata/person/thechai-supnithi	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Hong+Gee+Kim	https://w3id.org/scholarlydata/person/hong-gee-kim	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Hideaki+Takeda	https://w3id.org/scholarlydata/person/hideaki-takeda	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Ognjen+Savkovic	https://w3id.org/scholarlydata/person/ognjen-savkovic	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Khai+Nguyen	https://w3id.org/scholarlydata/person/khai-nguyen	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Kouji+Kozaki	https://w3id.org/scholarlydata/person/kouji-kozaki	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Guilin+Qi	https://w3id.org/scholarlydata/person/guilin-qи	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Jeff+Z.+Pan	https://w3id.org/scholarlydata/person/jeff-z-pan	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Ryutaro+Ichise	https://w3id.org/scholarlydata/person/ryutaro-ichise	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Jin+Dong+Kim	https://w3id.org/scholarlydata/person/jin-dong-kim	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Rub%C3%A9n+Francisco+Manrique	https://w3id.org/scholarlydata/person/ruben-francisco-manrique	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Takahiro+Kawamura	https://w3id.org/scholarlydata/person/takahiro-kawamura	100.0%	
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Martin+Giese	https://w3id.org/scholarlydata/person/martin-giese	100.0%	

Prev | 1 | Next

Generate Links: UI to run the linking task and check the results

Silk Workbench Start | Workspace | Status | Editor | Generate Links | Reference Links | About

Start | Stop | Generate Links finished in 578ms

Expand All | Collapse All | Prev | 1 | Next | Filter:

Source:	Target:	Score	Correct
JIST	ISWC	100.0%	✓ ? ✎
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Arild+Waaler	https://w3id.org/scholarlydata/person/arild-waaler	100.0%	✓ ? ✎
Comparison:levenshteinDistance (unnamed_6) 100.0%			
Transform:lowerCase (unnamed_4) arild waaler			
Input:?a/<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Arild Waaler			
Transform:lowerCase (unnamed_5) arild waaler			
Input:?b/<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Arild Waaler			
all.github.io/jist2018-tutorial/jist.ttl#Evgeny+Kharlamov		100.0%	✓ ? ✎
n:levenshteinDistance (unnamed_6) 100.0%			

Each result

https://lod4all.github.io/jist2018-tutorial/jist.ttl#Arild+Waaler | https://w3id.org/scholarlydata/person/arild-waaler

Comparison:levenshteinDistance (unnamed_6) 100.0%

 Transform:lowerCase (unnamed_4) arild waaler

 Input:?a/<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Arild Waaler

 Transform:lowerCase (unnamed_5) arild waaler

 Input:?b/<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Arild Waaler

https://lod4all.github.io/jist2018-tutorial/jist.ttl#Jin+Dong+Kim	https://w3id.org/scholarlydata/person/jin-dong-kim	100.0%	✓ ? ✎
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Ryutaro+Ichise	https://w3id.org/scholarlydata/person/ryutaro-ichise	100.0%	✓ ? ✎
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Takahiro+Kawamura	https://w3id.org/scholarlydata/person/takahiro-kawamura	100.0%	✓ ? ✎
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Rub%C3%A9n+Francisco+Manrique	https://w3id.org/scholarlydata/person/ruben-francisco-manrique	100.0%	✓ ? ✎
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Martin+Giese	https://w3id.org/scholarlydata/person/martin-giese	100.0%	✓ ? ✎

Prev | 1 | Next

Generate Links: UI to run the linking task and check the results

Silk Workbench Start | Workspace | Status | Editor | Generate Links | Reference Links | About

Start | Stop | Generate Links finished in 578ms

Expand All | Collapse All | Prev | 1 | Next | Filter:

Source: JIST	Target: ISWC	Score	Correct
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Arild+Waaler	https://w3id.org/scholarlydata/person/arild-waaler	100.0%	  

Comparison:levenshteinDistance (unnamed_6) 100.0%

- Transform:lowerCase (unnamed_4) arild waaler
 - Input: ?a/<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Arild Waaler
- Transform:lowerCase (unnamed_5) arild waaler
 - Input: ?b/<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Arild Waaler

Detail of each result

https://lod4all.github.io/jist2018-tutorial/jist.ttl#Arild+Waaler | https://w3id.org/scholarlydata/person/arild-waaler

Comparison:levenshteinDistance (unnamed_6) 100.0%

- Transform:lowerCase (unnamed_4) arild waaler
 - Input: ?a/<https://dblp.org/rdf/schema-2017-04-18#primaryFullPersonName> (unnamed_2) Arild Waaler
- Transform:lowerCase (unnamed_5) arild waaler
 - Input: ?b/<https://w3id.org/scholarlydata/ontology/conference-ontology.owl#name> (unnamed_3) Arild Waaler

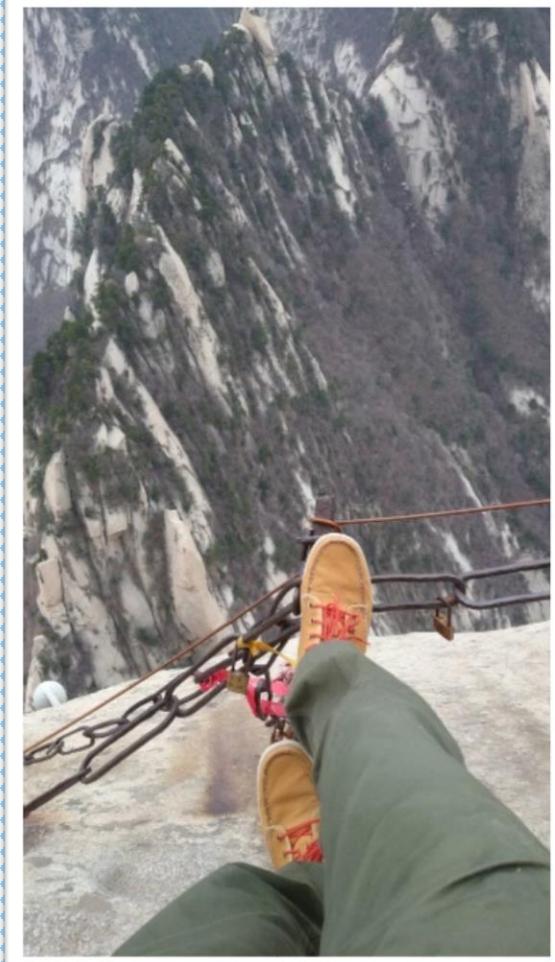
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Ryutaro+Ichise	https://w3id.org/scholarlydata/person/ryutaro-ichise	100.0%	  
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Jin-Dong+Kim	https://w3id.org/scholarlydata/person/jin-dong-kim	100.0%	  
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Rub%C3%A9n+Francisco+Manrique	https://w3id.org/scholarlydata/person/ruben-francisco-manrique	100.0%	  
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Takahiro+Kawamura	https://w3id.org/scholarlydata/person/takahiro-kawamura	100.0%	  
https://lod4all.github.io/jist2018-tutorial/jist.ttl#Martin+Giese	https://w3id.org/scholarlydata/person/martin-giese	100.0%	  

Prev | 1 | Next

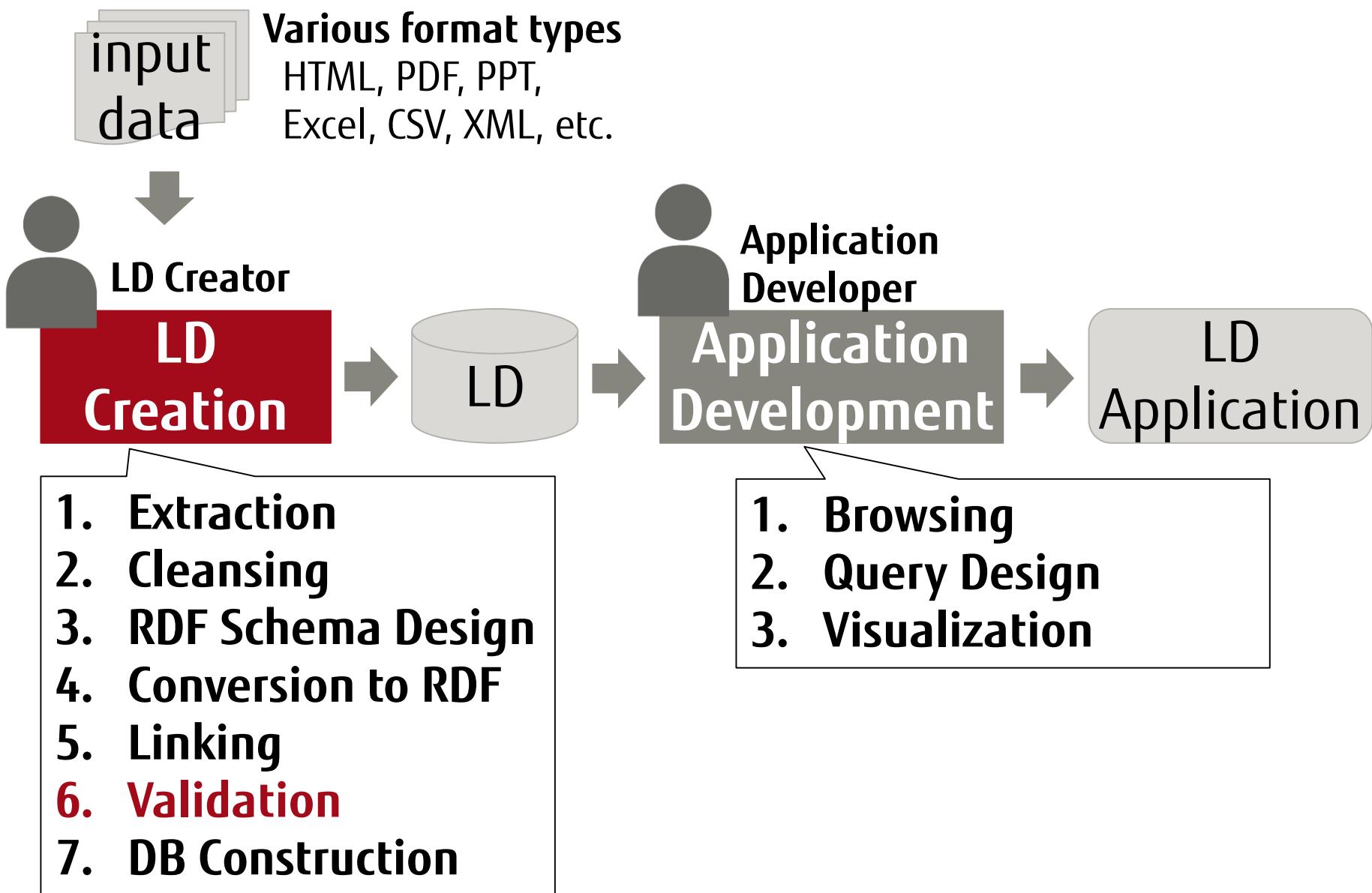
My Favorite Things

FUJITSU

Mount Hua / 華山 (China)



Outline



Validation

To validate the syntax and the schema for the created RDF files

```
jist:xxx  
dblp:identifier "id_x"  
dblp:authorBy dblp:author_x ;  
dblp:title "title_x".  
dblp:author_x  
dblp:authorOf jist:xxx ;  
dblp:authorOf jist:yyy .
```

RDF(.ttl)



Validation



line 2:
syntax error

Validation
Report

■ Syntax validation

- rapper★, Serd, arq

■ Schema validation

- Simple validation: SPARQL

- Languages for validation: SHACL★, ShEx

■ A RDF parsing and serializing utility

- Raptor RDF Syntax Library

■ Usage

- RDF serializing
 - rapper –output <format> –input <format> <inputfile>

- GraphViz DOT Serializer
 - rapper --output dot --input <format> <file> | dot -Tpng -o output.png

- Validating RDF content 

- rapper -c –input <format> <file>

■ Problem

- Official site version difficult to handle huge files

- *Latest version: 2.0.15

- *Github version improved this problem by reading input in chunks

- <https://github.com/dajobe/raptor>

■ A RDF parsing and serializing utility

- Free Software released under the extremely liberal ISC license

■ Features

- It can handle huge files
 - Adopting stream processing
 - Comparing with rapper

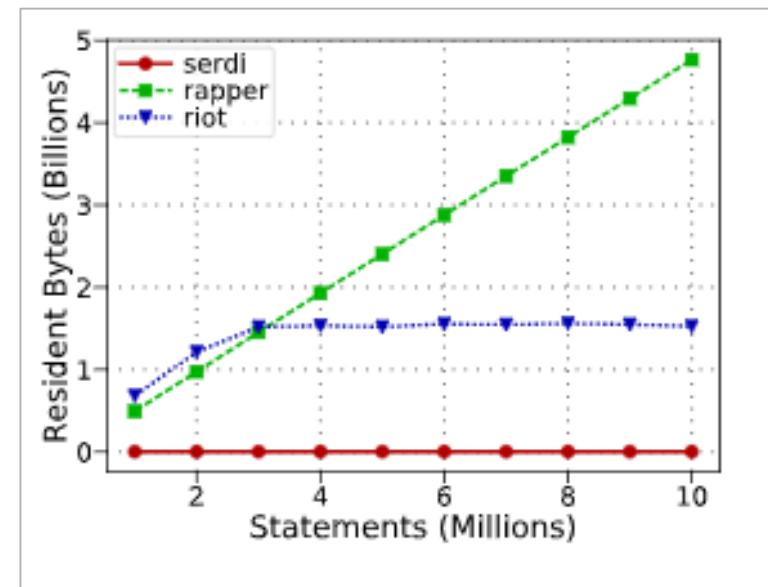


■ Usage

- RDF serializing
 - serdi -i <format> -o <format> input_file
 - serdi -i <format> -o <format> -s input_string
 - format:turtle or Ntriples
 - STD output

■ Others

- Making URI of Blank node



■ Command for executing SPARQL queries for RDF files

- One of Jena package's command-line tools
- arq is used for validating IRI (rapper CANNOT validate IRI)

■ Usage

- arq --data=<file> --query=<queryfile>
- arq --data=<file> '<query>'

■ Comparison of arq and rapper for IRI validation

arq

```
$ arq -data arq_vs_rapprt.ttl --query sample.rq
10:38:40 WARN riot :: [line: 6, col:
1 ] Bad IRI: <http://bad://example.com#bob> Code:
12/PORT_SHOULD_NOT_BE_EMPTY in PORT: The colon
introducing an empty port component should be omitted
entirely, or a port number should be specified.
```

found Bad IRI
(has double slashes)

rapper

```
$ rapper -c --input turtle arq_vs_rapprt.ttl
rapper: Parsing URI file:///root/arq_vs_rapprt.ttl with
parser turtle
rapper: Parsing returned 3 triples
```

NOT found Bad IRI

- It is a means w/o validation tools
- Invalid RDF Data

```
# cat jist_ap-json.include.invalidation.data.ttl | head -n 10  
<https://lod4all.github.io/9> dc:identifier "9" , "99999";
```

■ SPARQL Query

```
# cat sparql/rule.rq  
CONSTRUCT { ?paper v:isValid false . }  
WHERE{ FILTER(?number_of_id != 1)  
{ SELECT ?paper (COUNT(DISTINCT ?id) AS ?number_of_id)  
WHERE {  
    ?paper a dblp:Publication ;  
           dc:identifier ?id .  
} GROUP BY ?paper  
}  
}
```



multiple IDs

check number of IDs

■ Result

```
# arq -data jist_ap-json.invalidation.data.ttl -query sparql/rule.rq  
<https://lod4all.github.io/9> v:isValid false .
```

- Describes structures (constraints) of an RDF graph as Shapes
 - ex) : a resource has (or doesn't have) specific predicates .

- Usage

```
sh:rule1
  a sh:NodeShape ;
  sh:targetClass dblp:Publication ;
  sh:property [
    sh:path dc:identifier ;
    sh:minCount 1 ;
    sh:maxCount 1 ;
    sh:nodetype xsd:string;
  ] .
```

Shape file

constraints for number of IDs

```
# pyshacl -s shacl/rule.ttl -d jist_ap-json-20181101.ttl
```

Validation Report

Conforms: False

Results (1):

Constraint Violation in MaxCountConstraintComponent

(<http://www.w3.org/ns/shacl#MaxCountConstraintComponent>):

Severity: sh:Violation

Source Shape: [sh:maxCount Literal("1", datatype=xsd:integer) ; ...]

Focus Node: <<https://lod4all.github.io/jist2018-tutorial/jist.ttl#9>>

Result Path: dc:identifier

Result

- Language for constraints with equivalent function to SHACL
 - Compact notation is implemented unlike SHACL
- Comparison for Shape files

SHACL

```
sh:property [  
    sh:path dc:identifier ;  
    sh:minCount 1 ;  
    sh:maxCount 1 ;  
    sh:nodetype xsd:string;  
] .
```

RDF syntax
(machine-readable)

ShEx

```
dc:identifier xsd:string{1}
```

ShexC syntax
(human-readable)

My Favorite Things

FUJITSU

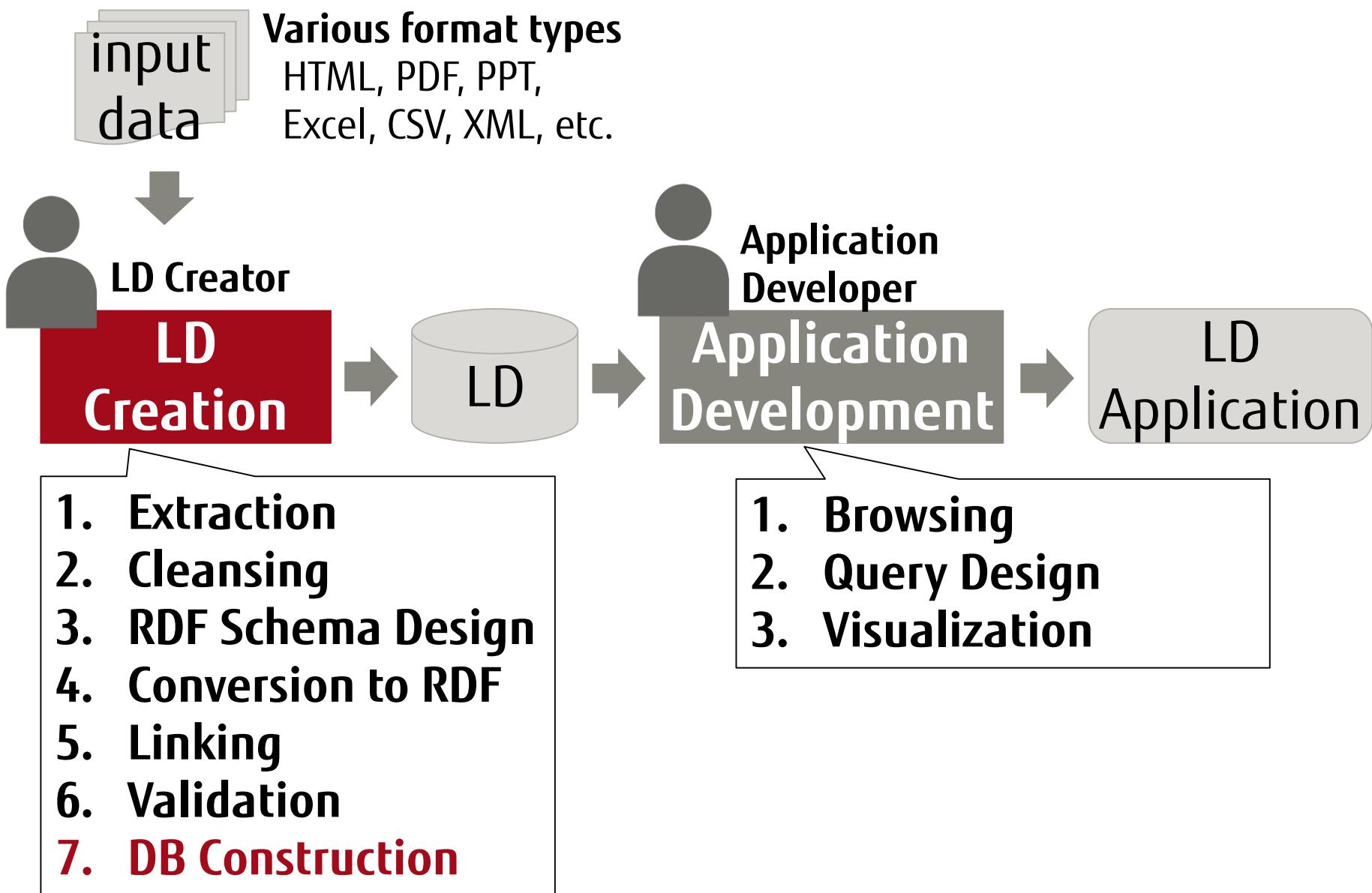
Soft Roe Nabe/白子鍋



Kasemasa Bar / 加瀬政
Tokyo, Japan

<https://tabelog.com/tokyo/A1323/A132301/13003798/>

Outline



Loading the dataset into the DB to enable users to access the SPARQL endpoint

Ex. JIST data

```
jist:xxx  
dblp:identifier "id_x" ;  
dblp:authorBy dblp:author_x ;  
dblp:title "title_x" .  
dblp:author_x  
dblp:authorOf jist:xxx ;  
dblp:authorOf jist:yyy .
```

RDF(.ttl)



DB Construction (cont.)

■ OpenLink Virtuoso★ <https://virtuoso.openlinksw.com/>

- Latest version: 07.20.3229
- License: Single Server Edition: Free, Universal Server: Commercial license
- Acquisition:
 - git clone -b stable/7 <https://github.com/openlink/virtuoso-opensource>

■ StarDog <https://www.stardog.com/>

- Latest version: Stardog 5.3.5 (3 Oct 2018)
- Acquisition: Requiring user registration

■ AllegroGraph <https://franz.com/agraph/allegrograph/>

- Latest version : Release 6.4.4
- Acquisition: <https://franz.com/agraph/downloads/server?ui=new>

■ GraphDB <http://graphdb.ontotext.com/documentation/standard/release-notes.html>

- Latest version : GraphDB 8.7.2 (22 October 2018)
- Acquisition: <https://www.ontotext.com/products/graphdb/>
 - Requiring user registration

■ High performance and large capacity RDF store

■ Single Server Edition:Free

■ Note: In case of clustering system, you need to contract a license

The screenshot shows the Virtuoso CONDUCTOR interface. At the top, there's a navigation bar with links like Home, System Admin, Database, Replication, Web Application Server, SPARQL, Sponger, Statistics, Graphs, Schemas, Namespaces, and View. Below the navigation bar, the main content area has a title "Quad Store Upload". It contains two input fields: one for "File*" with a placeholder "ファイルを選択 選択されていません" and another for "Resource URL*". There's also a checkbox for "Create graph explicitly" and a field for "Named Graph IRI*" containing "http://localhost:8890/DAV". At the bottom right of this form are "Cancel" and "Upload" buttons.

RDF loading UI

The screenshot shows the Virtuoso SPARQL Query Editor. The title bar says "Virtuoso SPARQL Query Editor". Below it, there's a "Default Data Set Name (Graph IRI)" input field. The main area is titled "Query Text" and contains the SPARQL query: "select distinct ?Concept where { [] a ?Concept} LIMIT 100". At the bottom, there are several configuration options: "Results Format" set to "HTML", "Execution timeout" set to "0 milliseconds", "Options" with checkboxes for "Strict checking of void variables" and "Log debug info at the end of output", and buttons for "Run Query" and "Reset". A note at the bottom states: "(The result can only be sent back to browser, not saved on the server, see [details](#))".

SPARQL endpoint UI

My Favorite Things

FUJITSU

Shark Diving



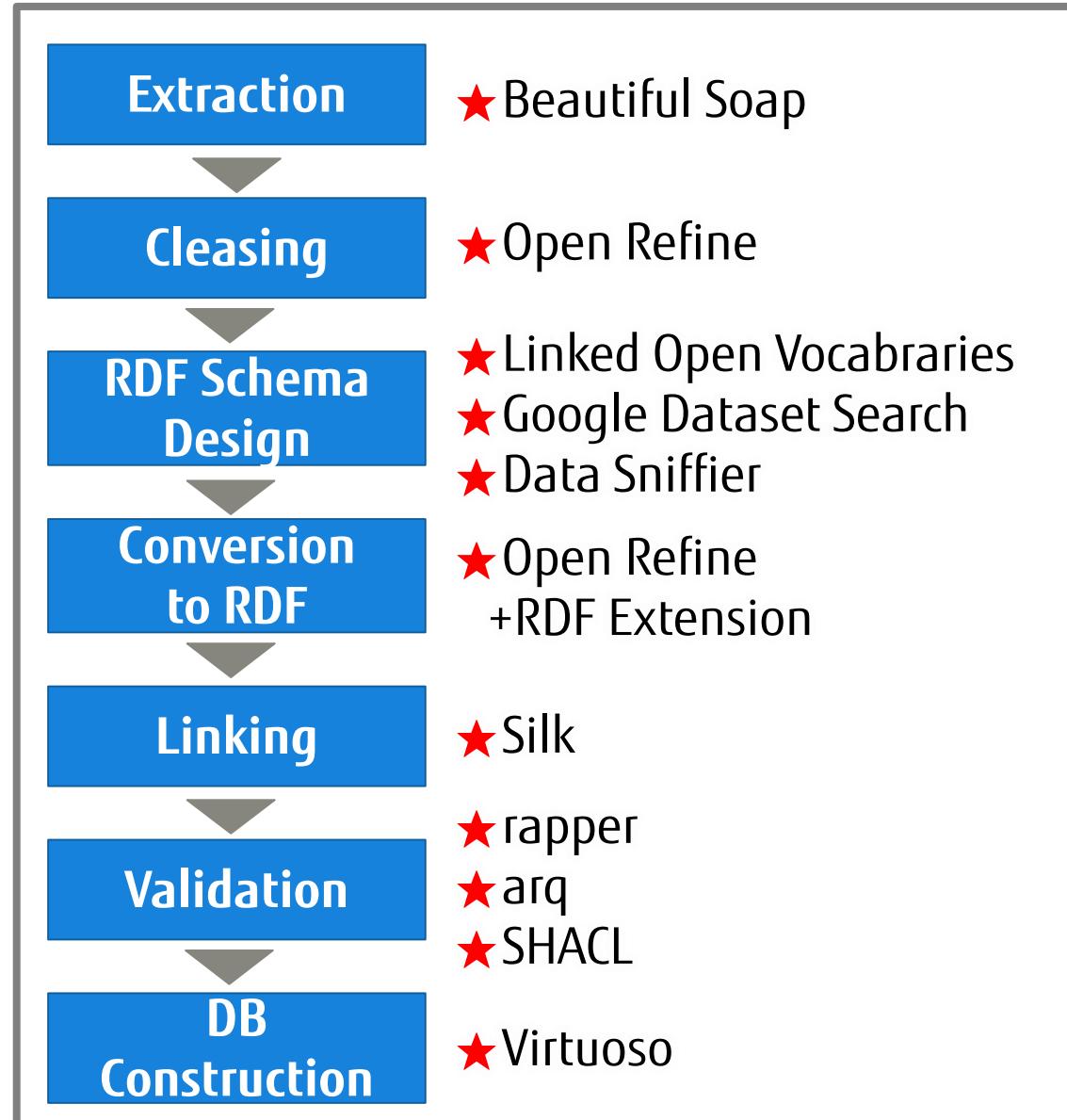
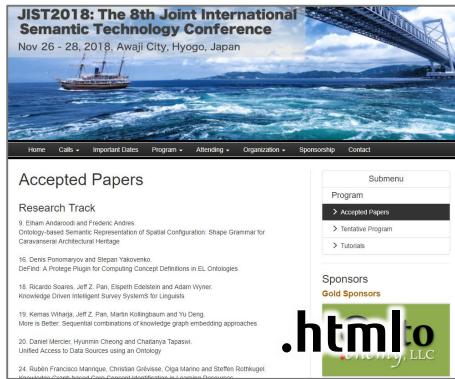
countless sharks...



Chiba, Japan

Review of LD Creation

FUJITSU



jist:xxx
dblp:identifier "id_x" ;
dblp:authorBy dblp:author_x ;
dblp:title "title_x".

SPARQL endpoint
DB



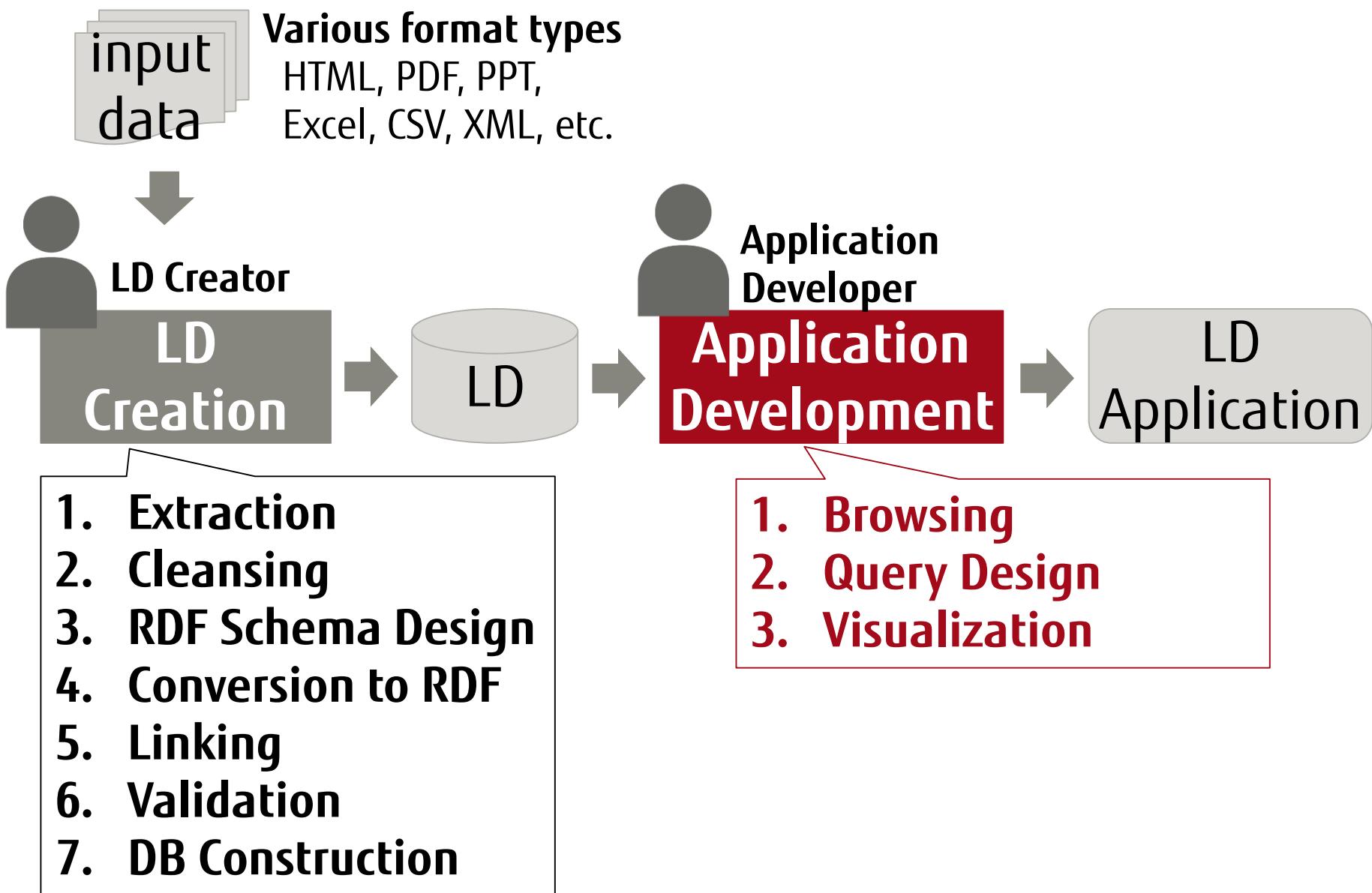
Let's have a break



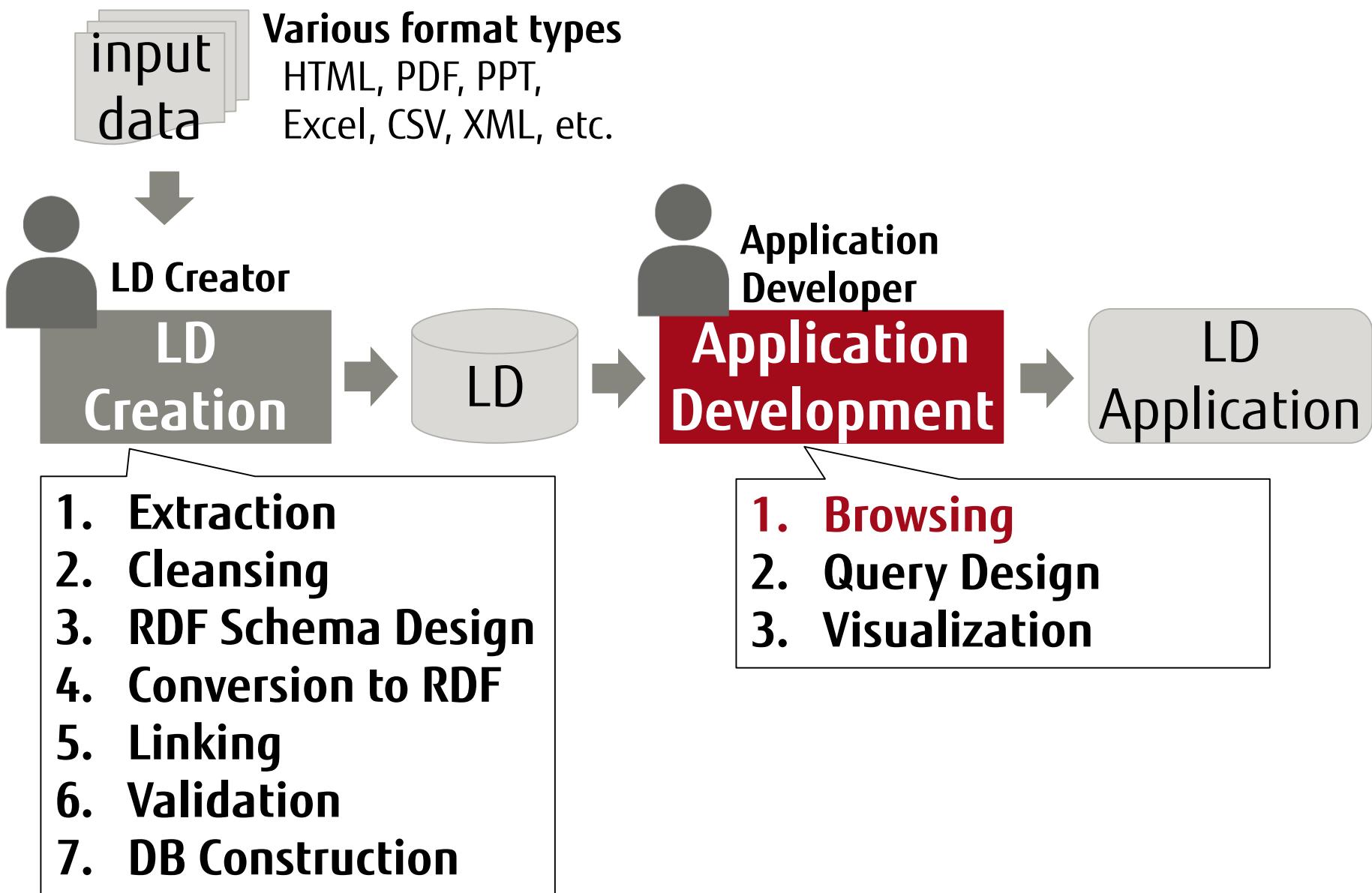
Morning	Introduction to LD tools	9:00-10:00 (60 mins)	Presentation
		10:00-10:10 (10 mins)	Break
		10:10-10:50 (40 mins)	Presentation
		10:50-11:10 (20 mins)	Use Case Studies
Afternoon	Hands-on	13:00-13:10 (10 mins)	Introduction
		13:10-14:55 (95 mins)	Trial 1: Let's visualize corporate data Trial 2: Let's develop a financial analysis application
		14:55-15:00 (5 mins)	Closing

2. Application Development

Outline



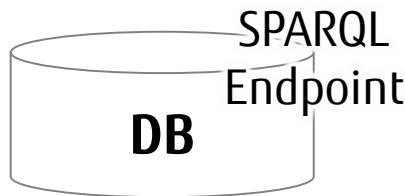
Outline



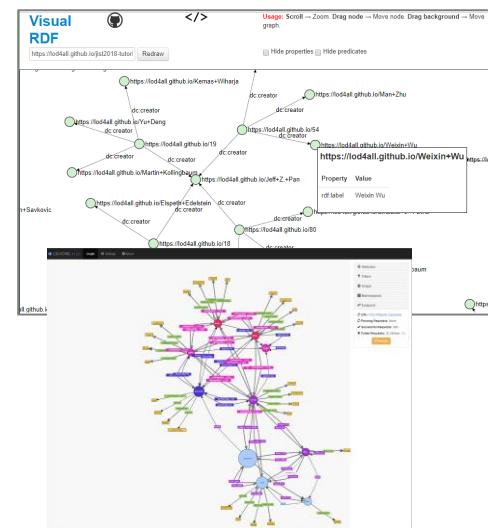
To browse dataset contents from SPARQL endpoints or RDF files directly

Ex. JIST data

```
jist:xxx
  dblp:identifier "id_x" ;
  dblp:authorBy dblp:author_x ;
  dblp:title "title_x" .
  dblp:author_x
    dblp:authorOf jist:xxx ;
    dblp:authorOf jist:yyy
RDF (.ttl)
```



Browsing



- Table: [Pubby★](#), [Elda](#), [DataSniffer](#), [Loupe★](#)
- Graph: [VisualRDF★](#), [Graphviz★](#), [Lodlive](#), [LD-VOWL](#)

■ A Linked Data Frontend for SPARQL Endpoints

- Mapping the original URIs to dereferenceable URIs
- Asking information about the resource via URIs
 - handle requests to the mapped URIs by connecting to the SPARQL endpoint
- Visualizing the response from the SPARQL endpoint as a table type view in one HTML page.

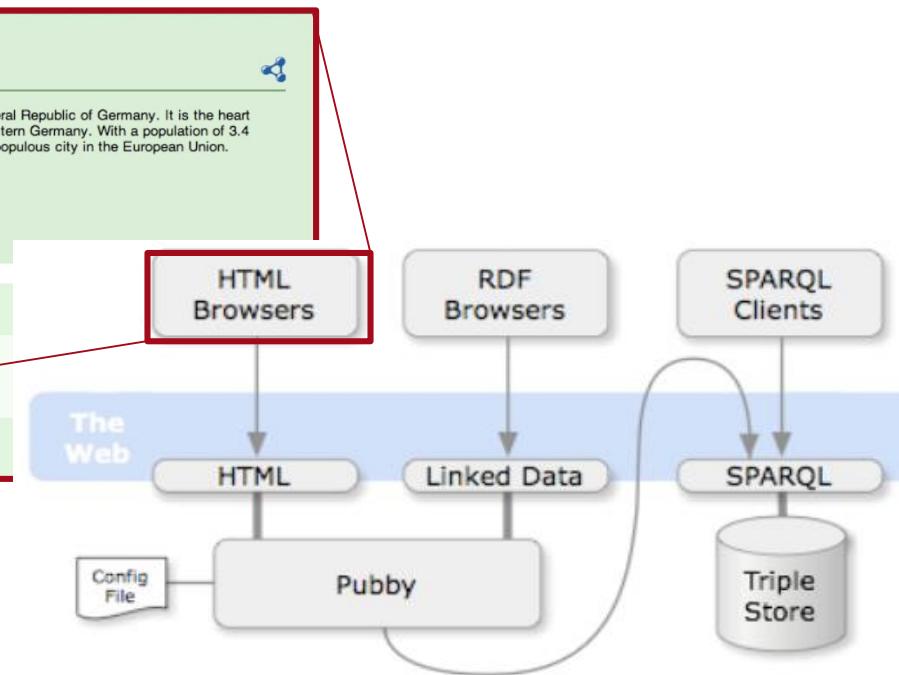
Berlin at DBpedia.org
<http://dbpedia.org/resource/Berlin>



The screenshot shows a table with three columns: Property and Value. The properties listed are 'is p:Origin of', 'is p:PLACE_OF_BIRTH of', and 'is p:PLACE_OF_DEATH of'. The corresponding values are:

Property	Value
is p:Origin of	dbpedia:Alec_Empire dbpedia:Clara_Hill dbpedia:Frank_Duval
is p:PLACE_OF_BIRTH of	dbpedia:Drafi_Deutscher dbpedia:Hannelore_Kohl dbpedia:Hartmut_Mehdorn dbpedia:Julius_Klaproth dbpedia:Otto_Devrient
is p:PLACE_OF_DEATH of	dbpedia:August_Borsig dbpedia:Heinrich_Gr%C3%BCnfeld dbpedia:Johannes_Rau dbpedia:Ludwig_Suthaus

**Table type view
in HTML page**



■ A linked data API for web developers to access linked data from a triple store

- It provides a table type view for RDF data
- The linked data can accessed both by SPARQL query, and directly from JavaScript via a simple, straightforward API.

The screenshot shows the "Linked data API" interface with three tables of game data:

A Brief History of the World	
label	A Brief History of the World
players	3 = , 4 = , 5 = , 6 =
pub year	2009 =
designed by	Ragnar Brothers =
games dataset	dataset =
same as	brief history of the world1.html =

A Few Acres of Snow	
label	A Few Acres of Snow =
play time minutes	60 =
players	2 =
pub year	2011 =
designed by	Martin Wallace =
games dataset	dataset =
same as	a few acres of snow 2 =

Age of Industry	
-----------------	--

Table type view in HTML page

■ Enable to confirm RDF and metadata in HTML

■ Chrome extension

DBLP website

[–] Publication search results

found 36 matches

2017

Franz Baader, Daniel Borchmann, Adrian Nuradiansyah:
Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding.
 g. JIST 2017: 102-117

RDF N-Triples

Turtle

Statement Collection #1	
Entity	https://dblp.uni-trier.de/rec/conf/jist/BaaderBN17
Attributes	https://dblp.org/rec/conf/jist/BaaderBN17

Statement Collection #2	
Entity	https://dblp.org/rec/conf/jist/BaaderBN17
Attributes	https://dblp.uni-trier.de/rdf/schema-2017-04-18#Publication https://doi.org/10.1007/978-3-319-70682-5_7
	The Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding. http://data.bibbase.org/ontology/#Inproceedings https://dblp.uni-trier.de/rdf/schema-2017-04-18#Inproceedings https://dblp.org/pers/b/Baader:Franz https://dblp.org/pers/b/Borchmann:Daniel https://dblp.org/pers/n/Nuradiansyah:Adrian https://doi.org/10.1007/978-3-319-70682-5_7 https://dblp.org/db/conf/jist/jist2017
Entity	JIST 102-117 2017

Statement Collection #3	
Entity	https://dblp.org/rec/conf/jist/BaaderBN17.nt

- A tool that obtains statistics for each dataset

- Searchable element

- class / property(predicate) / triple pattern / named graph

Ex. Class statistic

List of used classes found in the **DBpedia (English)** dataset (941 classes)

Show entries

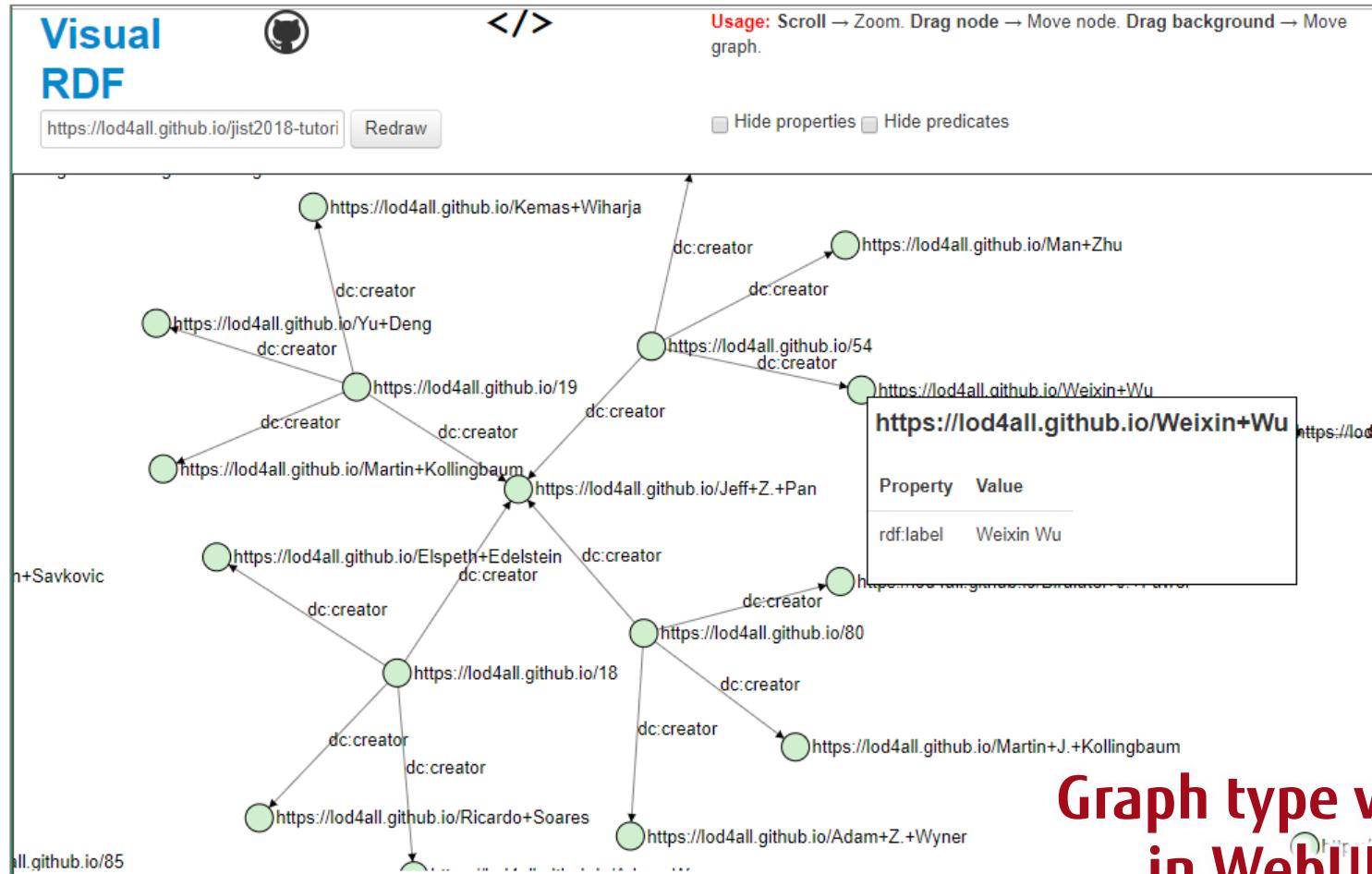
Class	Instance Count
http://dbpedia.org/ontology/Person	2135040 
http://www.wikidata.org/entity/Q5	2135040 
http://dbpedia.org/ontology/Organisation	220219 
http://dbpedia.org/ontology/MusicalWork	162397 

Count of Instances
for each Class

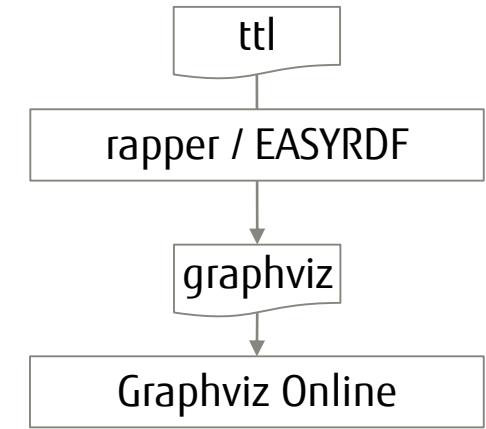
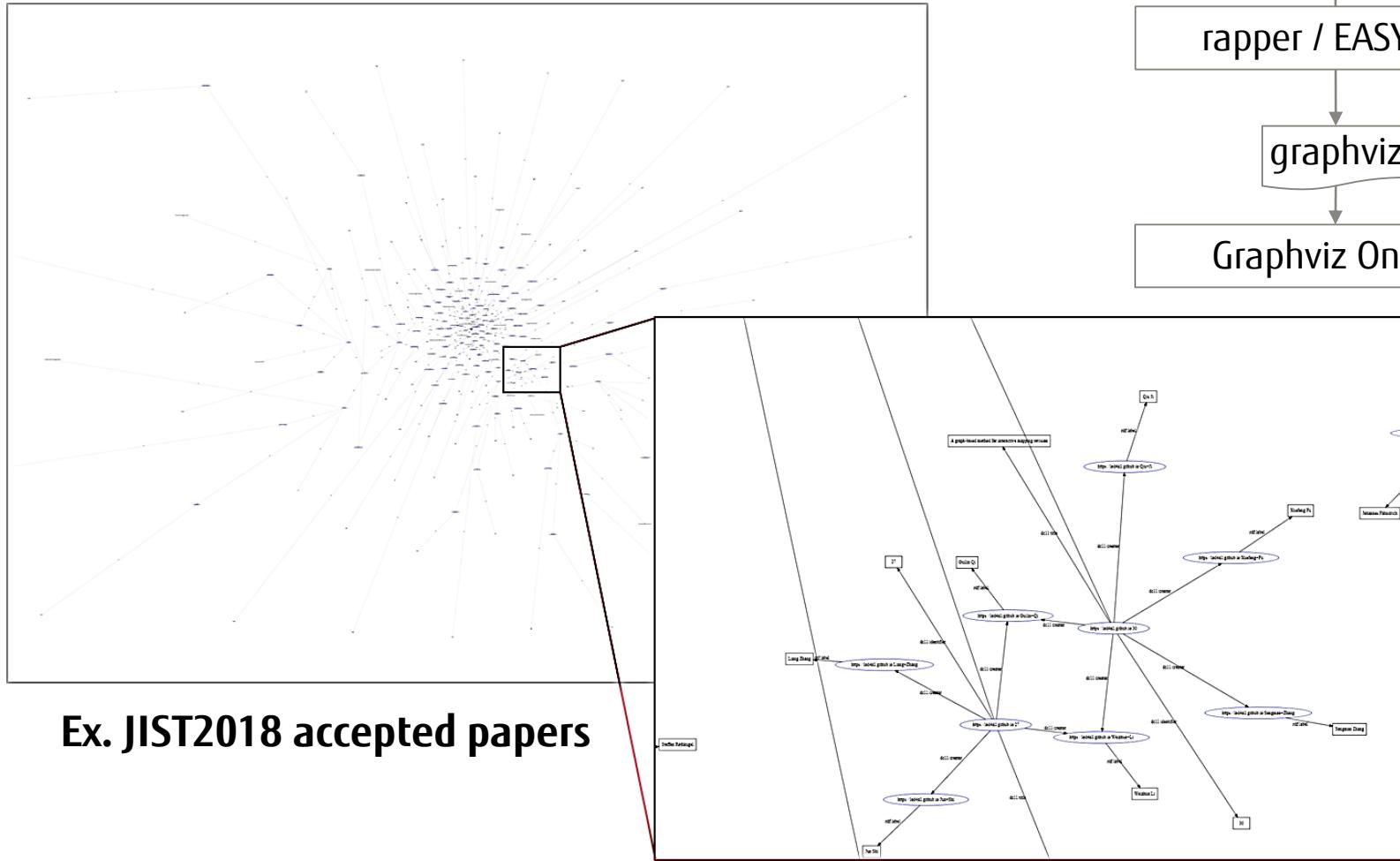
Table type view for statistics

■ Web Service that describe graph from a RDF file easily

Ex. JIST2018 accepted papers



■ Easily visualizing the Dot file on online



■ A web application for LOD visualization

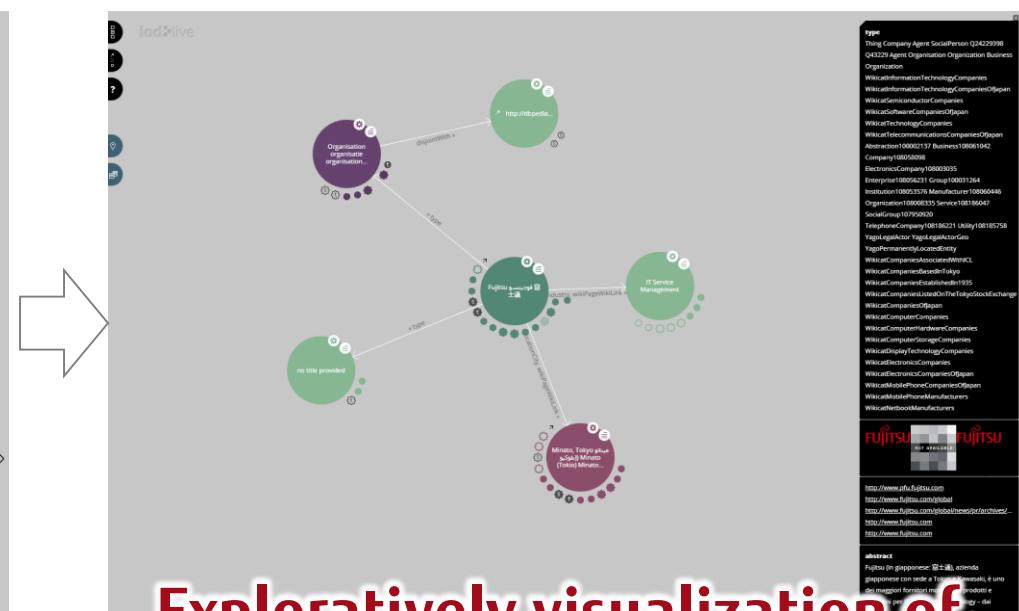
- It is developed with JavaScript and can be embedded in other applications

- Selecting a SPARQL endpoint
- Keyword search for interesting class

The LodLive interface consists of three main panels:

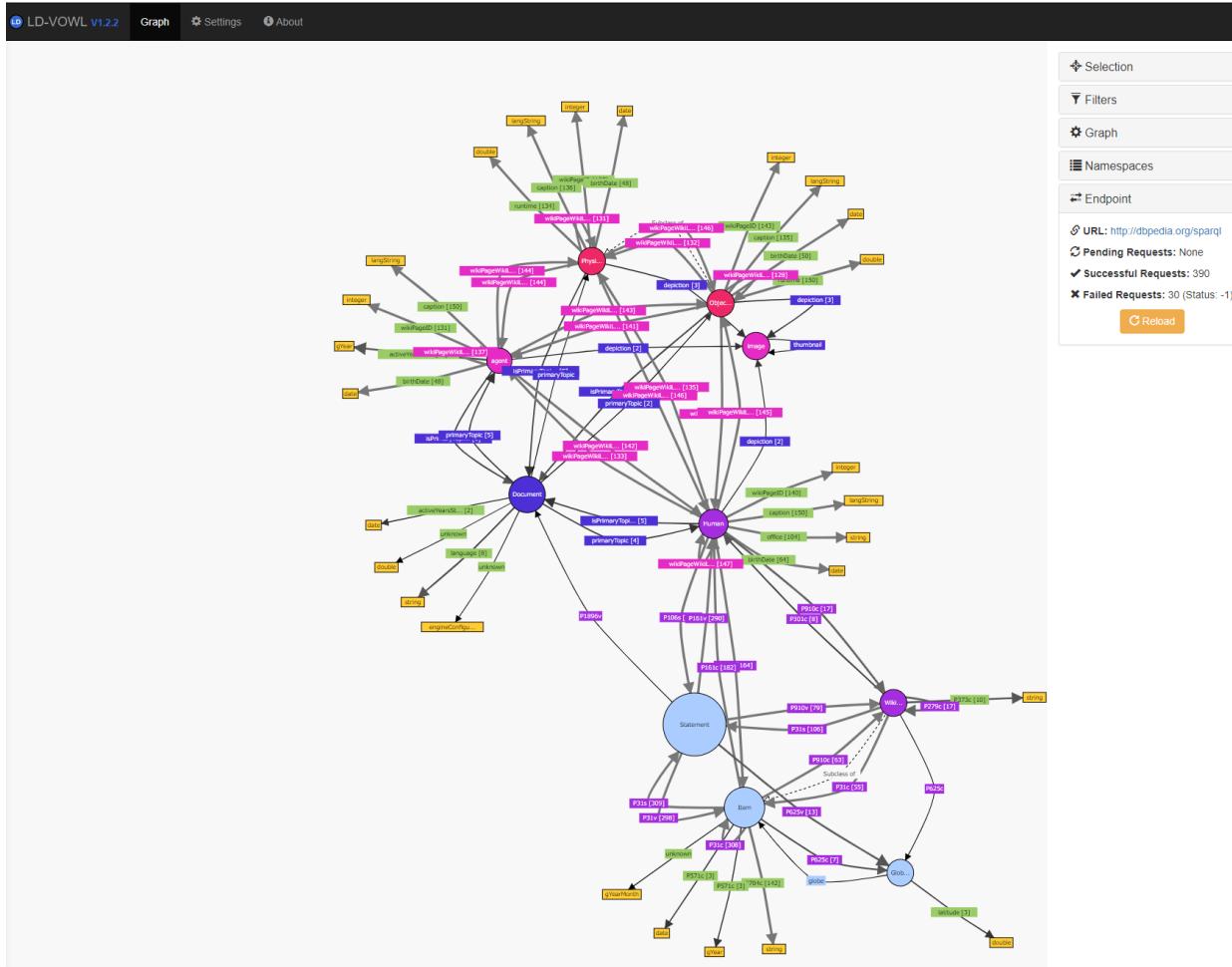
- simple search :** Includes a dropdown for "CHOOSE A DATASET" (set to dbpedia.org), a text input for "INSERT A KEYWORD" (set to Fujitsu), and a "start >" button.
- insert URI :** Includes a dropdown for "PASTE A RESOURCE ADDRESS" and a "start >" button.
- browse your data :** Includes a dropdown for "PASTE SOME RDF DATA" with a placeholder "coming soon" and a "start >" button.

A large red arrow points from the "simple search" panel towards the right side of the slide, indicating the flow of the application's functionality.



Exploratively visualization of the LOD from a selected class

■ Steamily extract the RDF data schemas from a specified SPARQL endpoint



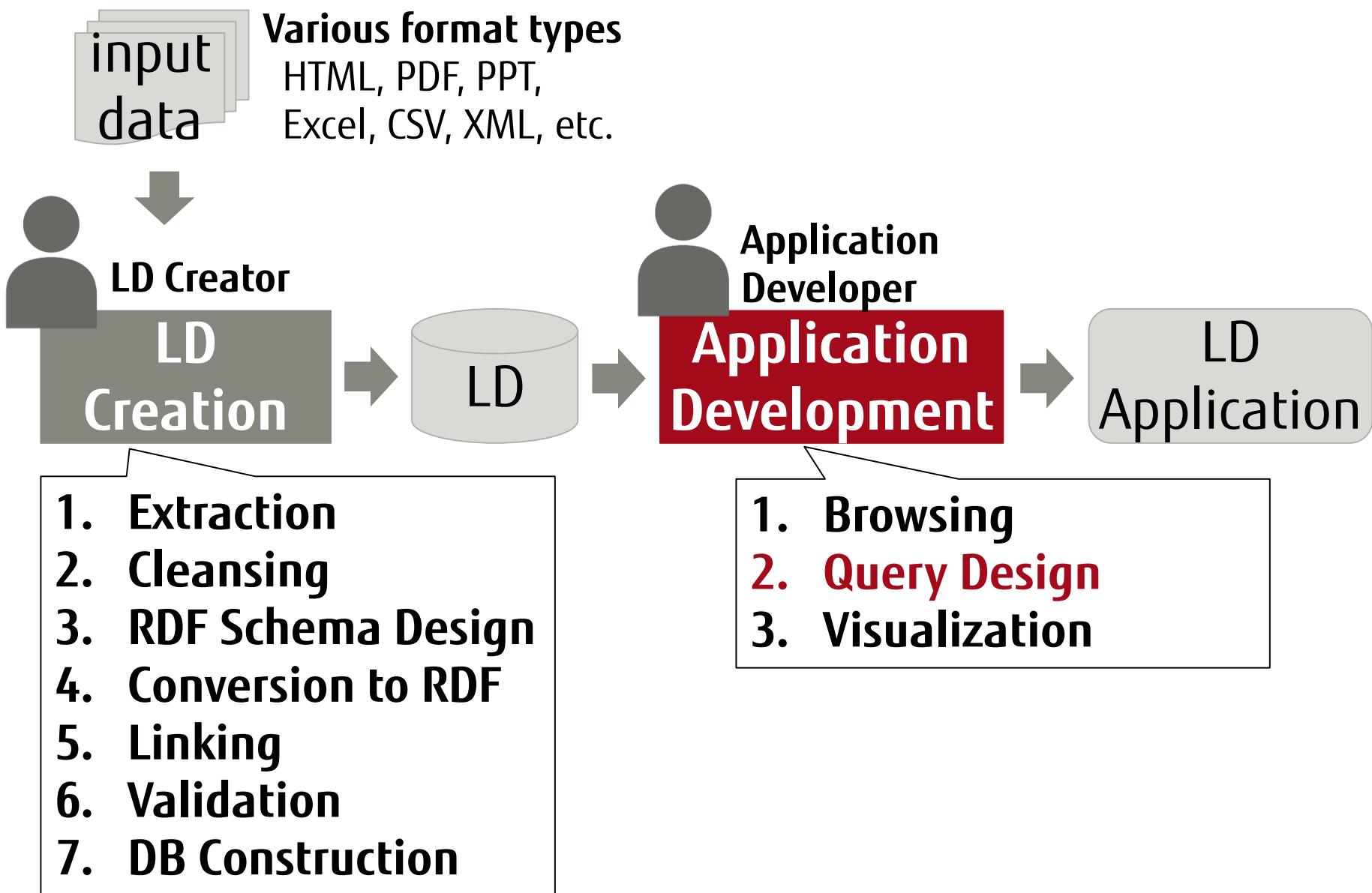
My Favorite Things

FUJITSU

Soft Shelled Turtle/すっぽん

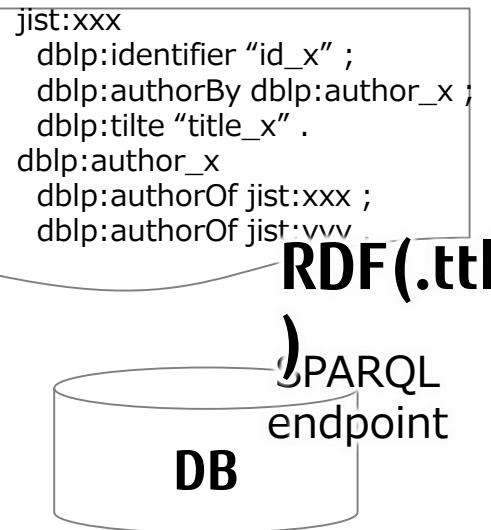


Outline



To design queries that search the necessary data for applications

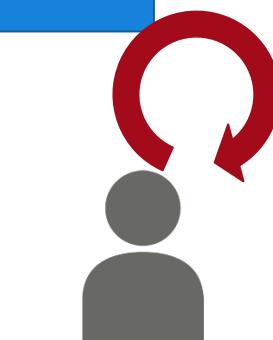
Ex. JIST data



the tools for requesting
SPARQL queries basically

Query Design

SPARQL Result Format



Application Developer

searching over and
over again for design
the queries

Query Design (cont.)

■ for SPARQL endpoints

- Command: [rsparql★ SPANG](#)
- WebUI: [sparql-kernel](#)
- library: [SPARQL Wrapper\(Python\)](#), [SPARQL\(R\)](#)
- library independence: [curl](#), [Excel](#)

■ for RDF files

- command: [arq](#)
- library: [rdflib](#)

■ Command-line tool to execute SPARQL for SPARQL endpoint

- One of Jena packages

■ Usage

- rsparql --service=<sparql_endpoint_uri> --query=<queryfile>

```
PREFIX dblp: <https://dblp.org/rdf/schema-2017-04-18#>
SELECT ?title WHERE {
    ?paper dblp:title ?title;
    dblp:publishedInBook ?book.
    FILTER(regex(?book, "JIST"))
}
```

**queryfile
(query.rq)**

```
$ rsparql --service http://[REDACTED] --query query.rq
```

Result

title
"G-Diff: A Grouping Algorithm for RDF Change Detection on MapReduce."
"A MapReduce-Based Approach for Prefix-Based Labeling of Large XML Data."
"Gathering Photos from Social Networks Using Semantic Technologies."
"Publishing Danish Agricultural Government Data as Semantic Web Data."

■ Command-line tool for executing queries to SPARQL endpoint

- Specify endpoint and prefix in advance

■ Usage

~/.spang/endpoints

```
dbpedia http://dbpedia.org/sparql  
jisttutorial http://xxxxxxxx/sparql
```

w/ queryfile

```
$ spang jisttutorial query.rq  
"G-Diff: A Grouping Algorithm for RDF Change Detection on MapReduce."  
"A MapReduce-Based Approach for Prefix-Based Labeling of Large XML Data."  
...
```

~/.spang/prefix

```
PREFIX dblp: <https://dblp.org/rdf/schema-2017-04-18#>
```

```
$ spang jisttutorial -P dblp:title | cut -f2  
"Integrated semantic model for complex disease network (Short Paper)"  
"Implementing LOD Surfer as a Search System for the Annotation of  
Multiple Protein Sequence Alignment (Short Paper)"  
...
```

w/o queryfile

■ A Jupyter kernel for SPARQL

The screenshot shows a Jupyter notebook interface with the title "jupyter JIST_Tutorial Last Checkpoint: 2時間前 (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. Below the menu is a toolbar with icons for file operations, run cell, and code cell type selection.

In [1]: %endpoint `http://XXXXXXXXXXXXXXXXXXXXXX/sparql`

Endpoint set to: http://XXXXXXXXXXXXXXXXXXXXXX /sparql

In [2]: `SELECT distinct ?id ?title ?creator_name {
 ?paper <http://purl.org/dc/elements/1.1/creator> ?creator .
 ?creator <http://www.w3.org/1999/02/22-rdf-syntax-ns#label> ?creator_name .
 ?paper <http://purl.org/dc/elements/1.1/title> ?title .
 ?paper <http://purl.org/dc/elements/1.1/identifier> ?id .
} LIMIT 3`

Total: 3, Shown: 3

id	title	creator_name
14	Publication of Statistical Linked Open data in Japan	Akie Mizutani
14	Publication of Statistical Linked Open data in Japan	Dan Yamamoto
14	Publication of Statistical Linked Open data in Japan	Fumihiro Kato

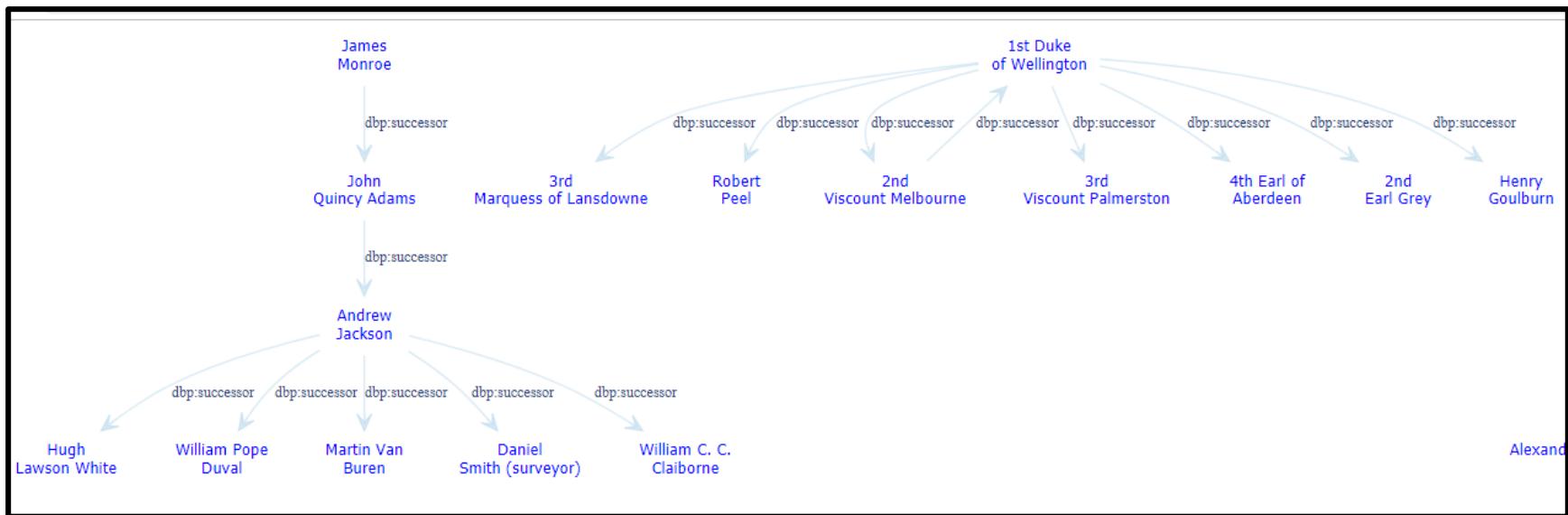
SPARQL kernel (cont.)

Query Design (3/9)

<https://github.com/paulovn/sparql-kernel>

FUJITSU

- It can output SVG in conjunction with GraphViz.
 - GraphViz must be installed on the same machine.



■ SPARQL endpoint interface for Python library

- converts the result into various formats

Installation

```
$ pip install SPARQLWrapper
```

Code

```
from SPARQLWrapper import SPARQLWrapper, JSON

sparql = SPARQLWrapper("http://*****/sparql")

sparql.setQuery("""
    select * where {
        ?paper dc:creator ?creator .
    } limit 3
""")
sparql.setReturnFormat(JSON)
print(sparql.query().convert())
```

- R also has a SPARQL library

```
> library(SPARQL)
> endpoint <- http://*****/sparql'
> q <- "PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT * WHERE {
    ?paper dc:creator ?creator .
> res <- SPARQL(endpoint, q)$results
```

■ Command-line tool for transferring data

```
$ curl -G -H 'Accept: application/sparql-results+json' --data-urlencode 'query@-'  
http://*****/sparql < query_curl.rq
```

```
{ "head": { "link": [], "vars": ["creator"] },  
  "results": { "distinct": false, "ordered": true, "bindings": [  
    { "creator": { "type": "uri", "value": "https://lod4all.github.io/Akie+Mizutani" }},  
    { "creator": { "type": "uri", "value": "https://lod4all.github.io/Dan+Yamamoto" }},  
    { "creator": { "type": "uri", "value": "https://lod4all.github.io/Fumihiro+Kato" }},  
    { "creator": { "type": "uri", "value": "https://lod4all.github.io/Hideaki+Takeda" }},  
    { "creator": { "type": "uri", "value":  
      "https://lod4all.github.io/Hiromu+Harada" } } ] }
```

■ Use the WEBSERVICE function

Screenshot of Microsoft Excel showing the use of the WEBSERVICE function to query data from a SPARQL endpoint.

The formula in cell B3 is:

```
=FILTERXML(WEBSERVICE("http://***** /sparql?default-graph-uri=&query=select+distinct+%3Ftitle+where+%7B%0D%0A%3Fpaper+dc%3Acreator%2Frdf%3Alabel+"""+$A3+""""+.%0D%0A%3Fpaper+dc%3Atitle+%3Ftitle%0D%0A%7D&format=application%2Fsparql-results%2Bxml&timeout=0&debug="), "/*")
```

	A	B
1	author	title
2	Biralatei J. Fawei	A Criminal Law and Procedure Ontology for Legal Question Answering
3	Adam Wyner	Knowledge Driven Intelligent Survey SystemS for Linguists
4	Adam Z. Wyner	A Criminal Law and Procedure Ontology for Legal Question Answering
5	Ahmet Soylu	Making Complex Ontologies End-User Accessible via Ontology
6	Akane Takezaki	Building the Core Vocabulary of Crop Names to Integrate the Vocabularies by Different Government Agencies
7	Akio Mizutani	Publication of Statistical Linked Open data in Japan

Cells B3 and B7 are highlighted with green borders.

■ Command line tool to execute SPARQL for RDF dump

- One of Jena packages

■ Usage

- arq --data=<file> --query=<queryfile>
- arq --data=<file> '<query>'

query.rq

```
SELECT distinct ?id ?title ?creator_name {  
    ?paper <http://purl.org/dc/elements/1.1/creator> ?creator .  
    ?creator <http://www.w3.org/1999/02/22-rdf-syntax-ns#label> ?creator_name .  
    ?paper <http://purl.org/dc/elements/1.1/title> ?title .  
    ?paper <http://purl.org/dc/elements/1.1/identifier> ?id .  
} LIMIT 3
```

```
$ arq --data jist_ap-json.ttl --query query.rq
```

id title creator_name		
-----	-----	-----
"48" "Construction and Reuse of Linked Agriculture Data: An Experience of Taiwan Government Open Data" "Dongpo Deng"		
"26" "IMI: A Common Vocabulary Framework for Open Government Data" "Shuichi Tashiro"		
"91" "On Enhancing Visual Query Building Over KGs Using Query Logs (Short Paper)" "Ahmet Soylu"		

■ Python library for RDF dump

- specifies resource URIs or RDF files
- used in SPARQL Wrapper library

Installation

```
$ pip install rdflib
```

Code

```
import rdflib
graph=rdflib.Graph()
graph.load("https://*****/jist_ap-json.ttl",format="turtle")
for s,p,o in graph:
    print(s,p,o)
```

The diagram illustrates the execution flow of the provided Python code. A red speech bubble labeled "RDF file" points to the line `graph.load("https://*****/jist_ap-json.ttl",format="turtle")`. A red rounded rectangle labeled "python collection" points to the line `for s,p,o in graph:`.

Result

```
https://lod4all.github.io/43 http://purl.org/dc/elements/1.1/creator https://lod4all.github.io/Pokpong+Songmuang
https://lod4all.github.io/30 http://purl.org/dc/elements/1.1/creator https://lod4all.github.io/Songmao+Zhang
https://lod4all.github.io/Khai+Nguyen http://www.w3.org/1999/02/22-rdf-syntax-ns#label Khai Nguyen
```

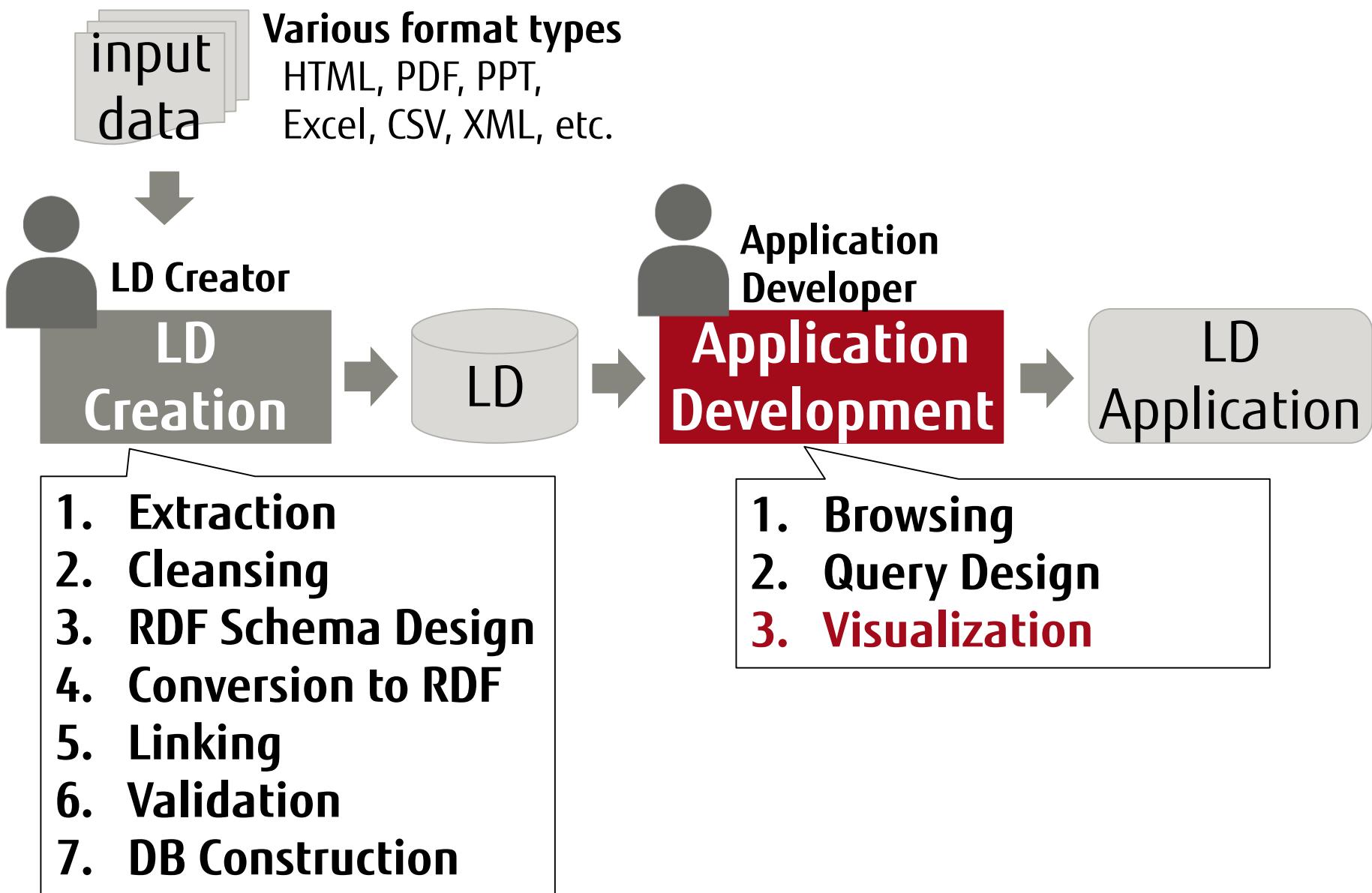
My Favorite Things

FUJITSU

Paragliding near Mt.Fuji



Outline

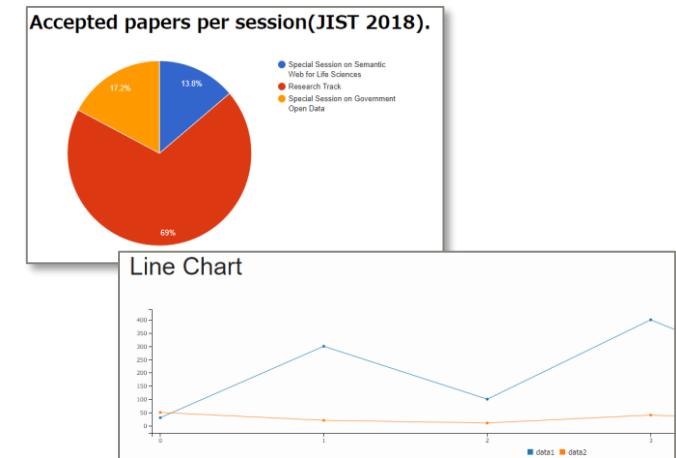


Visualizing the results of SPARQL queries using visualization libraries

SPARQL Result Format



Visualization



- Specialized in SPARQL
 - [d3sparql★](#), [Sgvizler](#)
- Standard
 - [D3.js★](#), [C3.js](#)
- Wikification
 - [TAGME★](#), [DBpedia Spotlight](#), [Illinois Wikifier](#)

■ A JavaScript library based on the standard visualization library D3.js.

- Executing SPARQL query
- Transforming the resulting JSON for visualization in D3.js layouts
- Major D3.js layouts
 - Chart: Bar, Pie, Graph: Force, Sankey
 - Tree: Round tree, Tree, map
 - Map, Table

SPARQL query

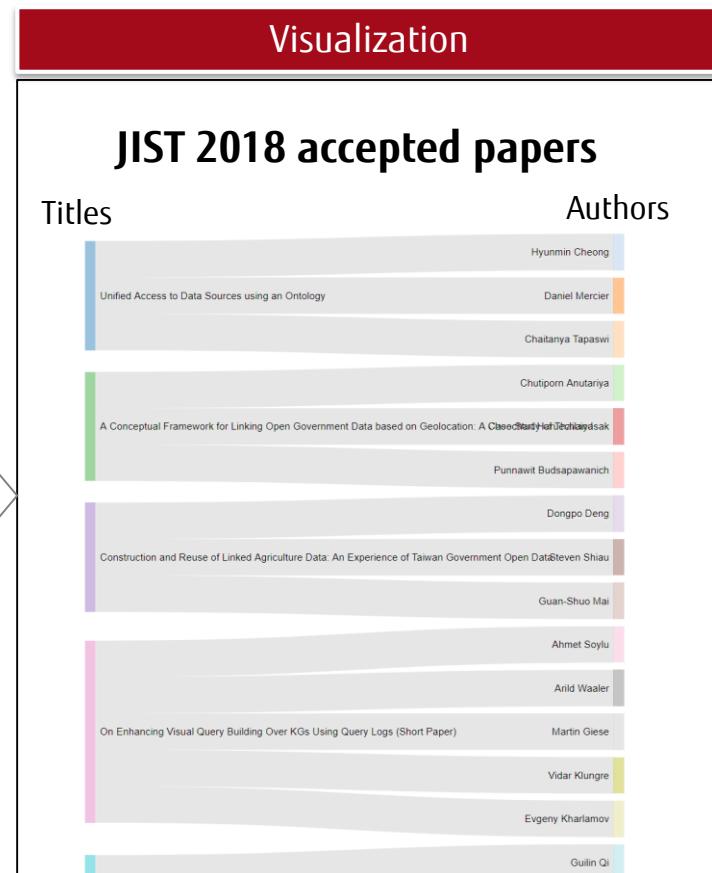
```
SELECT distinct ?paper
  ?paper_title
  ?author
  ?author_name
  WHERE{
    ?paper dc:title ?title;
           dc:creator ?author.
    ?author rdfs:label ?author_name.
  }
```

Searching for JIST 2018 paper titles and authors

Visualization definition

```
function render(json) {
  var config = {
    // for d3sparql.graph()
    "key1": "paper",
    "key2": "authour",
    "label1": "paper_title",
    "label2": "authour_name",
    // for d3sparql.sankey()
    "width": 750,
    "height": 1200,
    "margin": 10,
    "selector": "#result"
  }
  d3sparql.sankey(json, config)
}
```

Specifying the output layout as a *Sankey* graph



- A JavaScript library Rendering the result of SPARQL SELECT queries into charts or HTML elements

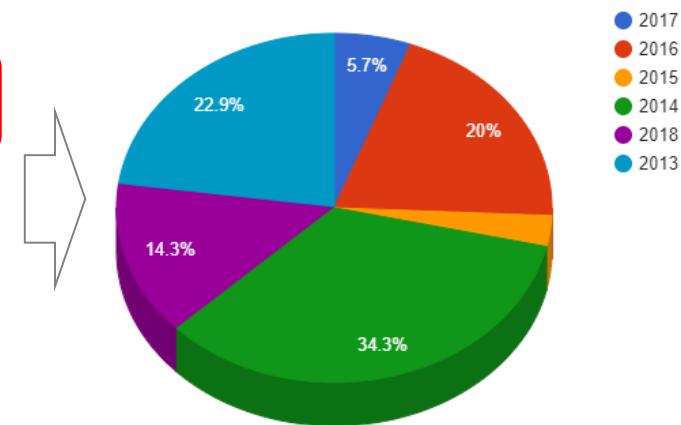
```
<html>
  <head>
    <title>MWE</title>
    <script type="text/javascript" src="//cdnjs.cloudflare.com/ajax/libs/jquery/1.9.0/jquery.min.js"></script>
    <script type="text/javascript" src="https://raw.githubusercontent.com/mgskjaeveland/sgvizler/v0.6/sgvizler.js"></script>
    <script>
      $(document).ready(
        function (){ sgvizler.containerDrawAll();
      });
    </script>
  </head>
  <body>

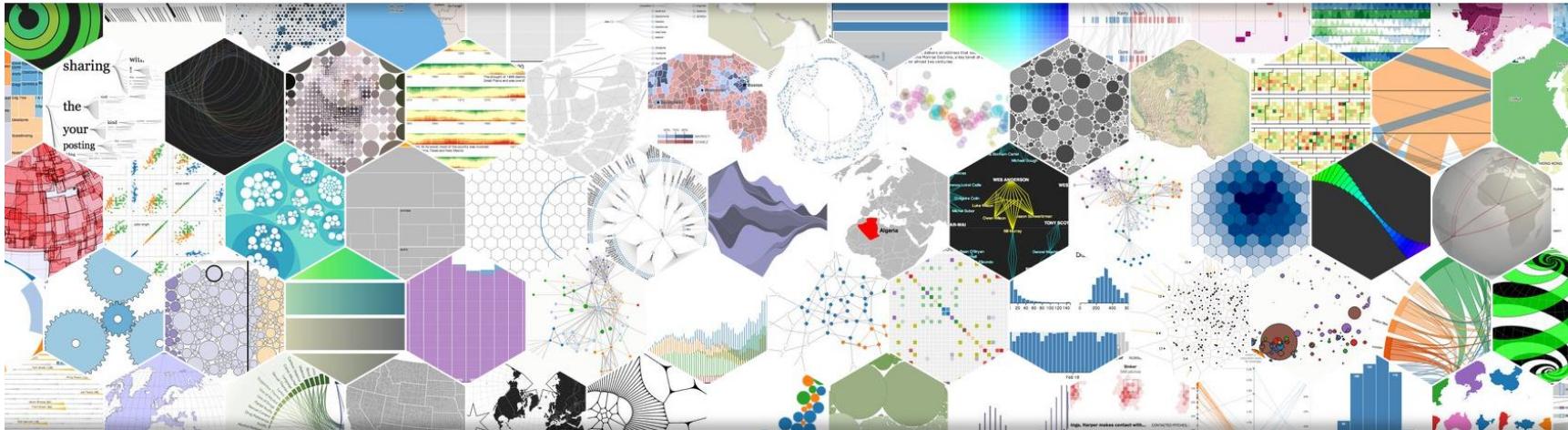
    <h1>Accepted papers per session(JIST 2017)</h1>
    <p></p>
    <div id="example">
      <pre>
        data-sgvizler-endpoint="http://localhost:3030/jist/query"
        data-sgvizler-query="PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
          PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
          SELECT ?subject (count(?paper) as ?c)
          WHERE {
            ?paper <http://purl.org/dc/elements/1.1/subject> ?subject .
          }
          group by ?subject"
        data-sgvizler-chart="google.visualization.PieChart"
        style="width:800px; height:400px;"></div>
    </body>
</html>
```

SPARQL endpoint

SPARQL query

Visualization definition:
google.visualization.PicChart



 Data-Driven Documents

- A JavaScript library for producing dynamic, interactive data visualizations in web browsers
 - Latest version: 5.7.0
 - It helps you bring data to life using HTML, SVG, and CSS

```
chart = {
  const root = pack(data);

  const svg = d3.select(DOM.svg(width, height))
    .style("width", "100%")
    .style("height", "auto")
    .attr("font-size", 10)
    .attr("font-family", "sans-serif")
    .attr("text-anchor", "middle");

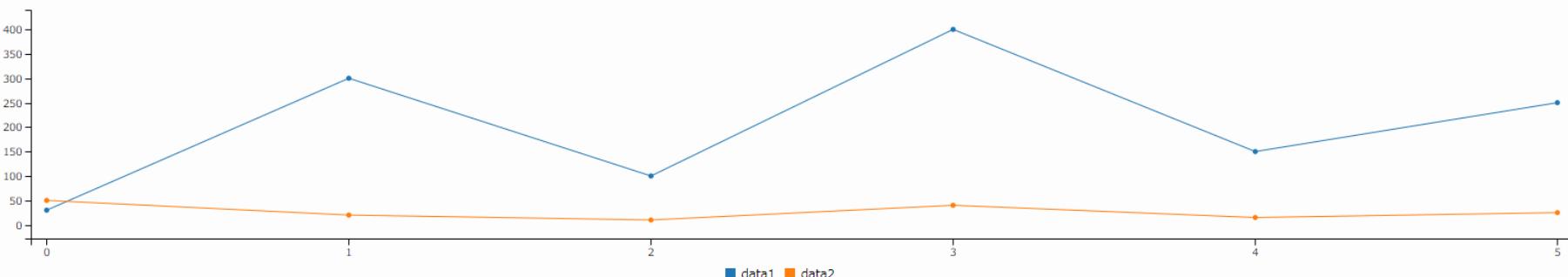
  const leaf = svg.selectAll("g")
    .data(root.leaves())
    .enter().append("g")
    .attr("transform", d => `translate(${d.x + 1},${d.y + 1})`);

  leaf.append("circle")
    .attr("id", d => (d.leafUid = DOM.uid("leaf")).id)
    .attr("r", d => d.r)
    .attr("fill-opacity", 0.7)
    .attr("fill", d => color(d.data.group));

  leaf.append("clipPath")
    .attr("id", d => (d.clipUid = DOM.uid("clip")).id)
    .append("use")
    .attr("xlink:href", d => d.leafUid.href);
}
```

- Makes it easy to generate D3-based charts by wrapping the code required to construct the entire chart. We don't need to write D3 code any more.

Line Chart



```
# simple_multiple.js
```

```
var chart = c3.generate({
  data: {
    columns: [
      ['data1', 30, 300, 100, 400, 150, 250],
      ['data2', 50, 20, 10, 40, 15, 25]
    ]
  }
});
```

■ Implementation of the Wikification systems

- available in English, German and in Italian and it is based on Wikipedia snapshots of April, 2016.

Example

In computing, linked data (often capitalized as Linked Data) is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries.^[1] It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the internet to become a global database.

The screenshot shows the TAGME web application interface. At the top, there's a logo and a brief description of what TAGME does: identifying meaningful short-phrases in unstructured text and linking them to Wikipedia pages. Below this, a message says "Try TAGME now!" and provides documentation links. A "NEWS! TAGME is now hosted by the D4Science infrastructure. Check the RESTful API page for details." A developer note credits Paolo Ferragina and Ugo Scialo at A3 Lab, Dipartimento di Informatica, University of Pisa. The main area has an "Input Text" field with language selection buttons for Italian, English, and Deutsche. The text "Deep Learning is one of the most representative technologies in recent AI and shows high performance in pattern recognition and analysis. However, as it cannot explain the reasons for its judgment, it is called "black box AI." Due to this restriction, it is difficult to apply AI to the fields requiring high reliability and persuasiveness such as healthcare, finance, and corporate management that especially need important decision-making." is entered. To the right, there are two vertical sliders: "Many links" (higher) and "Few links" (lower). Below the text, a "Reset" button and a "TAGME!" button are visible. A smaller window below shows the same text with blue underlined links, indicating where the system has found "links".

■ Ex. Annotated titles

- Ontology-based Semantic Representation of Spatial Configuration: Shape Grammar for [Caravanserai](#) Architectural Heritage
- DeFind: A Protege Plugin for Computing Concept Definitions in EL [Ontologies](#)

DBpedia Spotlight

Visualization (6/7)

<https://www.dbpedia-spotlight.org/demo/index.html>

FUJITSU

- A tool for automatically annotating mentions of DBpedia resources in text

Example

In computing, [linked data](#) (often capitalized as [Linked Data](#)) is a method of publishing structured data so that it can be interlinked and become more useful through [semantic queries](#).^[1] It builds upon standard Web technologies such as [HTTP](#), [RDF](#) and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of [linked data](#) is for the internet to become a global [database](#).

The screenshot shows the DBpedia Spotlight demo interface. At the top, there's a logo with the text "DBpedia Spotlight". Below it is a form with a "Confidence:" slider set to 0.5, a "Language:" dropdown set to "English", and two buttons: "SELECT TYPES..." and "ANNOTATE". There's also a checkbox for "n-best candidates" which is unchecked. The main area contains the input text from the previous slide, followed by annotated results. The annotated text includes links like [HTTP](#), [RDF](#), [semantic queries](#), [W3C](#), [Semantic Web](#), [linked open data \(LOD\)](#), and [database](#). Below the input text, there's a note about Tim Berners-Lee and the Semantic Web project, and another note about [open data](#). At the bottom, there's a "BACK TO TEXT" button, a "How to cite this work" link, and a "You should know:" section with some tips.

■ also makes links with Wikipedia

Example

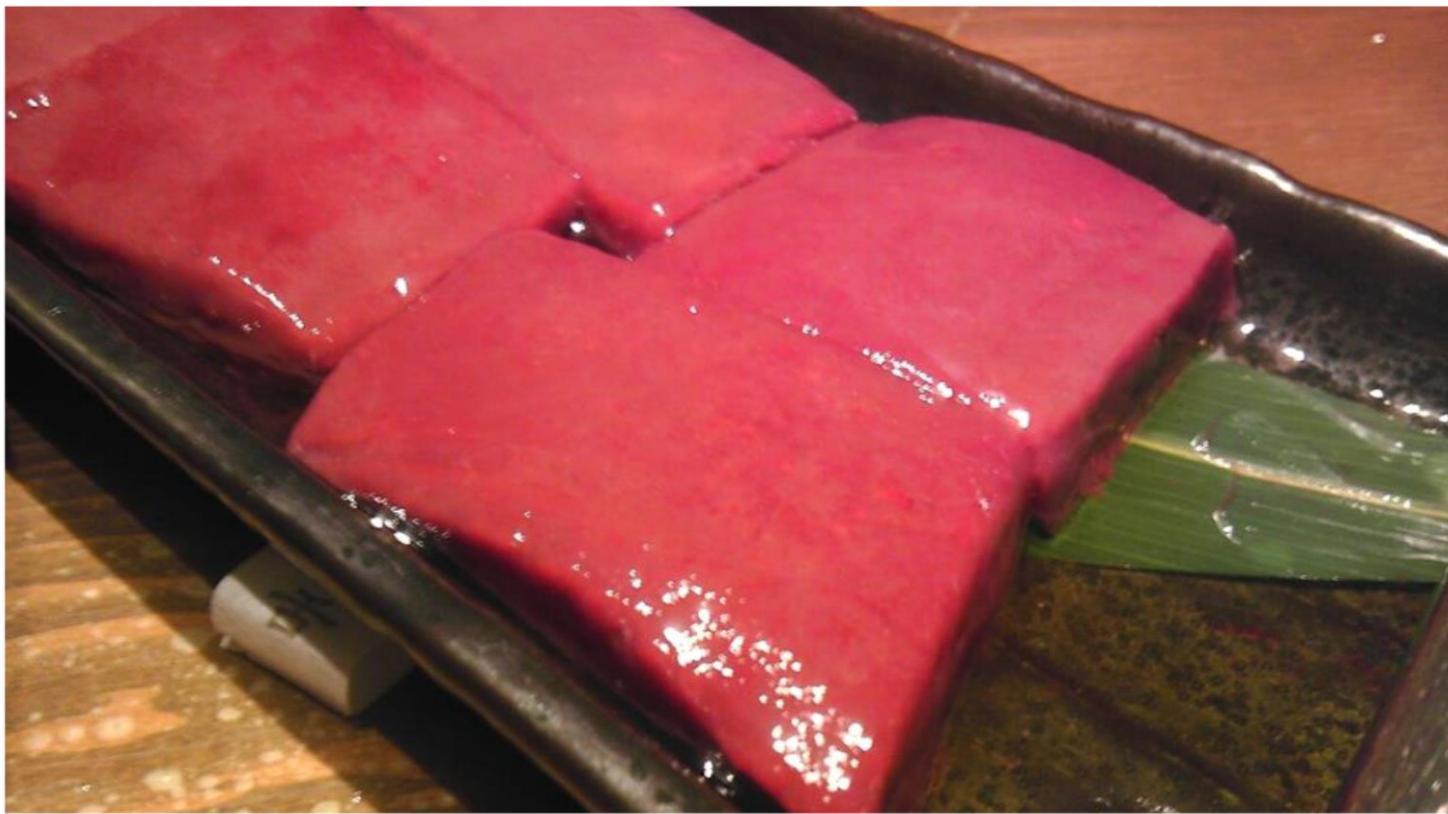
In computing, linked data (often capitalized as Linked Data) is a method of publishing structured data so that it can be interlinked and become more useful through semantic **queries**.^[1] It builds upon standard Web technologies such as [HTTP](#), [RDF](#) and [URIs](#), but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the internet to become a global database.

The screenshot shows a web page titled "COGNITIVE COMPUTATION GROUP" with a sub-section for "Wikifier Demo". The page includes a navigation bar with links to Research, People, Software, Demos, Publications, Resources, and Schedule. Below the navigation is a search bar with placeholder text "Please enter a passage below to be Wikified". A large text area contains the explanatory text about linked data. At the bottom of this area, there is a note about Tim Berners-Lee and the Semantic Web project. A green callout box highlights a section where the system has identified entities with Wikipedia articles. The footer of the page includes links to Home, Demos, and Wikifier, along with contact information and a note about restarting the demo.

My Favorite Things

FUJITSU

Liver Sashimi



Final Remarks

- Let's Review Through The Sample Application

JIST LD Application

FUJITSU

Specifying conference name

COMMON KEYWORDS EXTRACTED FROM PAPERS TITLES

localhost

Search

Common keywords extracted from paper titles

Checking the trends of the keywords by clicking directly.

TRENDS OF CLICKING KEYWORDS

Trend of users clicking the keyword

Year	rdf	ontology	semantic web
2013	8	5	2
2014	12	5	4
2015	2	3	3
2016	5	5	5
2017	2	2	2
2018	5	1	1

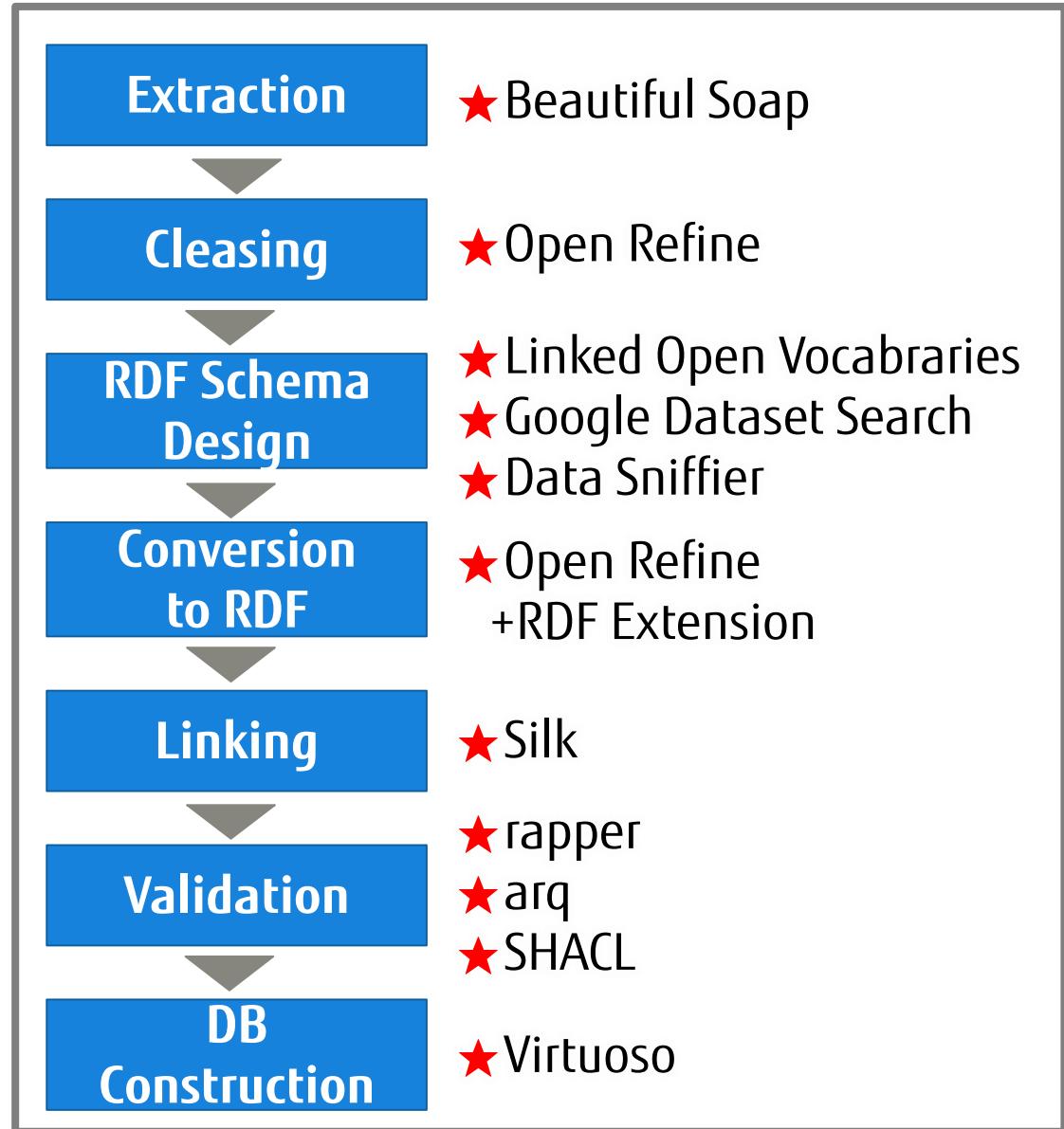
PAPER TITLES CONTAIN A SELECTED KEYWORD

Paper titles containing the keyword

- Building Wikipedia Ontology with More Semantically Enriched Information Resources
- Ontology Construction for Specialized Books
- SQuaRE: A Visual Ontology Editor
- Development of Faceted and Synonym Search Support for the Ontology Application Management Framework
- Development of the Belief Culture Ontology and Its Application: Case Study of the GreaterMekong Subregion
- Automatic and Dynamic Book Category Assignment Using Concept-Based Book Ontology
- Ontology Based Suggestion Distribution System
- Constructing City Ontology from Expert for Smart City Management
- Ontology Construction Support for Specialized Books
- A Graph-Based Approach to Ontology Debugging in DL-Lite
- Utilizing Weighted Ontology Mappings on Federated SPARQL Querying
- A Framework of Automatic Alignment of Concept in Ontology with Confidence Score based on Inner Concept Information
- MAPSOM: User Involvement in Ontology Matching
- Designing of Ontology for Domain Vocabulary on Agriculture Ontology (AAO) and a Lesson Learned
- An Information Literacy Ontology and Its Use for Guidance Plan Design - An Example on Problem Solving
- Ontology Schema-specific Rule Materialization
- Ontology Refinement System for Improving Consistency of Classification among Brother Concepts
- Enriching Heterogeneous Biological Resources on the Web: A Case Study on TDP Channel Ontology Construction

JIST LD Creation

FUJITSU



RDF(.ttl)

```
jist:xxx
dblp:identifier "id_x" ;
dblp:authorBy dblp:author_x ;
dblp:title "title_x" .
```

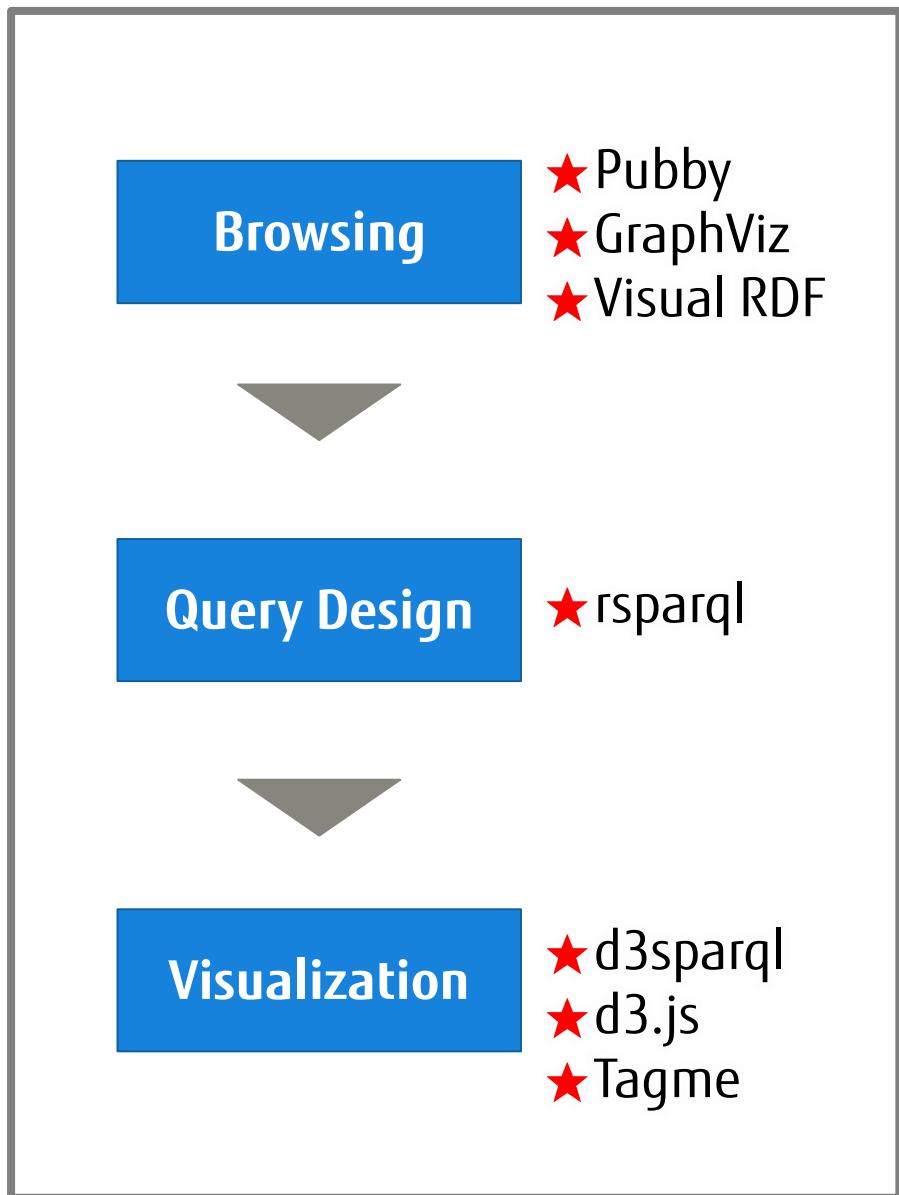
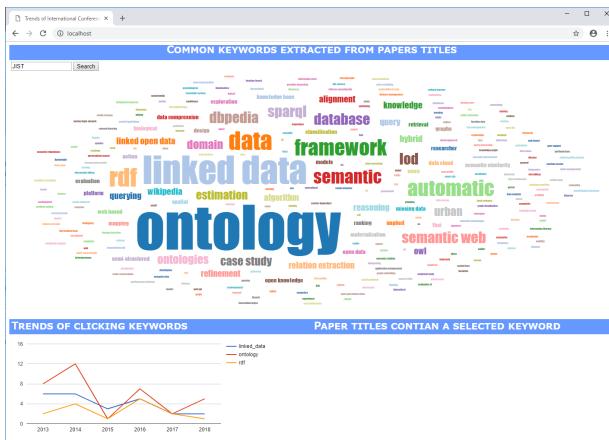
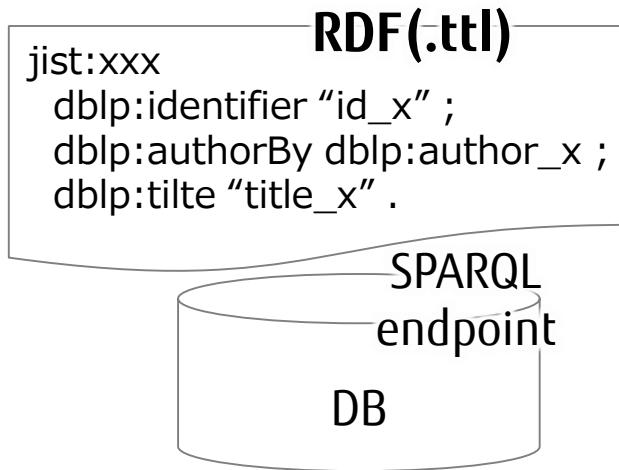
SPARQL endpoint

DB



JIST LD Application Development

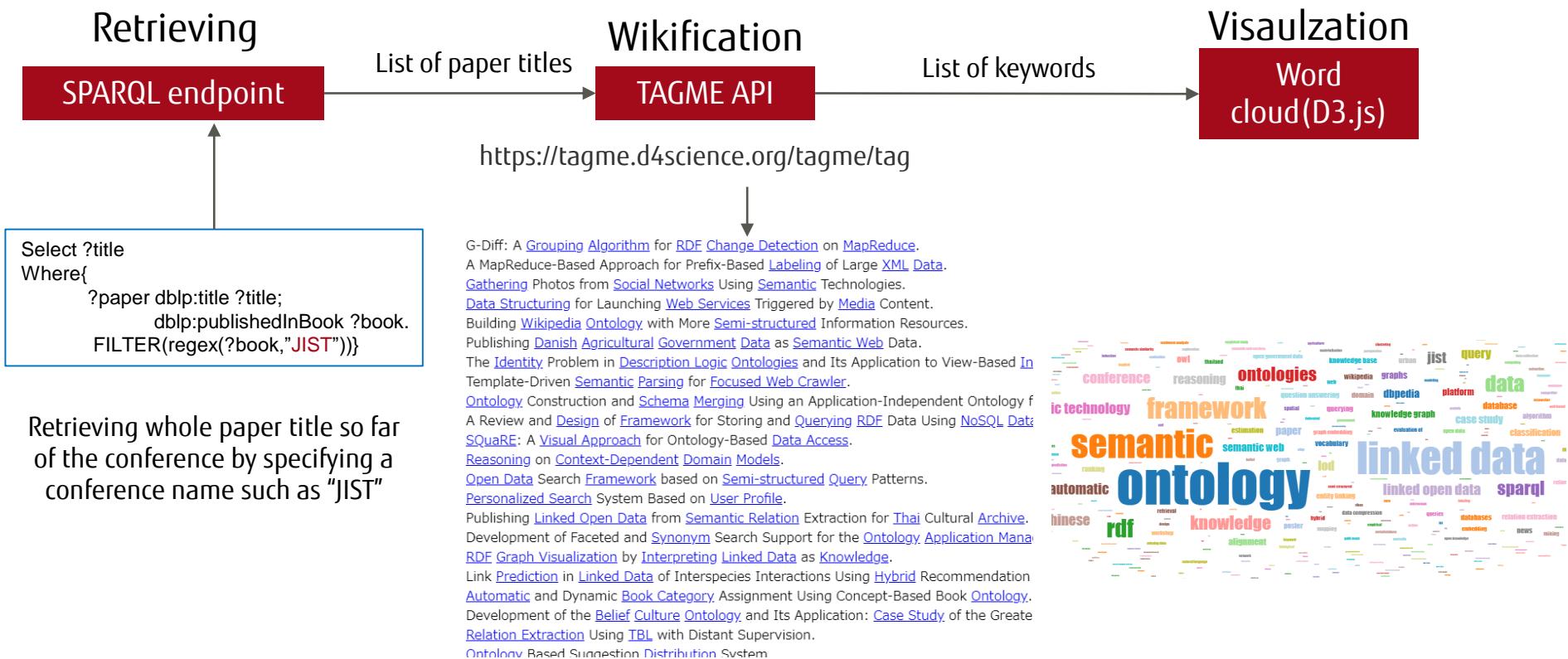
FUJITSU



JIST LD Application : Detail

FUJITSU

TAGME + D3.js

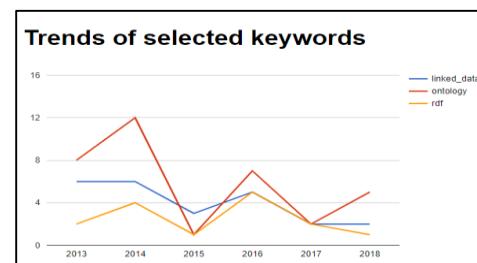
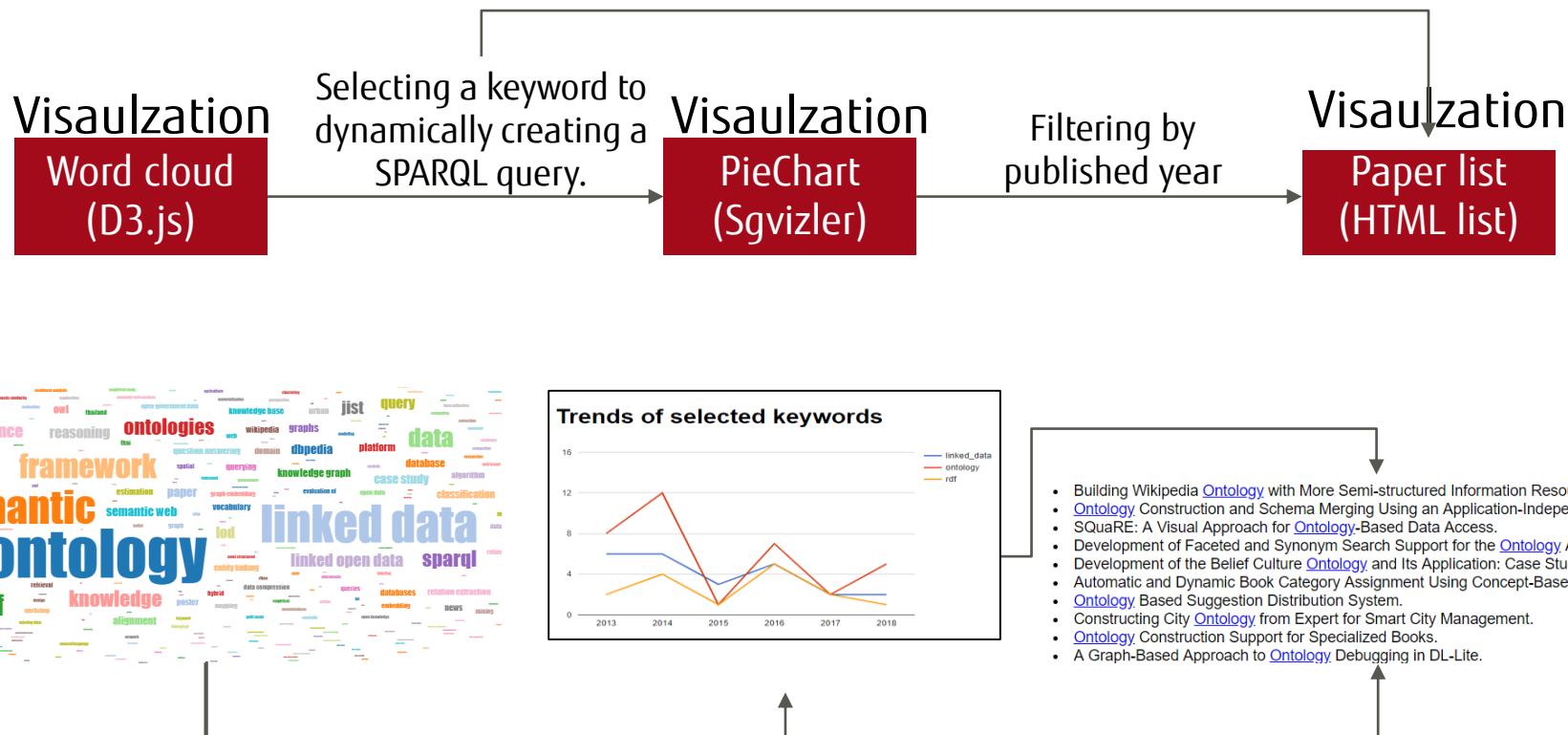


Retrieving whole paper title so far of the conference by specifying a conference name such as "JIST"

JIST LD Application : Detail (cont.)

FUJITSU

D3.js + Sgvizler



- Building Wikipedia [Ontology](#) with More Semi-structured Information Resources.
 - [Ontology](#) Construction and Schema Merging Using an Application-Independent Ontology for I-
 - SQuaRE: A Visual Approach for [Ontology](#)-Based Data Access.
 - Development of Faceted and Synonym Search Support for the [Ontology](#) Application Manager.
 - Development of the Belief Culture [Ontology](#) and Its Application: Case Study of the GreaterMel.
 - Automatic and Dynamic Book Category Assignment Using Concept-Based Book [Ontology](#).
 - [Ontology](#) Based Suggestion Distribution System.
 - Constructing City [Ontology](#) from Expert for Smart City Management.
 - [Ontology](#) Construction Support for Specialized Books.
 - A Graph-Based Approach to [Ontology](#) Debugging in DL-Lite.

Introduction : Schedule

Morning	Introduction to Linked Data tools	9:00-10:00 (60 mins)	Presentation
		10:00-10:10 (10 mins)	Break
		10:10-10:50 (40 mins)	Presentation
		10:50-11:10 (20 mins)	Use Case Studies
Afternoon	Hands-on	13:00-13:10 (10 mins)	Introduction
		13:10-14:55 (95 mins)	Trial 1: Let's visualize corporate data Trial 2: Let's develop a financial analysis application
		14:55-15:00 (5 mins)	Closing

FUJITSU

shaping tomorrow with you

Glossary

- cURLA: command line Open Source/Free Software client that can transfer data, including machine readable RDF, from or to a server using one of its many supported protocols.
- DBLP:DBLP is a computer science bibliography website. Starting in 1993 at the University of Trier, Germany, it grew from a small collection of HTML files[1] and became an organization hosting a database and logic programming bibliography site. DBLP listed more than 3.66 million journal articles, conference papers, and other publications on computer science in July 2016, up from about 14,000 in 1995.[2] All important journals on computer science are tracked. Proceedings papers of many conferences are also tracked. It is mirrored at three sites across the Internet
- Dbpedia: DBpedia is a community effort to extract structured information from Wikipedia and make it available on the Web. DBpedia is often depicted as a hub for the Data Cloud. An RDF representation of the metadata held in Wikipedia and made available for SPARQL query on the World Wide Web.
- Dublin Core Metadata Element Set : Dublin Core Metadata Element Set refers to a vocabulary of fifteen properties for use in resource descriptions, such as may be found in a library card catalog (creator, publisher, etc). The Dublin Core Metadata Element Set, also known as "DC Elements", is the most commonly used vocabulary for Linked Data applications. See also Dublin Core Element Set, Version 1.1 Specification. [DCMI]
- Dublin Core Metadata Terms : A vocabulary of bibliographic terms used to describe both physical publications and those on the Web. An extended set of terms beyond those basic terms found in the Dublin Core Metadata Element Set. See also Dublin Core Metadata Terms
- Entity : In the sense of an entity-attribute-value model, an entity is synonymous with the Subject of an RDF Triple.
- FOAF (Friend of a Friend) A Semantic Web vocabulary describing people and their relationships for use in resource descriptions.
- GitHub: **GitHub Inc.** is a web-based [hosting service](#) for [version control](#) using [Git](#). It is mostly used for [computer code](#). It offers all of the [distributed version control](#) and [source code management](#) (SCM) functionality of Git as well as adding its own features. It provides [access control](#) and several collaboration features such as [bug tracking](#), [feature requests](#), [task management](#), and [wikis](#) for every project

Glossary (cont.)

- **International Semantic Web Conference (ISWC)**: The International Semantic Web Conference (ISWC) is a series of academic conferences and the premier international forum, for the Semantic Web, Linked Data and Knowledge Graph Community.[1][2][3] Here, scientists, industry specialists, and practitioners meet to discuss the future of practical, scalable, user-friendly, and game changing solutions.[4] Its proceedings are published in the Lecture Notes in Computer Science by Springer-Verlag.
- **JSON** : JavaScript Object Notation (JSON) is syntax for storing and exchanging text based information. JSON has proven to be a highly useful and popular object serialization and messaging format for the Web. See also: the application/json Media Type for JavaScript Object Notation (JSON) [RFC4627].
- **Linked Data (LD)**: A pattern for hyperlinking machine-readable data sets to each other using Semantic Web techniques, especially via the use of RDF and URIs. Enables distributed SPARQL queries of the data sets and a browsing or discovery approach to finding information (as compared to a search strategy). Linked Data is intended for access by both humans and machines. Linked Data uses the RDF family of standards for data interchange (e.g., RDF/XML, RDFa, Turtle) and query (SPARQL). If Linked Data is published on the public Web, it is generally called [Linked Open Data](#). See also [[Linked Data Principles](#)].
- **Linked Open Data (LOD)**: Linked Data published on the public Web and licensed under one of several open licenses permitting reuse. Publishing Linked Open Data enables distributed SPARQL queries of the data sets and a "browsing" or "discovery" approach to finding information, as compared to a search strategy. See also: "Linked Data: Structured Data on the Web" [LD-FOR-DEVELOPERS] and "Linked Data: Evolving the Web into a Global Data Space" [HOWTO-LODP]
- **Ontology**: A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms).

Glossary (cont.)

- **Web Ontology Language(OWL):** The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. Ontologies are a formal way to describe taxonomies and classification networks, essentially defining the structure of knowledge for various domains: the nouns representing classes of objects and the verbs representing relations between the objects. Ontologies resemble class hierarchies in object-oriented programming but there are several critical differences. Class hierarchies are meant to represent structures used in source code that evolve fairly slowly (typically monthly revisions) whereas ontologies are meant to represent information on the Internet and are expected to be evolving almost constantly. Similarly, ontologies are typically far more flexible as they are meant to represent information on the Internet coming from all sorts of heterogeneous data sources. Class hierarchies on the other hand are meant to be fairly static and rely on far less diverse and more structured sources of data such as corporate databases
- **Python:** Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.[27] In July 2018, Van Rossum stepped down as the leader in the language community after 30 years
- **Resource Description Framework (RDF):** A family of international standards for data interchange on the Web produced by W3C. Resource Description Framework (RDF) is based on the idea of identifying things using Web identifiers or HTTP URLs, and describing resources in terms of simple properties and property values. See also [RDF 1.1 Concepts and Abstract Syntax]
- **RDF Schema(RDFS):**The simplest RDF vocabulary description language. It provides much less descriptive capability than the Simple Knowledge Organization System (SKOS) or the Web Ontology Language (OWL). A standard of the W3C [RDFS]
- **SPARQL:** SPARQL Protocol and RDF Query Language (SPARQL) defines a query language for RDF data, analogous to the Structured Query Language (SQL) for relational databases. A family of standards of the World Wide Web Consortium. See also SPARQL 1.1 Overview [SPARQL-11].

Glossary (cont.)

- SPARQL endpoint: A service that accepts SPARQL queries and returns answers to them as SPARQL result sets. It is a best practice for datasets providers to give the URL of their SPARQL endpoint to allow access to their data programmatically or through a Web interface. A list of some endpoints status is available at <http://labs.mondeca.com/sparqlEndpointsStatus/>
- **Semantic Web Rule Language (SWRL)** The **Semantic Web Rule Language (SWRL)** is a proposed language for the [Semantic Web](#) that can be used to express rules as well as logic, combining [OWL DL](#) or OWL Lite with a subset of the [Rule Markup Language](#) (itself a subset of [Datalog](#)).
- YAML: YAML (YAML Ain't Markup Language) is a human-readable data serialization language. It is commonly used for configuration files, but could be used in many applications where data is being stored (e.g. debugging output) or transmitted (e.g. document headers). YAML targets many of the same communications applications as XML but has a minimal syntax which intentionally breaks compatibility with SGML.[1] It uses both Python-style indentation to indicate nesting, and a more compact format that uses [] for lists and {} for maps[1] making YAML 1.2 a superset of JSON.[2]

References

- Wikipedia: <https://en.Wikipedia.org/>
- Linked Data Glossary: <https://www.w3.org/TR/ld-glossary/>