

Dunhuang Mural Thesaurus

敦煌壁画主题词表

Xiaoguang Wang



Center for Digital Humanities
Wuhan University, China



Dunhuang Mogao Grottoes Cave

Dunhuang Mogao caves, located south-east of the Dunhuang oasis on the Silk Road, in Gansu province, China. With 492 caves, the total size of the murals reaches more than 45,000 square meters .

The numerous huge number of Dunhuang murals contain very exceedingly rich content, making this site a significant academic treasure with an abundance of vivid materials depicting various aspects of medieval politics, economics, culture, arts, religion, ethnic relations, and daily dress in Western China.



Dunhuang Mogao Grottoes Cave



Dunhuang Mogao Grottoes Cave 285

Purpose

The absence of Dunhuang Mural vocabularies not only hinders the reasonable and efficient organization of relevant Dunhuang mural resource, but also limits the studies and explorations of its values.



Approach

A combination of top-down and bottom-up approach is used. The top-down approach starts at the general, conceptual levels, and establishes a general framework for the thesaurus. A bottom-up approach builds up important categories from the concepts that are extracted from the corpus using NLP technology.



Benefits

The thesaurus provides a controlled vocabulary for semantic annotation, knowledge organization and information retrieval of digital resources of Dunhuang murals. As a professional and high-quality corpus resource, it can improve the automation and intelligence level of information resource organization.

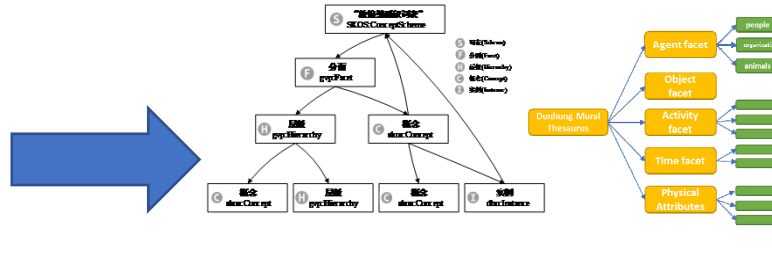
Work flow



- Term management
- Scope note & definition
- Quality control

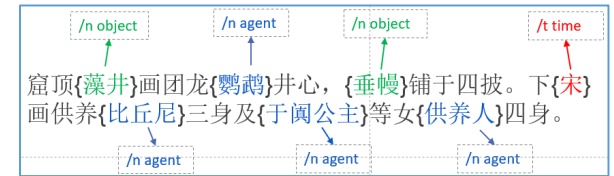


Thesaurus Management and Maintenance



Top-down Structure Design

Dunhuang Mural Thesaurus



- Word Segmentation
- POS Tagging
- Grouping & Classification
- Establishing relationship
- Thesaurus expansion
- Structure adjustment

Bottom-up Thesaurus Optimization

Features

- **5 facets:** Agents, Objects, Activities, Time, and Physical attributes
- **25 second-level categories**
- **4276 terms:** 3199 concepts, 989 instances
- **Chinese and English authority notes** and image notes.
- **Covering the Mural protection and restoration, religion, iconography, cave archaeology, humanity, culture and other research perspectives.**



Facets and second-level categories

Thesaurus Management and Visualization

菩萨

Home / 代理者 / 人物 / <依宗教区分> / <佛教相关人物> / 佛家神祇 / 菩萨

Term Notes **3** Metadata

Scope note

- 梵文音译全称为“菩提萨埵”，意为“觉有情”，即“上求菩提(觉悟)、下化有情(众生)”的人。佛教指达到自觉(自身得到解脱)、觉他(使众生达到解脱)两项修行果位者。原为释迦牟尼尚未成佛时的称号。后指修行到了一定程度，地位仅次于佛的人。在民间，有时也称佛教神像为菩萨。——汉语大词典
- 大乘佛教里，指已达成佛陀开悟，却选择延后进入涅槃，留在世上帮助人们开悟的慈悲者。南传佛教里，此词用来专指佛陀们的前生。——AAT台湾

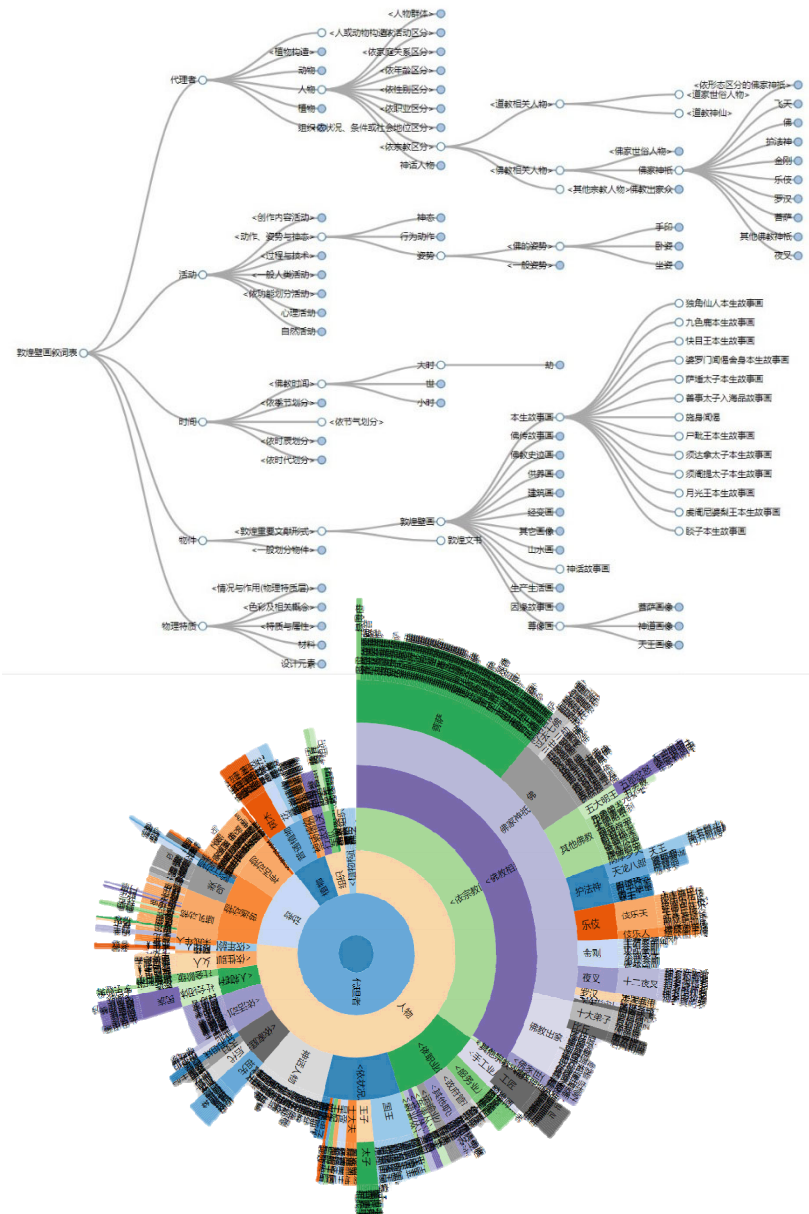
Scope note (en)

- In Mahayana Buddhism, one who attains the enlightenment of a Buddha but chooses to postpone entering nirvana and instead stays in the world for the compassionate purpose of helping lesser beings attain enlightenment. In Theravada Buddhism, the term is exclusively used to identify historical Buddhas in their previous lives.——AAT台湾
- One destined to become a Buddha; a Buddhist saint who, having only one birth to undergo before attaining Nirvana, consents to be reborn for the sake of suffering mankind; a superhuman being of infinite wisdom and compassion.——Oxford English Dictionary

图片注释



Tema Tres
Vocabulary Server



Linked Data Publishing

敦煌壁画主题词表关联数据SPARQL查询

选择查询的模块，在编辑框中输入SPARQL查询语句，点击右上角查询按钮执行查询。

1 主题词表本体相关查询

- 1.1 本体包含的类
敦煌壁画主题词表本体模型的类。
`select DISTINCT ?class
where{ ?class rdfs:subClassOf* skos:Concept }`
- 1.2 查询所有谓词
敦煌壁画主题词表本体模型使用的属性。
`select DISTINCT ?
property where{ ?subject rdfs:domain ?object }`

2 主题词相关查询

- 2.1 主题词模糊查询
查询首选项中含有“菩萨”的主题词。
`select * where{ ?concept skos:prefLabel ?label.
filter(contains(?label, '菩萨')) }`
- 2.2 主题词精确查询
查询“菩萨”的主题词。
`select * where{ ?concept skos:prefLabel '菩萨' . }`

```
1 PREFIX app: <http://vocab.gstty.edu/ontology#>  
2 PREFIX dc: <http://purl.org/dc/element/1.1/>  
3 PREFIX foaf: <http://purl.org/foaf/terms/1.1/>  
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
5 PREFIX owl: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
6 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>  
7 PREFIX dhvocab: <http://dh.whu.edu.cn/dhvocab#>  
8 SELECT ?term1 ?term2 WHERE {  
9 skos:broader(dhvocab:tema100) skos:broader* ?term1.  
10 skos:broader(dhvocab:tema100) skos:broader* ?term2.  
11 ?term1 rdfs:label ?label1.  
12 ?term2 rdfs:label ?label2.  
13 filter(?label1 <> ?label2).  
14 }
```

查询结果:

facets	label
1 dhvocab:tema100	"代理者"@zh
2 dhvocab:tema181	"物理特性"@zh
3 dhvocab:tema182	"活动"@zh
4 dhvocab:tema183	"物件"@zh
5 dhvocab:tema247	"时间"@zh

SPARQL query



菩萨

Subject	Predicate	Object	All
词表的属性列表:			
skosaltLabel		菩萨汉语	rdf:langString
skosscopeNote		梵文音译全称“菩萨摩訶萨”，意为“觉有情”，即“上求菩提(觉悟)下化有情众生”的人。佛教徒达到解脱轮回解脱行果位者，即为释迦牟尼佛成佛前的称号。此称号打到了大乘佛教。地位仅次于佛的人。在汉语，有时也简称菩萨摩訶萨。——《AAT词条》 In Mahayana Buddhism, one who attains the enlightenment of a Buddha but chooses to postpone entering Nirvana and instead stays in the world for the compassionate purpose of helping lesser beings attain enlightenment. In Theravada Buddhism, the term is exclusively used to identify historical Buddhas in their previous lives.——AAT词条 One destined to become a Buddha; a Buddhist saint who, having only one birth to undergo before attaining Nirvana, consents to be reborn for the sake of suffering mankind; a superhuman being of infinite wisdom and compassion.——Oxford English Dictionary	rdf:langString
skosprefLabel		菩萨	rdf:langString
rof:type	skos:Concept		rdf:langString
skosbroader	dhvocab:tema204	佛教神祇	
skosnarrower	dhvocab:tema236	弥勒菩萨	
	dhvocab:tema240	妙吉祥菩萨	
	dhvocab:tema241	地藏菩萨	
	dhvocab:tema242	观世音菩萨	
	dhvocab:tema243	大势至菩萨	
	dhvocab:tema244	虚空	
skosexactMatch	aat:300264360		
skosinScheme	<http://118.197.150.55/81/tema3.0/vocab/>		

RDF Statements

<http://dh.whu.edu.cn/dhvocab/home>

敦煌壁画主题词表查询

输入关键词搜索

FIVE FACETS

敦煌壁画主题词表五大方面

- 代前菩萨像
- 敦煌壁画
- 活动
- 物件
- 物理特性

FEATURES

敦煌壁画主题词表特征

- 人机协同机制构建
- 关联数据发布
- 高质量规范词典

ABOUT THE PROJECT

项目介绍

精选词汇

- 弥勒
- 观世音

敦煌莫高窟与敦煌壁画

Linked with AAT



Getty Vocabularies: LOD

SPARQL Queries Any Search... Search Brief

Query:

```
1 prefix skos:<http://www.w3.org/2004/02/skos/core#>
2 prefix gvp:<http://vocab.getty.edu/ontology#>
3 select ?subj ?label ?title where {
4 ?subj skos:prefLabel | skos:altLabel "菩薩"@zh-Hant.
5 ?subj gvp:parentString ?title.
6 ?subj skos:prefLabel ?label. FILTER(lang(?label)= 'en').
7 }
```

SPARQL query

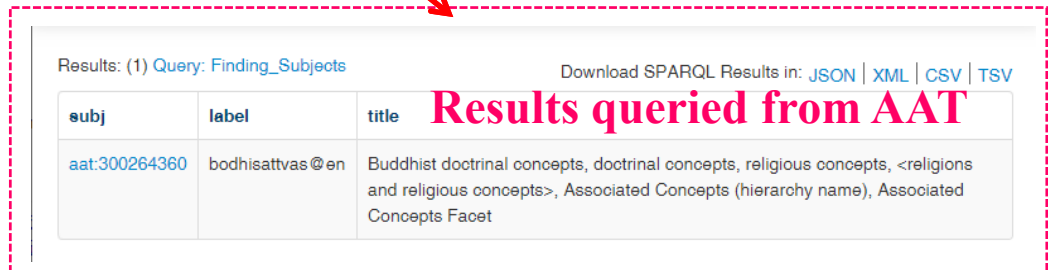
Include inferred
 Expand results over equivalent URIs

Submit

Enriching the entities by linking to external resources such as Getty vocabularies which can enhance the interoperability, reusability and findability of the Dunhuang mural thesaurus.

Query: Using the “菩薩” as the keyword to query the preferred label and alternative label of AAT through a SPARQL query endpoint.

Query results



Results: (1) Query: Finding_Subjects Download SPARQL Results in: JSON | XML | CSV | TSV

subj	label	title
aat:300264360	bodhisattvas@en	Buddhist doctrinal concepts, doctrinal concepts, religious concepts, <religions and religious concepts>, Associated Concepts (hierarchy name), Associated Concepts Facet

Results queried from AAT

Linked with AAT

Enriching the Dunhuang mural thesaurus by linking the resources with AAT.



Dunhuang Mural Thesaurus
<http://dh.whu.edu.cn/dhvocab/tema239>
 (Subject)

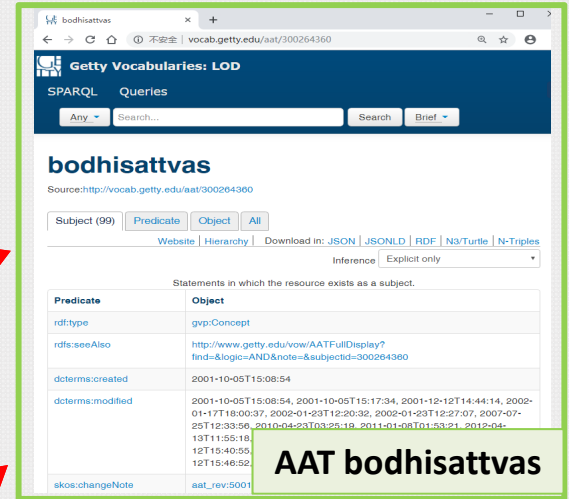
Based on the Getty Vocabulary Program (GVP) ontology, SKOS standard is used to describe and transform the thesaurus.

Field (Property)	Value (Object)
skos:prefLabel	菩萨@zh
skos:altLabel	菩提萨埵
skos:broader	tema204 (佛家神祇)
skos:narrower	tema2387 (文殊菩萨) tema2388 (普贤菩萨)
skos:scopeNote	大乘佛教里·指已达成佛陀开悟....@zh
skos:scopeNote	In Mahayana Buddhism, one who attains the....@en

- Oxford English Dictionary ← *dc:source*
- 《佛学大辞典》 ← *dc:source*
- 《敦煌大辞典》 ← *dc:source*

skos:exactMatch

skos:scopeNote

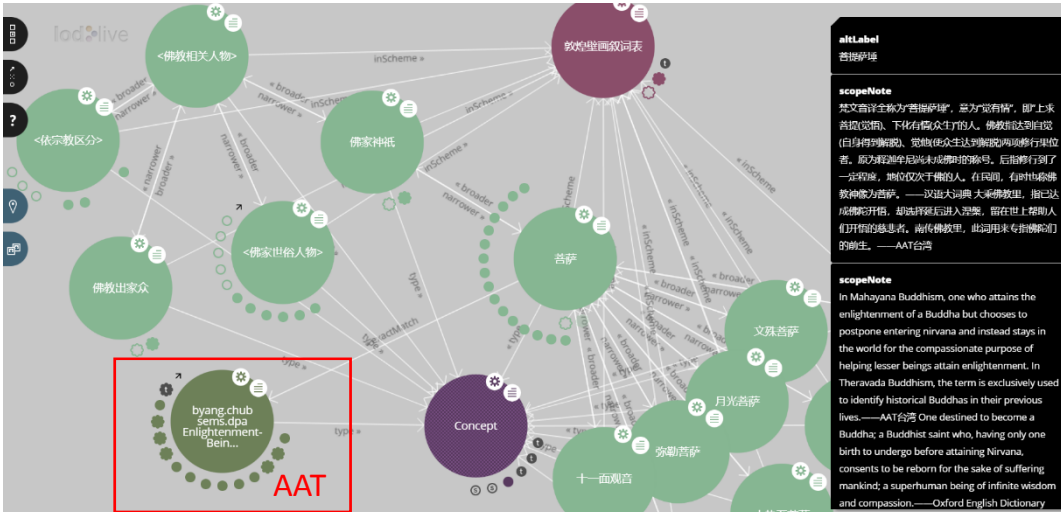


The concepts were linked with AAT by using the *skos:exactMatch* property after querying the equality terms. The scope Notes of some terms also comes from the AAT .

Linked Data Publishing

Facets	Concepts	Linked with AAT	Percentage
Agents	911	115	12.6%
Time	131	36	27.5%
Activities	494	61	12.3%
Objects	2075	167	8.0%
Physical Attributes	242	51	21.1%
Total	3853	430	11.2%

Type	Number
Facets	5
Hierarchy terms	83
Concepts	3199
Instances	989
Total Terms	4276



Website:

<http://dh.whu.edu.cn/dhvocab/home>

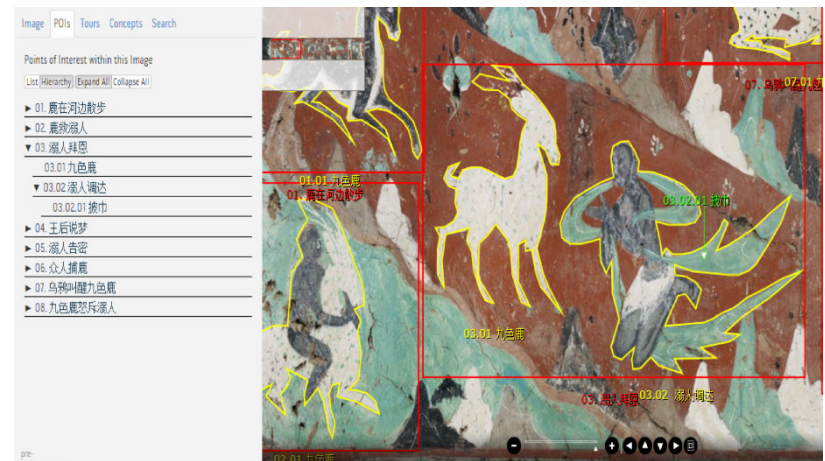
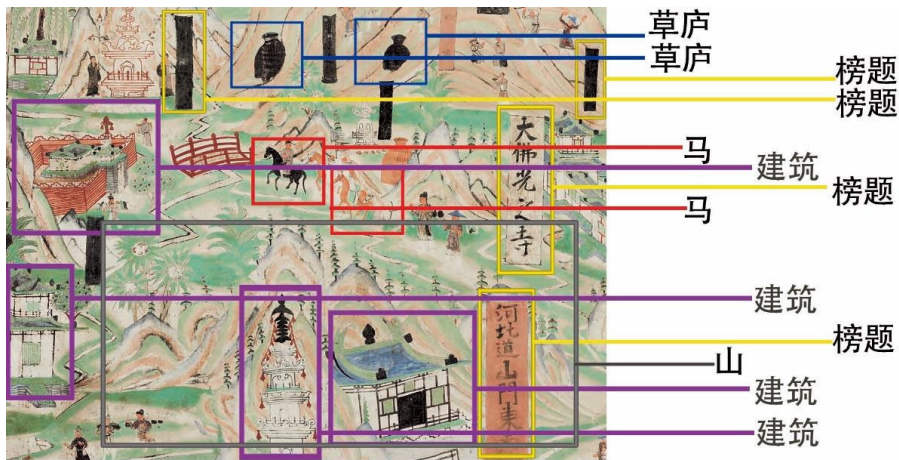
SPARQL Endpoint:

<http://dh.whu.edu.cn:7480/sparql>

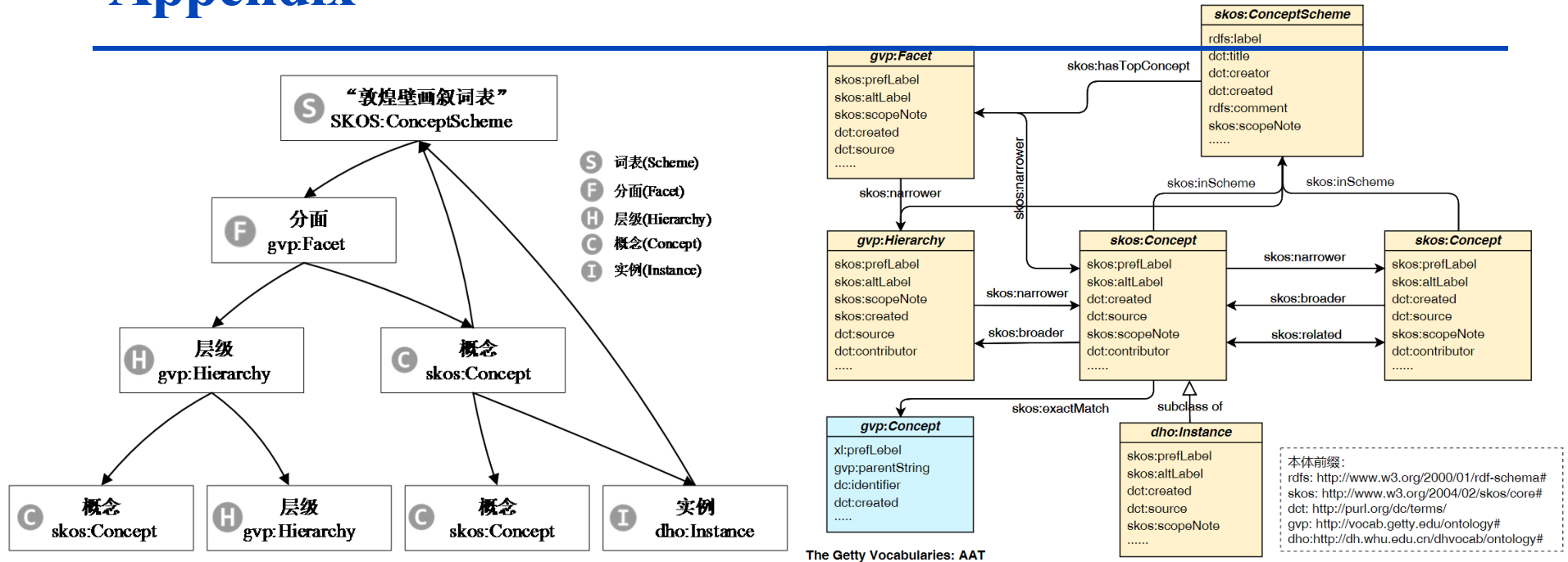
Application programs

Semantic annotation of the mural digital images with the Dunhuang

Mural thesaurus.

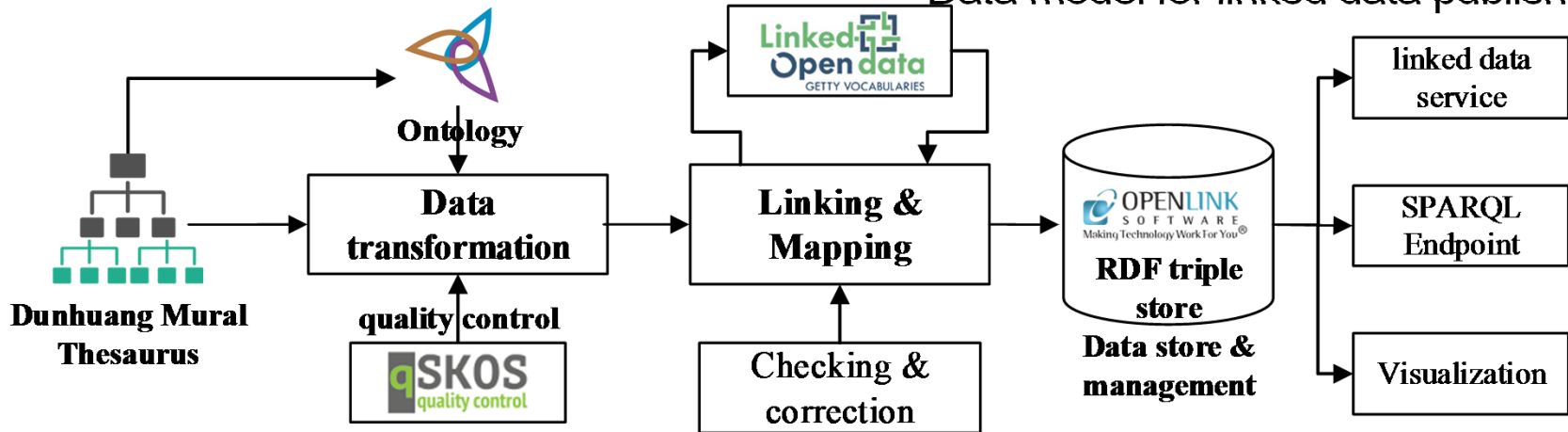


Appendix



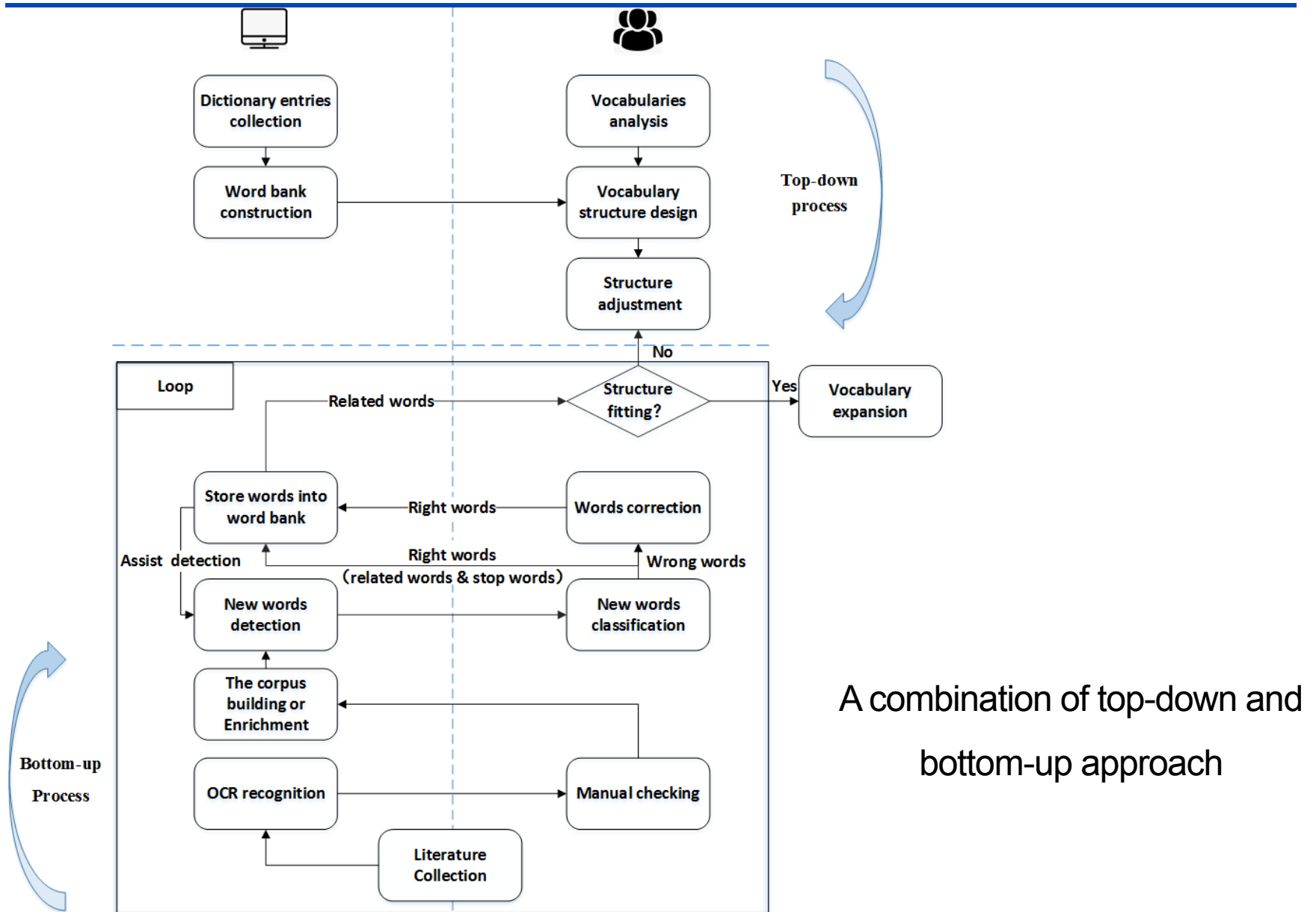
Hierarchical structure of the thesaurus

Data model for linked data publishing



work flow of linked data publishing

Appendix



Appendix

敦煌壁画主题词表可视化浏览

dh.whu.edu.cn/dhvocab/ontodia

DH 敦煌壁画主题词表关联数据

首页 词表可视化 可视化浏览 SPARQL查询 本体模型

Clear All Layout Data Language - English

Classes

- Concept 2723
- ConceptScheme 1
- Facet 5
- Hierarchy 82
- Instance 1080

Instances

Has type Concept

- 孙吴
- 晋朝
- 西晋
- 十六国
- 前赵
- 后赵
- 前燕
- 前凉
- 前秦

Graph Structure:

- 敦煌壁画主题词表 (ConceptScheme) hasTopConcept 活动 (Facet)
- 敦煌壁画主题词表 (ConceptScheme) hasTopConcept 代理人 (Facet)
- 敦煌壁画主题词表 (ConceptScheme) hasTopConcept 物理特质 (Facet)
- 敦煌壁画主题词表 (ConceptScheme) hasTopConcept 时间 (Facet)
- 敦煌壁画主题词表 (ConceptScheme) hasTopConcept 物件 (Facet)
- 代理人 (Facet) hasTopConcept 植物 (Concept)
- 代理人 (Facet) hasTopConcept 组织 (Concept)
- 植物 (Concept) narrower/broader 植物构造 (Hierarchy)
- 植物构造 (Hierarchy) narrower/broader 动物 (Concept)
- 植物构造 (Hierarchy) narrower/broader 人或动物构造 (Hierarchy)
- 动物 (Concept) narrower/broader 组织 (Concept)
- 物件 (Facet) exactMatch 菩萨 (Concept)
- 菩萨 (Concept) exactMatch 事物 (Thing 300264360)

Visual browsing of the thesaurus based on *Ontodia*