

Publishing the Paul Kendel Fonoroff Collection for Chinese Film Studies as Linked Data: An Architecture of Distributed Data Services

Haiqing Lin

C.V. Starr East Asian Library

University of California, Berkeley

PRRLA 2019 MEETING

Korea University

September 1-4, 2019.

Outlines

- Background
- Why linked data? An architecture
- Design aspects of ontology, defined classes and singleton properties
- Query and services, SPARQL endpoint, relationships and data integration
- Challenges

Background

In 2016, UC Berkeley C.V. Starr East Asian Library has acquired a significant Chinese film studies collection. The collection contains over 70,000 items. The highlight of the collection includes 436 old film periodicals in 5910 issues and a large amount of film ephemera, i.e. 4195 film posters, 21233 lobby cards in 2,194 sets, 3332 theatre flyers, 9,214 photographic negatives & slides, 4,145 stills & publicity photos.

To make these valuable resources available to researchers, we are proposing a collaboration project with Shanghai Library to develop a Chinese film knowledge base based on two great valuable Chinese film collections, Shanghai Library film collection and our collection.

Example of materials: Film Posters



1957?



导演: 刘国权
1957



导演: 石挥
1957



导演: 司徒
1958



导演: 黄佐临
1958



导演: 王家乙
1959



导演: 郑君里
1959



编导: 董列, 张永枚等
1959



导演: 李俊
1959

Scripts, Dialogues



星星月亮太陽 "Sun,
Moon and Star" (对白本)

空中小姐: 对白本
易文

借女幽魂: 对白本

给我一个吻: 剧本本
李翰祥; 编剧: 程刚

夜半歌声 (对白本)
主要演员: 乐蒂, 赵雷等



火烧红莲寺: 第三集
监制: 庄明枢; 编剧: 凌云; 导演: 王天林; 助导: 石堃

挺进! 红卫兵: 征求意见稿
编剧: 崔梓

Flyers



Complexities of the Collection

The collection contains a large amount of various materials including monographs, periodicals, still images, posters, lobby cards, and other prints, and objects.

- Enormous quantity, more than 70000 items
- Various formats and types
- Covers entire live cycle of film, from production to exhibition.

Linked Data, Definitions

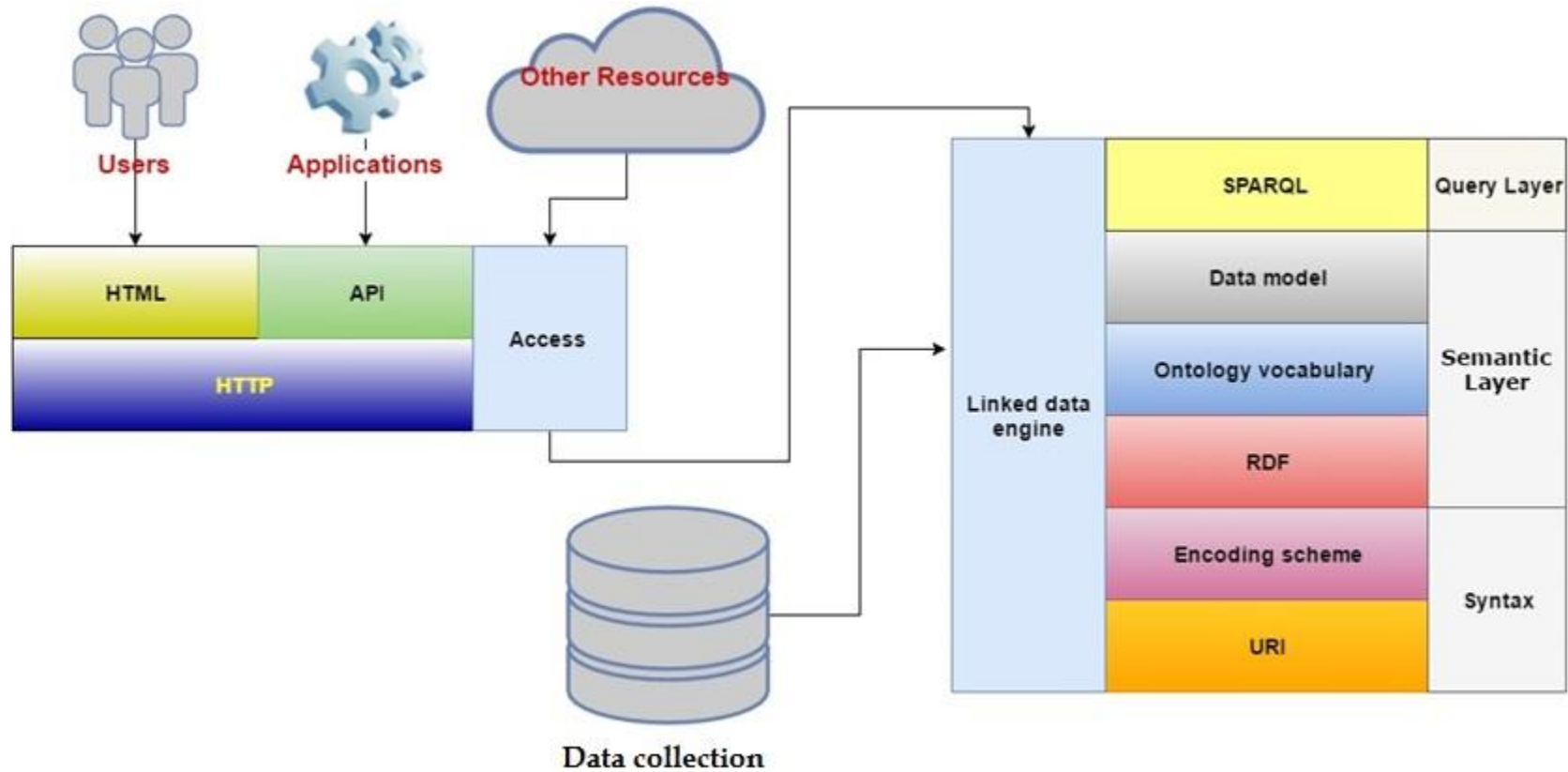
- Old definition in Wikipedia:

a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data , information , and knowledge on the Semantic Web using URIs and RDF.

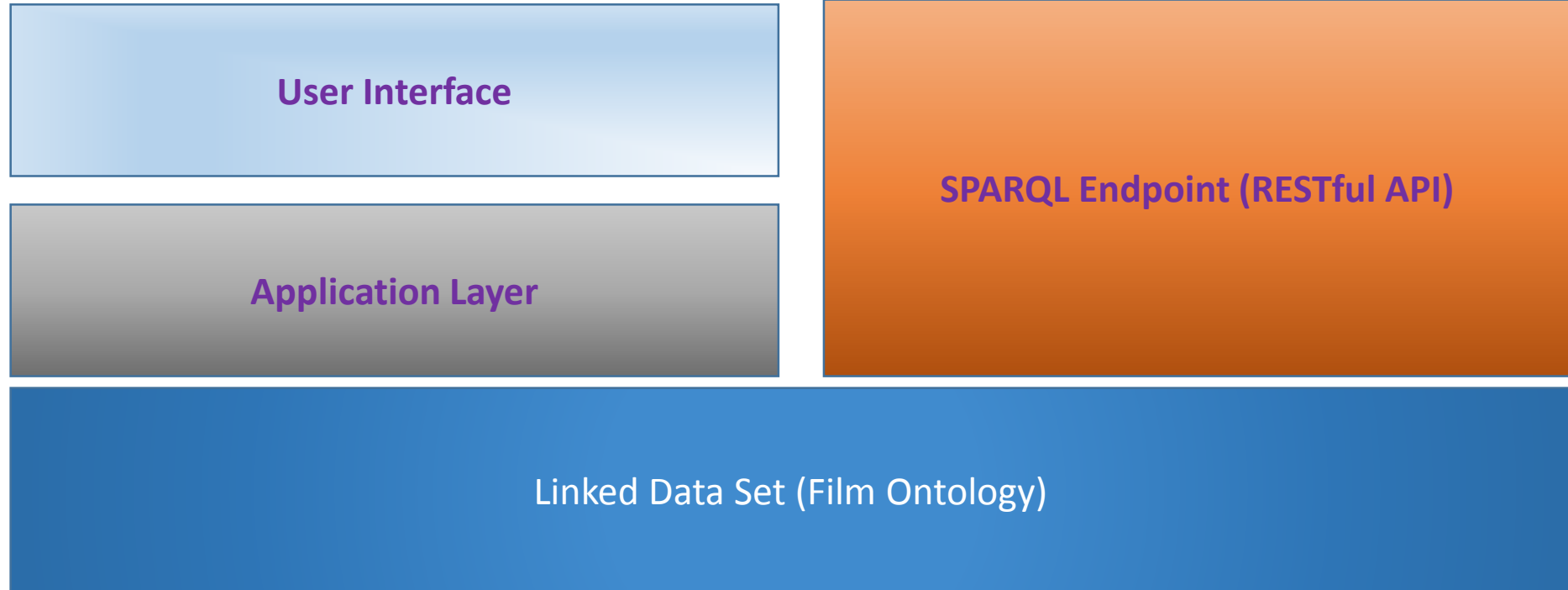
- New definition in Wikipedia:

In computing, linked data (often capitalized as Linked Data) is structured data which is interlinked with other data so it becomes 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**.

Linked Data as a Solution



Architecture

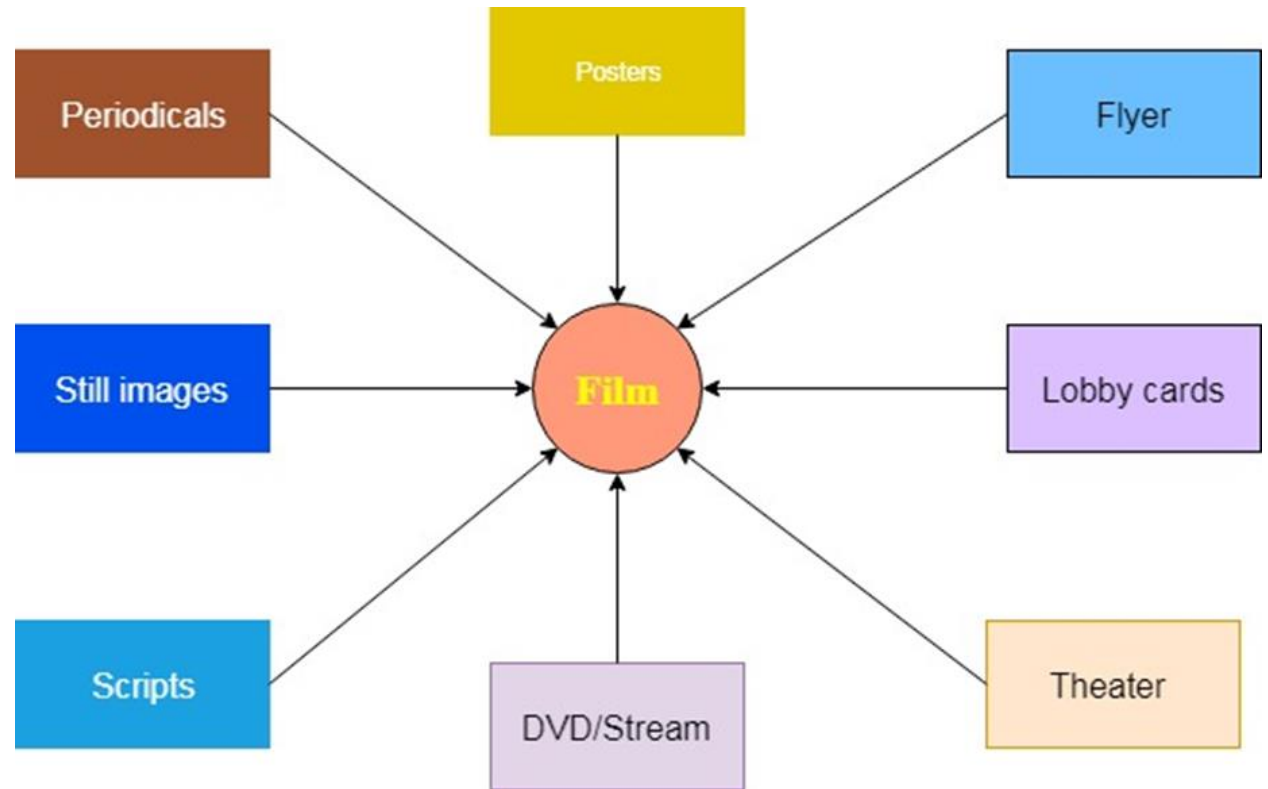


Virtuoso as Triple Store

- Virtuoso Universal Server is a middleware and database engine hybrid that combines the functionality of a traditional Relational database management system (RDBMS), Object-relational database (ORDBMS), virtual database, RDF, XML, free-text, web application server and file server functionality in a single system. The software has been developed by OpenLink Software. (Wikipedia)

Data Model

- Film-centered approach
- Film represented by film materials
- Focused on relationships among film entities
- Life-cycle modeling of films



Film-centered

- RDF: subject predicate object

A film is always a subject when describing something related to a film .

Film Ontology

- Ontology is an unified framework to describe the resources and formalize relationships among film entities as well as film resources, not just the film itself. It is a foundation of query and inference algorithms of the system.
- The ontology will also be used to achieve semantic interoperability with other collections. (OWL:SameAs)

Protégé View

The image displays the Protégé OWL editor interface. On the left, a class hierarchy is shown under 'owl:Thing', including categories like Language, Movie, Organization, Person, Place, and Work. The main area features a network graph with nodes representing classes and properties, connected by arrows. A 'Graph - Overview' window at the bottom left provides a smaller view of the graph. The top of the window shows various tabs for ontology management and a status bar at the bottom indicating the reasoner settings.

Active Ontology x Entties x Classes x Object Properties x Data Properties x Individuals by class x OWLViz x DL Query x NavigOwl x

Class hierarchy: NavigOwl: Asserted

Controls Search

Graph Operations

- Circle Layout
- Random Layout
- Force Layout
- Spring Layout
- Power Layout

Show Nodes Labels

Default Zoom

Reload

Graph Legend

- Class Node
- Instance Node
- Datatype Property Node
- Object Property Node
- Property Node
- Collection Node
- Literal Node
- Default Node

Graph - Overview

Overview

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

Generate RDF triples

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
2	POS02354	'91童黨							20 x 30 cm	MO003654	'91童黨	'91童黨	Gamgs '91		曹建南	鄭嘯
3	POS02355	'93 似夢年華							21 x 31 cm	MO003655	'93 似夢年華	'93 似夢年華				
4	POS02356	'93 似夢年華							21 x 31 cm	MO003655	'93 似夢年華	'93 似夢年華				
5	POS00004	13号地区							103 x 74 cm	MO001349	13號地區	13号地区			陈家林	
6	POS00005	1918年苦難的歷程(第二部)				C111343211		內容簡	53 x 38 cm	MO001350	1918年苦難的歷程(第二部)	1918年苦難的歷程(第二部)				
7	POS00006	2 x 2=5				C111345079			54 x 39 cm	MO001351	2 x 2=5	2 x 2=5				
8	POS00008	306号案件		1E+09	2/13/2C	C118638860			78 x 54 cm	MO001352	306號案件			36号案件	A-雷巴蘭夫	
9	POS00007	306号案件	b23388842	1E+09	Cabine	C111344297			77 x 53 cm	MO001352	306號案件			36号案件	A-雷巴蘭夫	
10	POS02357	3個受傷的警察							40 x 61 cm	MO003664	3個受傷的	3个受伤的	The Log			
11	POS02358	3個茶煲1個佬							40 x 61 cm	MO003665	3個茶煲1個佬	Multiplici	3个茶煲1	夏茶藍米斯		
12	POS02359	405謀殺案							31 x 40 cm	MO003666	405謀殺案	405謀殺案		沈耀庭		
13	POS02360	405謀殺案							21 x 30 cm	MO003666	405謀殺案	405謀殺案		沈耀庭		
14	POS02361	405謀殺案							21 x 30 cm	MO003666	405謀殺案	405謀殺案		沈耀庭		
15	POS00009	45号地区		2/13/2C	C111359031		国营天津人民印刷厂		104 x 77 cm, 77 x 53	MO001354	45號地區			45号地区	米克申斯基	
16	POS00010	45号地区				C111540817	国营天津人民印刷厂		104 x 78 cm	MO001354	45號地區			45号地区	米克申斯基	
17	POS02362	4個婚禮一個葬禮							40 x 61 cm	MO003669	4個婚禮一個葬禮	Four Wed	4个婚礼	米克里維		
18	POS02363	4面夏娃							40 x 61 cm	MO003670	4面夏娃	4面夏娃	4 Faces of Eve	甘國亮, 林		
19	POS00011	51号兵站				C111347758			78 x 53 cm	MO001356	51號兵站	51号兵站		刘琼		
20	POS02364	666魔鬼復活							26 x 40 cm	MO003671	666魔鬼復	666魔鬼复	Satan Returns	林偉倫	蔡忠	
21	POS02366	666魔鬼復活							20 x 30 cm	MO003671	666魔鬼復	666魔鬼复	Satan Returns	林偉倫	蔡忠	
22	POS02365	666魔鬼復活							27 x 40 cm	MO003671	666魔鬼復	666魔鬼复	Satan Returns	林偉倫	蔡忠	
23	POS02367	666魔鬼復活							20 x 30 cm	MO003671	666魔鬼復	666魔鬼复	Satan Returns	林偉倫	蔡忠	
24	POS02368	7金剛							13 x 19 cm	MO003675	7金剛	7金剛	Wonder Seven	程小東	劉儀	
25	POS02369	91家有囍事							40 x 61 cm	MO003676	91家有囍	91家有囍	All's Well End's Well	張堅庭		
26	POS02370	92黑玫瑰與黑玫瑰							30 x 20 cm	MO003677	92黑玫瑰	92黑玫瑰	92 Legendary La Ros	劉鎮偉		
27	POS02371	97古惑仔戰無不勝							40 x 61 cm	MO003678	97古惑仔	97古惑仔	Young and Dangerou	劉偉強	張敬	
28	POS02372	97古惑仔戰無不勝							27 x 40 cm	MO003678	97古惑仔	97古惑仔	Young and Dangerou	劉偉強	張敬	
29	POS02378	B計劃							25 x 39 cm	MO003685	B計劃	B计划	Extreme Crisis	羅禮賢		

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:flm="http://cnfilmstudies.online/ontology/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <flm:poster rdf:about="http://cnfilmstudies.online/filmdata/POS00001">
    <flm:isAbout rdf:resource="http://cnfilmstudies.online/filmdata/M0001343" />
    <flm:title>"特快"列车</flm:title>
    <flm:size>53 x 34 cm</flm:size>
  </flm:poster>
  <flm:movie rdf:about="http://cnfilmstudies.online/filmdata/M0001343">
    <flm:originaltitlePOS00001 xml:lang="zh">特快列车</flm:originaltitlePOS00001>
    <flm:originaltitlePOS00001 xml:lang="zh">特快列车</flm:originaltitlePOS00001>
    <flm:distributedByPOS00001 rdf:resource="http://cnfilmstudies.online/filmdata/ORG00938" />
    <flm:producedByPOS00001 rdf:resource="http://cnfilmstudies.online/filmdata/ORG00192" />
    <flm:description rdf:about="http://cnfilmstudies.online/ontology/originaltitlePOS00001">
      <rdfs:subPropertyOf rdf:resource="http://cnfilmstudies.online/ontology/originaltitle" />
      <flm:hasSource rdf:resource="http://cnfilmstudies.online/filmdata/POS00001"/>
    </rdf:Description>
  </rdf:Description>
  <rdf:Description rdf:about="http://cnfilmstudies.online/ontology/distributedByPOS00001">
    <rdfs:subPropertyOf rdf:resource="http://cnfilmstudies.online/ontology/distributedBy" />
  </rdf:Description>
  <rdf:Description rdf:about="http://cnfilmstudies.online/ontology/producedByPOS00001">
    <rdfs:subPropertyOf rdf:resource="http://cnfilmstudies.online/ontology/producedBy" />
  </rdf:Description>
  <flm:hasSource rdf:resource="http://cnfilmstudies.online/filmdata/POS00001"/>
</rdf:Description>
  <flm:poster rdf:about="http://cnfilmstudies.online/filmdata/POS00002">
    <flm:isAbout rdf:resource="http://cnfilmstudies.online/filmdata/M0001343" />
    <flm:title>"特快"列车</flm:title>
    <flm:size>73 x 53 cm</flm:size>
  </flm:poster>
  <flm:movie rdf:about="http://cnfilmstudies.online/filmdata/M0001343">
    <flm:originaltitlePOS00002 xml:lang="zh">特快列车</flm:originaltitlePOS00002>
    <flm:originaltitlePOS00002 xml:lang="zh">特快列车</flm:originaltitlePOS00002>
    <flm:distributedByPOS00002 rdf:resource="http://cnfilmstudies.online/filmdata/ORG00938" />
    <flm:producedByPOS00002 rdf:resource="http://cnfilmstudies.online/filmdata/ORG00192" />
  </flm:movie>
</rdf:RDF>
```


Size of the data set

Materials

Poster	Flyer	Script	Pamphlet	Theaters (not finished)
4180	5	154	1342	41

Objects

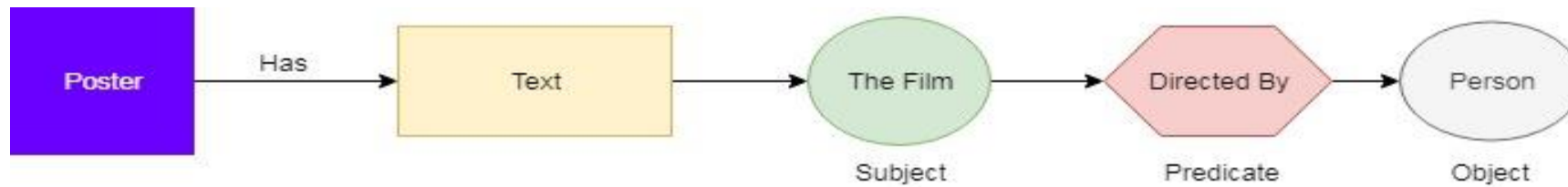
Film	Person	Organization	Country
3980	9840	1850	56

Triples

Triples
251609

Provenance

- Representing complicated sentences like
“The text on the poster said ‘XXXX directed by YYY’ ”



五朵金花 导演 王家乙



Singleton property

Subject	Predicate	Object
BobDylan	isMarriedTo	SaraLownds
BobDylan	isMarriedTo#1	SaraLownds
isMarriedTo#1	rdf:singletonPropertyOf	isMarriedTo
isMarriedTo#1	hasStart	1965-11-22
isMarriedTo#2	hasEnd	1973-12-03

Source: Don't Like RDF Reification? Making Statements about Statements Using Singleton Property by Vinh Nguyen, Olivier Bodenreider, and Amit Sheth, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350149/> on 10/23/2018 Subject

Singleton Property Presentation

- How to represent



五朵金花 导演 王家乙

```
<http://cnfilmstudies.online/filmdata/MO001551>  
  a          flm:movie ;  
  flm:hasDirectorPOS00203 <http://cnfilmstudies.online/filmdata/PN00003356> ;  
  flm:originaltitlePOS00203 "五朵金花"@zh .
```

```
<http://cnfilmstudies.online/ontology/originaltitlePOS00203>  
  flm:hasSource <http://cnfilmstudies.online/filmdata/POS00203> ;  
  rdfs:subPropertyOf <http://cnfilmstudies.online/ontology/originaltitle> .
```

```
<http://cnfilmstudies.online/filmdata/POS00203>  
  a          flm:poster ;  
  flm:isAbout <http://cnfilmstudies.online/filmdata/MO001551> .
```

```
<http://cnfilmstudies.online/ontology/hasDirectorPOS00203>  
  flm:hasSource <http://cnfilmstudies.online/filmdata/POS00203> ;  
  rdfs:subPropertyOf <http://cnfilmstudies.online/ontology/hasDirector> .
```

Defined Classes

- $\text{Director} \equiv \text{person} \cap \exists \text{ hasDirector.movie}$

```
<owl:Class rdf:about="http://cnfilmstudies.online/ontology/director">
  <rdfs:label xml:lang="en">Director</rdfs:label>
  <rdfs:comment xml:lang="en">The person with the overall responsibility for all creative and technical aspects of making
a film</rdfs:comment>
  <rdfs:subClassOf rdf:resource="http://cnfilmstudies.online/ontology/person"/>
  <owl:equivalentClass><owl:Restriction><owl:someValuesFrom
rdf:resource="http://cnfilmstudies.online/ontology/person"/>
  <owl:onProperty
rdf:resource="http://cnfilmstudies.online/ontology/hasDirector"/></owl:Restriction></owl:equivalentClass>
</owl:Class>
```

SPARQL Endpoint

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

```
select distinct ?Concept where {[] a ?Concept} LIMIT 100
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format:

Execution timeout: milliseconds *(values less than 1000 are ignored)*

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#).)

Access Sparql Endpoint via Excel

Function sparql()

Dim sparqlUrl As String

sparqlUrl = ["http://45.34.14.119:8890/sparql?default-graph-uri=&query=%0D%0Aselect+distinct+%3FConcept+where+%7B%5B%5D+a+%3FConcept%7D+LIMIT+100&format=text%2Fhtml&timeout=0&debug=on"](http://45.34.14.119:8890/sparql?default-graph-uri=&query=%0D%0Aselect+distinct+%3FConcept+where+%7B%5B%5D+a+%3FConcept%7D+LIMIT+100&format=text%2Fhtml&timeout=0&debug=on)

Set sparqlReq = CreateObject("MSXML2.XMLHTTP")

With sparqlReq

.Open "GET", sparqlUrl, False

.Send

End With

sparql=sparqlReq.ResponseText

Architecture of Application Layer



Application Layer Example


45.34.14.119:8890/sparql?defa x 45.34.14.119/cgi-enabled/dat x +

← → ↻ ⌂ ⓘ Not secure | 45.34.14.119/cgi-enabled/dataengine.py?name=赵丹&type=actor ☆

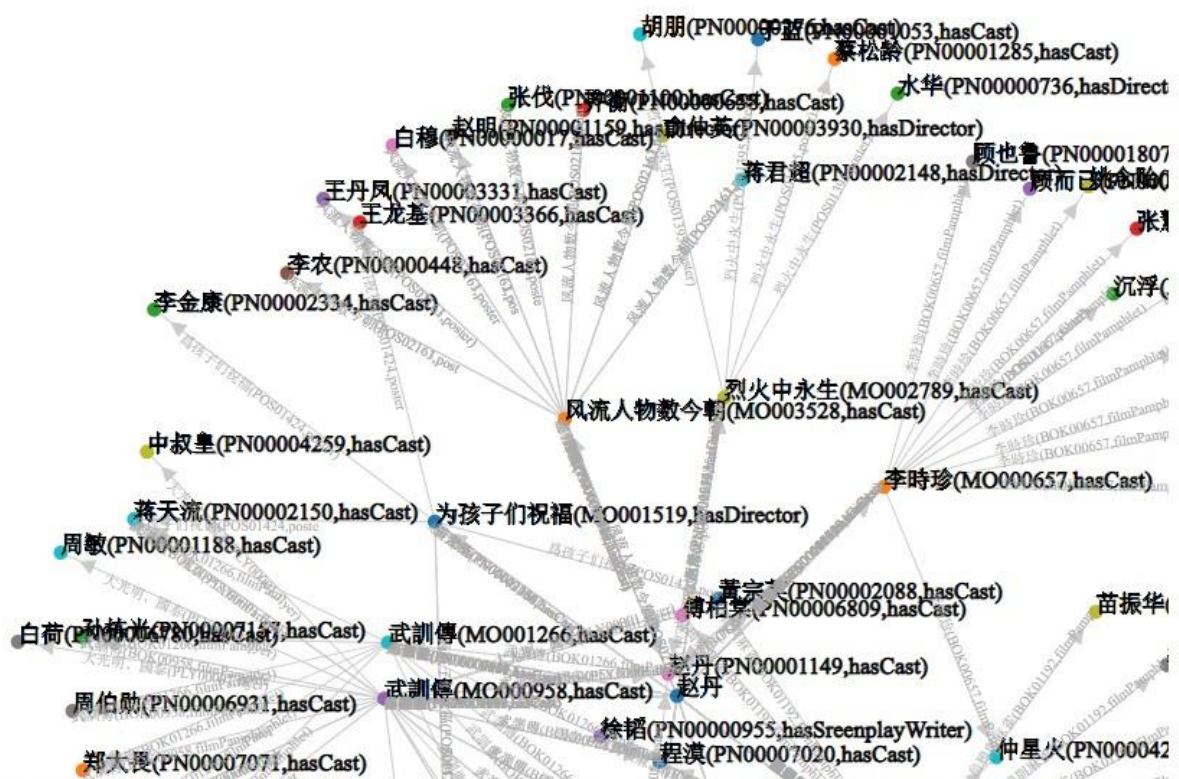
Apps Suggested Sites Imported From IE 馆 Excel用户窗体技术... Google Develop... Using PHP/MySQL... Presentation topi... features >> Other bookmarks

17:01:37.641421 赵丹 17:01:40.941349

赵丹
Zhao Dan (June 27, 1915 - October 10, 1980) was a Chinese actor popular in the golden age of Chinese Cinema. ---from [Wikipedia](#)



- hasCast 李時珍 from: 李時珍 (filmPamphlet)
- hasDirector 沉浮 from: 李時珍 (filmPamphlet)
- hasCast 舒适 from: 李時珍 (filmPamphlet)
- hasCast 顾而已 from: 李時珍 (filmPamphlet)
- hasCast 康泰 from: 李時珍 (filmPamphlet)
- hasCast 赵丹 from: 李時珍 (filmPamphlet)
- hasCast 顾也鲁 from: 李時珍 (filmPamphlet)
- hasCast 姚念贻 from: 李時珍 (filmPamphlet)
- hasCast 仲星火 from: 李時珍 (filmPamphlet)
- hasCast 舒绣文 from: 李時珍 (filmPamphlet)
- hasCast 钱千里 from: 李時珍 (filmPamphlet)
- hasCast 程之 from: 李時珍 (filmPamphlet)
- hasScreenplayWriter 张慧剑 from: 李時珍 (filmPamphlet)



胡朋 (PN00009253, hasCast) 蔡松龄 (PN00001285, hasCast) 水华 (PN00000736, hasDirect)

张伐 (PN00001100, hasCast) 俞仲英 (PN00003930, hasDirector) 顾也鲁 (PN00001807) 顾而已 (姚念贻)

白穆 (PN00000017, hasCast) 王丹凤 (PN00003331, hasCast) 王龙基 (PN00003366, hasCast) 蒋君超 (PN00002148, hasDirect)

李金康 (PN00002334, hasCast) 李农 (PN00000448, hasCast) 烈火中永生 (MO002789, hasCast) 风流人物数今朝 (MO003528, hasCast) 李時珍 (MO000657, hasCast)

中叔皇 (PN00004259, hasCast) 蒋天流 (PN00002150, hasCast) 为孩子们祝福 (MO001519, hasDirector) 周敏 (PN00001188, hasCast)

白荷 (PN00000715, hasCast) 武训传 (MO001266, hasCast) 傅柏如 (PN00006809, hasCast) 苗振华

周伯勋 (PN00006931, hasCast) 武训传 (MO000958, hasCast) 赵丹 (PN00001149, hasCast) 徐韬 (PN00000955, hasScreenplayWriter) 程漠 (PN00007020, hasCast) 仲星火 (PN0000042)

郑大畏 (PN00007071, hasCast)

SPARQL Template for relationships

```
PREFIX flm:<http://cnfilmstudies.online/ontology/>
```

```
PREFIX shl:<http://www.library.sh.cn/ontology/>
```

```
select distinct ?qsstitle ?qsst ?c ?fm ?qs ?bn ?ws ?wsstitle ?wsst where {{?a a flm:person.
```

```
    ?b a flm:person.
```

```
    ?c a flm:movie.
```

```
    ?c ?q ?a.
```

```
    ?c ?w ?b.
```

```
    ?a flm:name "赵丹"@zh.
```

```
    ?b flm:name ?bn.
```

```
    ?q rdfs:subPropertyOf ?qs.
```

```
    ?q flm:hasSource ?qss.
```

```
    ?qss a ?qsst.
```

```
    ?qss flm:title ?qsstitle.
```

```
    ?w rdfs:subPropertyOf ?ws.
```

```
    ?w flm:hasSource ?wss.
```

```
    ?wss a ?wsst.
```

```
    ?wss flm:title ?wsstitle.
```

```
  } OPTIONAL {
```

```
    ?c ?t ?fm.
```

```
    ?t rdfs:subPropertyOf flm:originaltitle}
```

```
  }
```

SPARQL Template for Defined Classes

```
prefix flm:<http://cnfilmstudies.online/ontology/>
```

```
select distinct ?cn ?c where {?a a flm:movie .
```

```
  ?a ?q ?c.
```

```
  ?q rdfs:subPropertyOf ?u.
```

```
  ?class owl:equivalentClass ?r.
```

```
  ?r owl:onProperty ?u.
```

```
  ?class rdfs:label ?cn} LIMIT 100
```

SPARQL Template for Data Integration : Federated Query

- Ability to take a query based on information from many different sources.

```
prefix shl: <http://www.library.sh.cn/ontology/>
```

```
prefix flm:<http://cnfilmstudies.online/ontology/>
```

```
select distinct * where{?a shl:actor ?b .
```

```
    ?b flm:name "赵丹"@zh.
```

```
    ?a foaf:name ?t. SERVICE <http://data1.library.sh.cn:8890/sparql> {?x dc:title ?t}
```

```
. SERVICE <http://dbpedia.org/sparql> {?w rdfs:label "赵丹"@zh.
```

```
    ?w a <http://dbpedia.org/ontology/Artist>.
```

```
    ?w rdfs:comment ?comm. filter(lang(?comm)="zh")}}
```

As a Service: an AJAX example

A Linked Data Set For Chinese Film Studies

Search for: Types: Relationships--Person Search

18:39:16.988559 刘晓庆 18:39:18.175919

刘晓庆
Liu Xiaoqing (born 30 October 1955) is a Chinese actress and businesswoman. She was one of the leading actresses in China in the 1980s.---from Wikipedia

- hasCast 潜网 from: 潜网 (poster)
- hasCast 林彬 from: 潜网 (poster)
- hasCast 刘晓庆 from: 潜网 (poster)
- hasCast 周森冠 from: 潜网 (poster)
- hasDirector 王好为 from: 潜网 (poster)
- hasCast 南海长城 from: 南海长城 (poster)
- hasCast 刘晓庆 from: 南海长城 (poster)
- hasCast 王心刚 from: 南海长城 (poster)
- hasCast 石韧 from: 南海长城 (poster)
- hasDirector 李俊 from: 南海长城 (poster)
- hasDirector 郝光 from: 南海长城 (poster)

```
<script> == $0
$(document).ready(function(){
  $("#button").click(function(){
    $.ajax({
      type: "GET",
      url: "cgi-enabled/dataengine.py",
      data: {name: $("#name").val(), type: $("#select").val()},
      timeout: (5 * 1000000),
      success: function(response){
        $("#datadisp").empty();
        $("#datadisp").html(response);
        //
        // ("#datadisp").find("script").each(function(){
        //   eval($(this).text());
        // });
      }
    });
  });
});
```

User Interface

A Linked Data Set For Chinese Film Studies

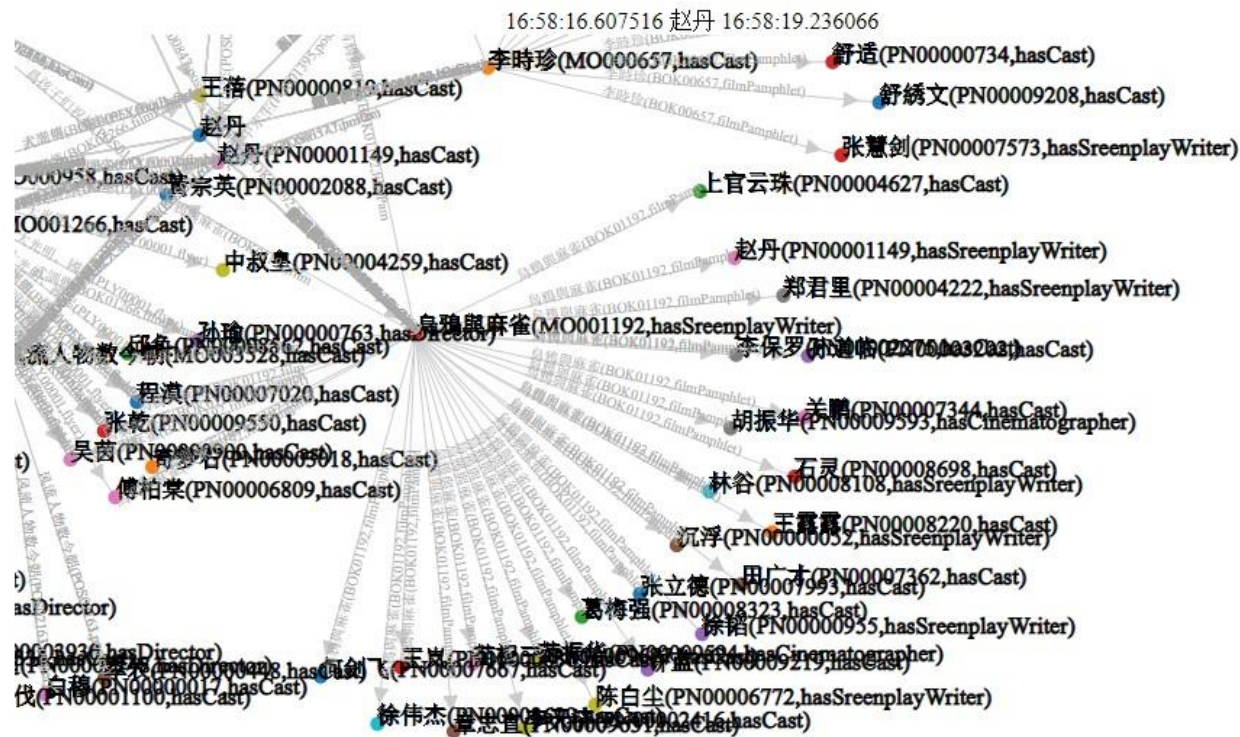
Search for: Types: Relationships--Person

赵丹

Zhao Dan (June 27, 1915 - October 10, 1980) was a Chinese actor popular in the golden age of Chinese Cinema.---from [Wikipedia](#)



- hasCast 李時珍 from: 李時珍 (filmPamphlet)
- hasDirector 沉浮 from: 李時珍 (filmPamphlet)
- hasCast 舒适 from: 李時珍 (filmPamphlet)
- hasCast 顾而已 from: 李時珍 (filmPamphlet)
- hasCast 康泰 from: 李時珍 (filmPamphlet)



Working with researchers

- Challenges
- Products vs. Platform
- Inputs from researchers

Thank PRRLA for The Karl Lo Award,
and
Thank you all

- <http://45.34.14.119/cgi-enabled/dataengine.py?name=%E8%B5%B5%E4%B8%B9&type=person>
- <http://45.34.14.119/cgi-enabled/dataengine.py?name=%E8%B5%B5%E4%B8%B9&type=actor>
- <http://45.34.14.119/cgi-enabled/dataengine.py?name=刘晓庆|姜文&type=actor>