

人工智能时代的知识组织方法 知识本体 (Ontology)

上海图书馆高级工程师

夏翠娟

2018年6月@****大学

1. 背景

The background of the slide is a deep blue and black night sky filled with numerous stars of varying brightness. A prominent feature is the Milky Way galaxy, which appears as a dense, glowing band of light and dust stretching diagonally from the bottom center towards the top right. The stars are scattered throughout the field, with some appearing as sharp points of light and others as soft, out-of-focus spots.

知识组织在图档博领域的发展

Knowledge organization (KO), organization of knowledge, organization of information, or information organization **is a branch of library and information science (LIS)** concerned with activities such as document description, indexing, and classification performed in libraries, databases, archives, etc.

https://en.wikipedia.org/wiki/Knowledge_organization

知识组织方法

受控词表（主题词）

索引

分类

分面

用户贡献内容

领域分析

知识组织体系

规范档authority files, 叙

词表 thesauri,

分类法classification ,

主题图topic maps,

分众分类Folksonomies

SKOS

知识组织在万维网上的应用

从 语义万维网（Semantic Web）

简单知识组织体系(SKOS)

到 关联数据（Linked Data）

规范档、主题词表、叙词表

到 知识图谱（Knowledge Graph）

谷歌知识图谱（google knowledge graph）

1960-70s

语义网络

Semantic Network

System Development
Corporation

1990-2000s

万维网

Web

Tim Berners-Lee

2006-2010s

关联数据

Linked Data

Tim Berners-Lee

1980-90s

描述逻辑

Description
Logic

Schild,
Schmiedel

2000s-2006

语义万维网

Semantic Web

W3C

2010s~

知识图谱

Knowledge
Graph

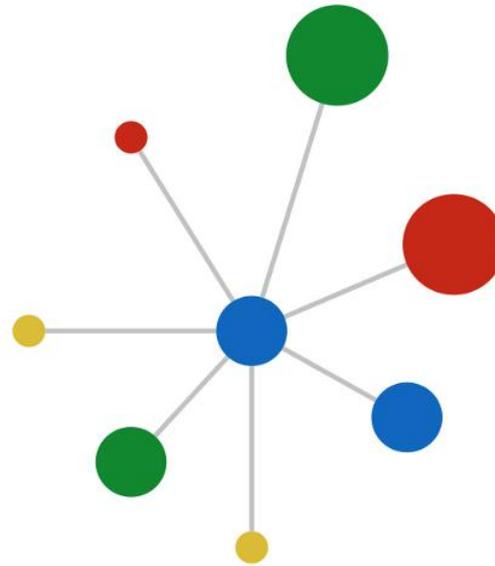
Google



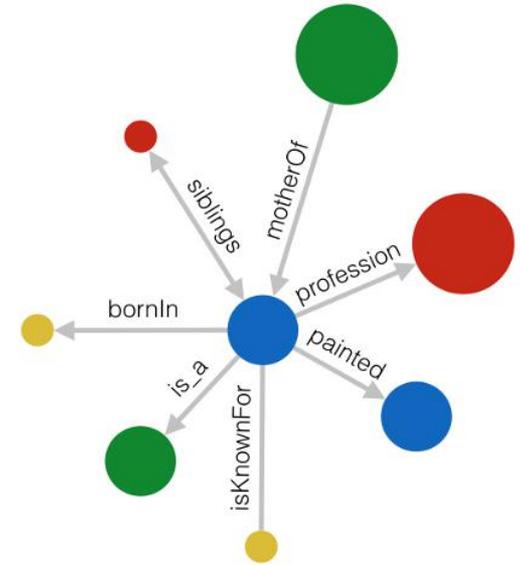
"leonardo da vinci"
String



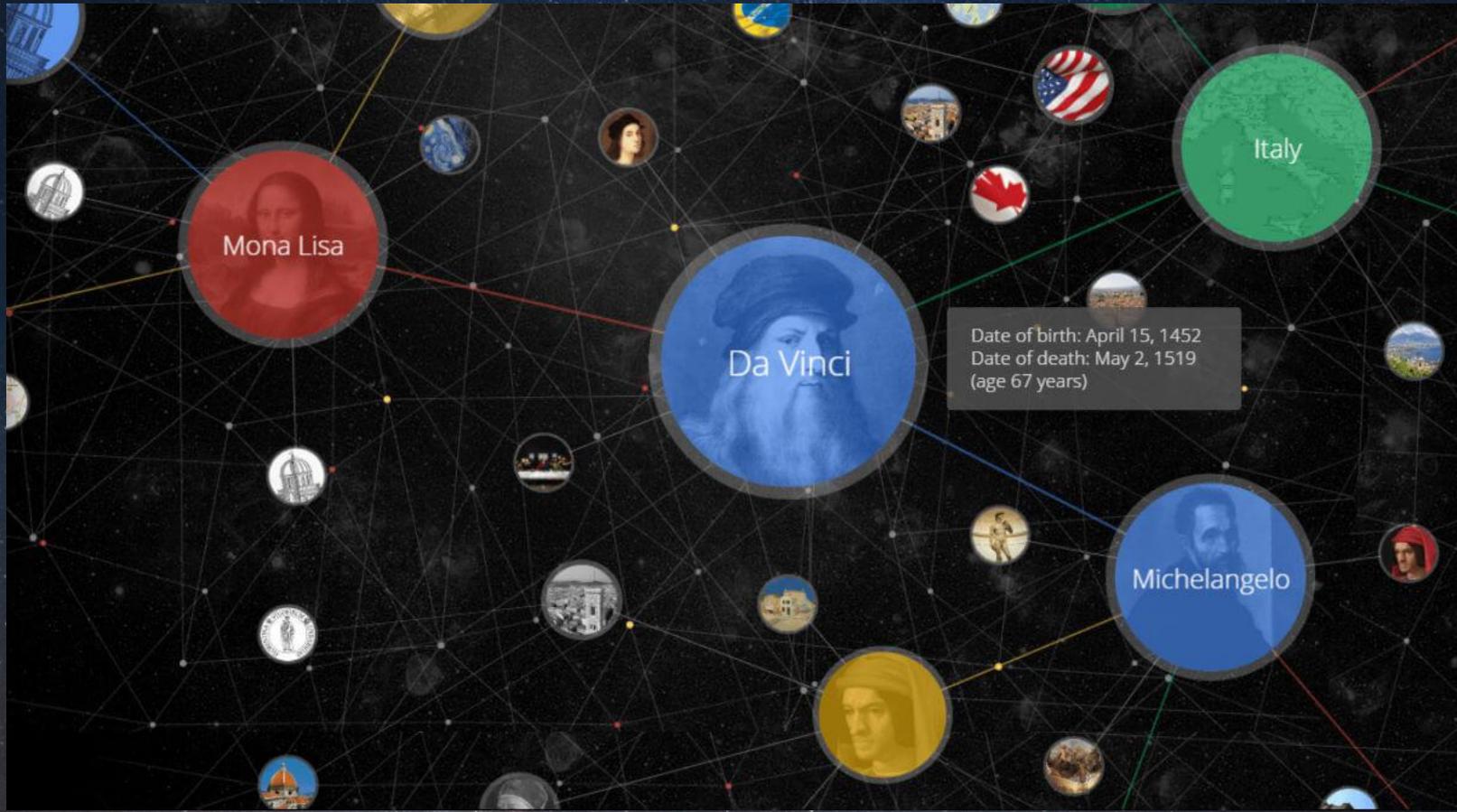
Leonardo da Vinci
Recognized entity



Leonardo da Vinci
Recognized entity
Related entities



Leonardo da Vinci
Recognized entity
Related entities
Named Relationship



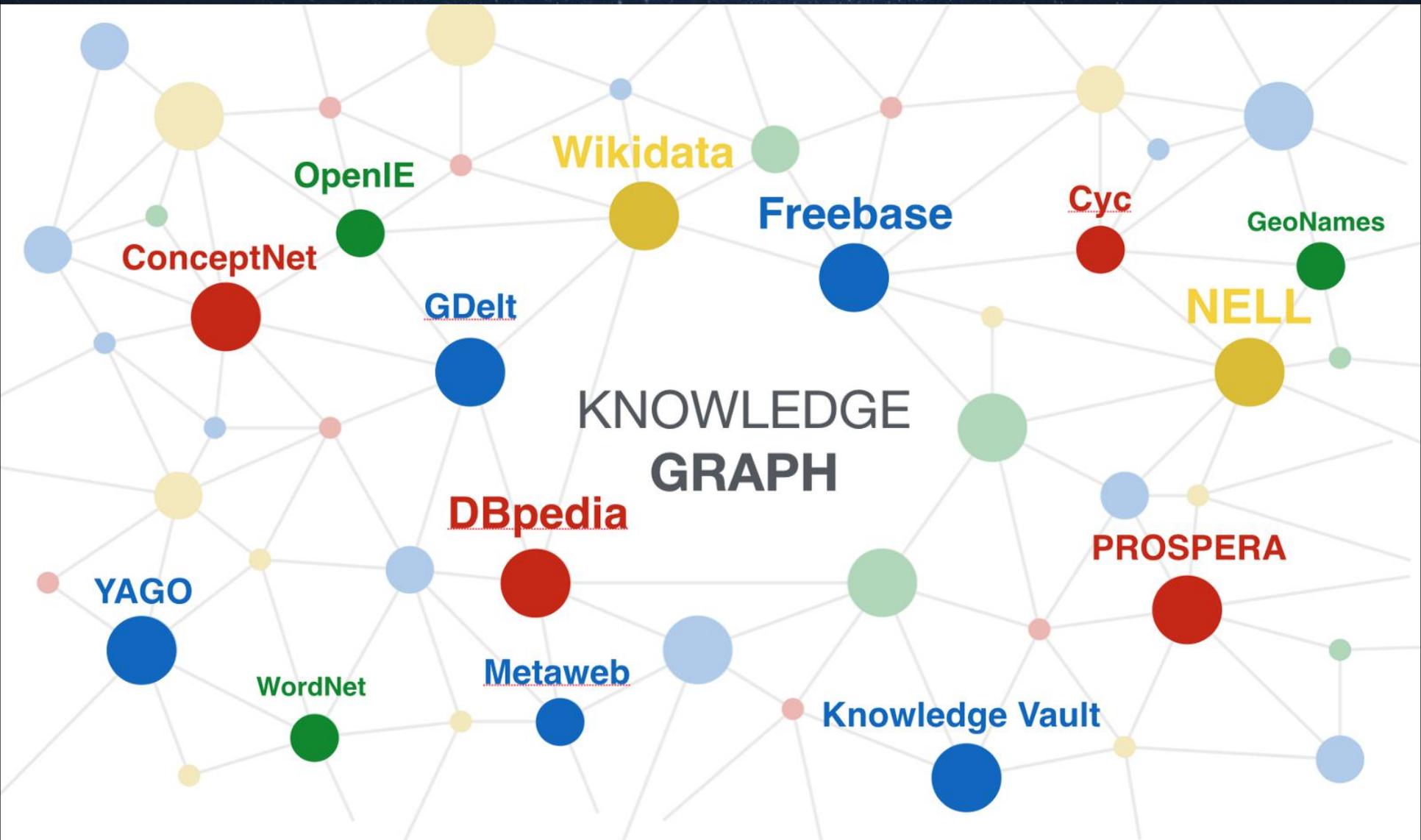
Mona Lisa

Da Vinci

Italy

Michelangelo

Date of birth: April 15, 1452
Date of death: May 2, 1519
(age 67 years)



	主要目的和用途	模型与技术	典型应用领域
语义网络 Semantic Network	表达概念间的关系 用于知识表示	节点——关系——节点	语词学 知识工程 早期人工智能（剽窃检测）
语义万维网 Semantic Web	知识表示、知识链接、数据开放、逻辑推理	<ul style="list-style-type: none"> • RDF • RDFs、OWL • Unifying Logic、Rule、Proof、Trust 	语义代理 语义搜索\推理 E-learning
关联数据 Linked Data	知识组织 知识链接 数据开放	<ul style="list-style-type: none"> • 元数据、本体、RDF、RDFa、Micro Format（Raw Data） • Web 技术（Http, URI, Rest... ..） • RDF Store、SPARQL 	政府、出版界、图情档博等文化继承机构
知识图谱 Knowledge Graph	知识组织 推理 职能问答 知识创新	<ul style="list-style-type: none"> • 本体、RDF、各种图数据模型 • 自然语言处理技术、规则系统、机器学习 • Graph DB、NoSQL, 专用查询语言 	搜索引擎 金融 医疗、生命科学

核心技术

RDF（资源描述框架）

Ontology（知识本体）

Graph Data Store（图数据存储）

2. 本体的概念

The background of the slide is a deep blue and black starry night sky. A prominent, vertical streak of light, likely a galaxy or nebula, runs through the center of the image, showing a gradient from dark blue at the top to a warmer, reddish-brown glow at the bottom. Numerous individual stars of varying brightness are scattered across the entire field of view.

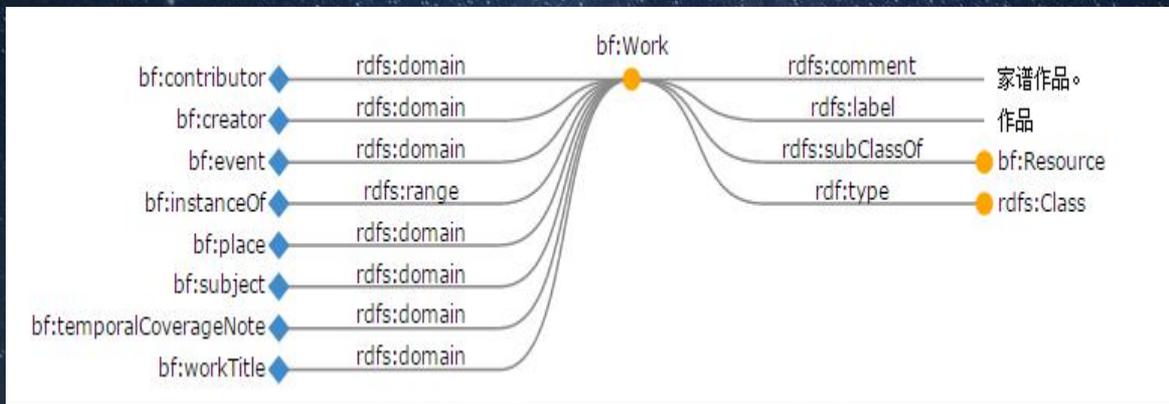
作品

机构

人

地

领域的
可共享的
概念模型



术语词表
类、属性 (关系)

RDF Schema 1.1

SKOS

OWL

形式化的

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix bf: <http://bibframe.org/vocab/> .
bf:Work a rdfs:Class ;
    rdfs:label "作品" ;
    rdfs:subClassOf bf:Resource ;
    rdfs:comment "家谱作品。" .
```

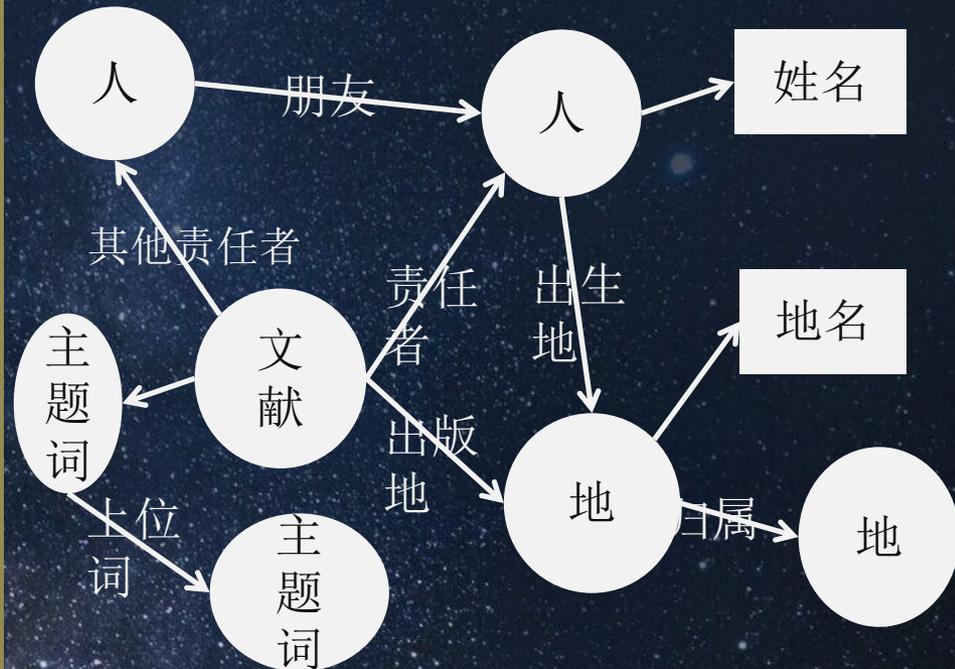
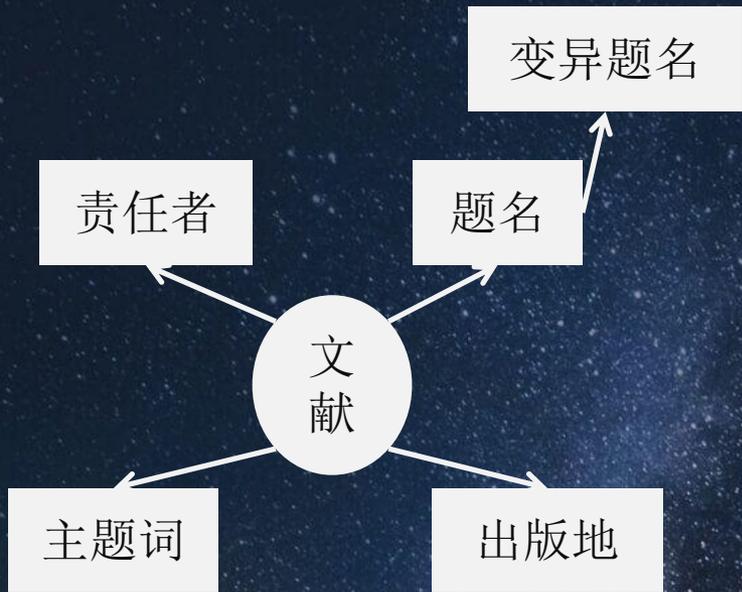
本体(Ontology)是对领域知识进行抽象的、形式化的概念模型。

3. 文献的知识组织



元数据方法和知识本体方法

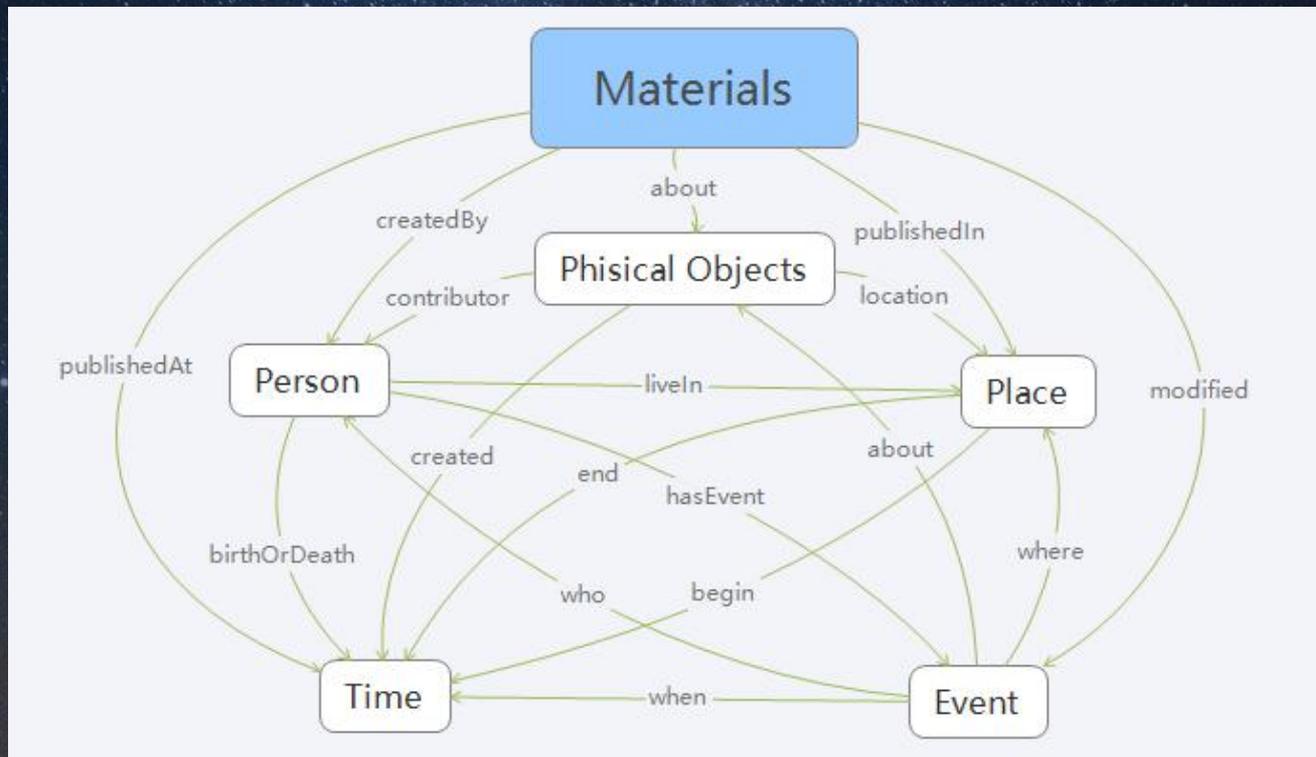
		元数据	知识本体
相同点		对信息的结构化描述	
不同点	定义及构成	“关于数据的数据。” 元数据方案（元素集）、置 标方案、元数据记录	“领域的可共享的概念及概念间 关系的形式化定义。” 模型、词表（类、属性）、实体/ 实例对象
	描述对象	文献	内容
	描述重点	文献的外在特征	内容对象的特征及其相关关系
	词表定义	元素	类、属性
	标准规范	MARC DC	FRBR BIBFRAME FOAF schema.org
	编码	RDF XML	RDFs OWL
	存储与查询	RDB SQL	RDF Store Graph DB SPARQL
联系		元数据元素可作为本体的属性，本体可看做“元”元数据。	



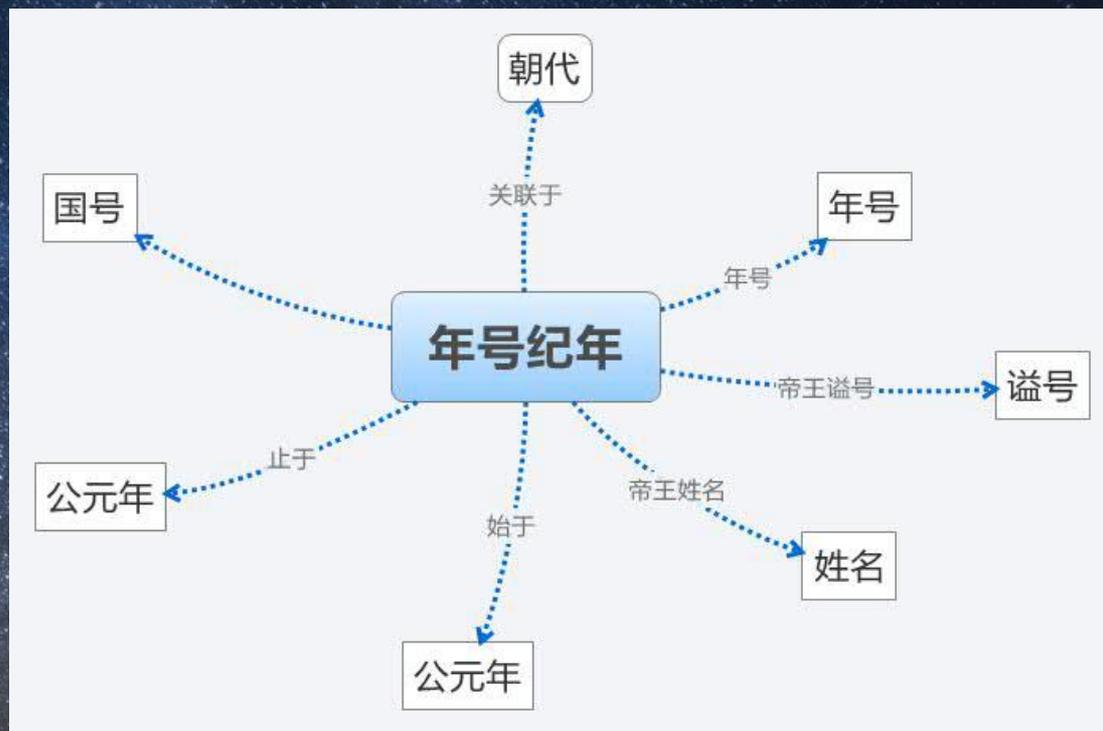
地名	序号	属性	古今属性	简称	曾用名	现名	别名	时代	位置
艾	(1)	古邑名	历史地名					春秋	今山东省新
	(2)	古邑名	历史地名					春秋	今江西省修
艾比湖	(1)	湖名	现今地名				库尔湖、布尔哈齐湖		新疆维吾尔
艾城镇	(1)	乡镇名	现今地名						江西省永修
艾丁湖	(1)	湖名	现今地名				觉洛浣		新疆维吾尔
艾浑	(1)								
艾家镇	(1)	城镇名	现今地名						湖北省宜昌
艾兰湖	(1)	湖名							
艾力西湖	(1)	乡镇名	现今地名						新疆维吾尔
艾陵	(1)	古邑名	历史地名					春秋	今山东省东
艾陵湖	(1)	古湖名	历史地名						今江苏省淮
艾门扎拉	(1)	区片名	现今地名						内蒙古自治
艾山	(1)	山名	现今地名						山东省胶东
艾塘	(1)	古地名	历史地名						今江苏省淮



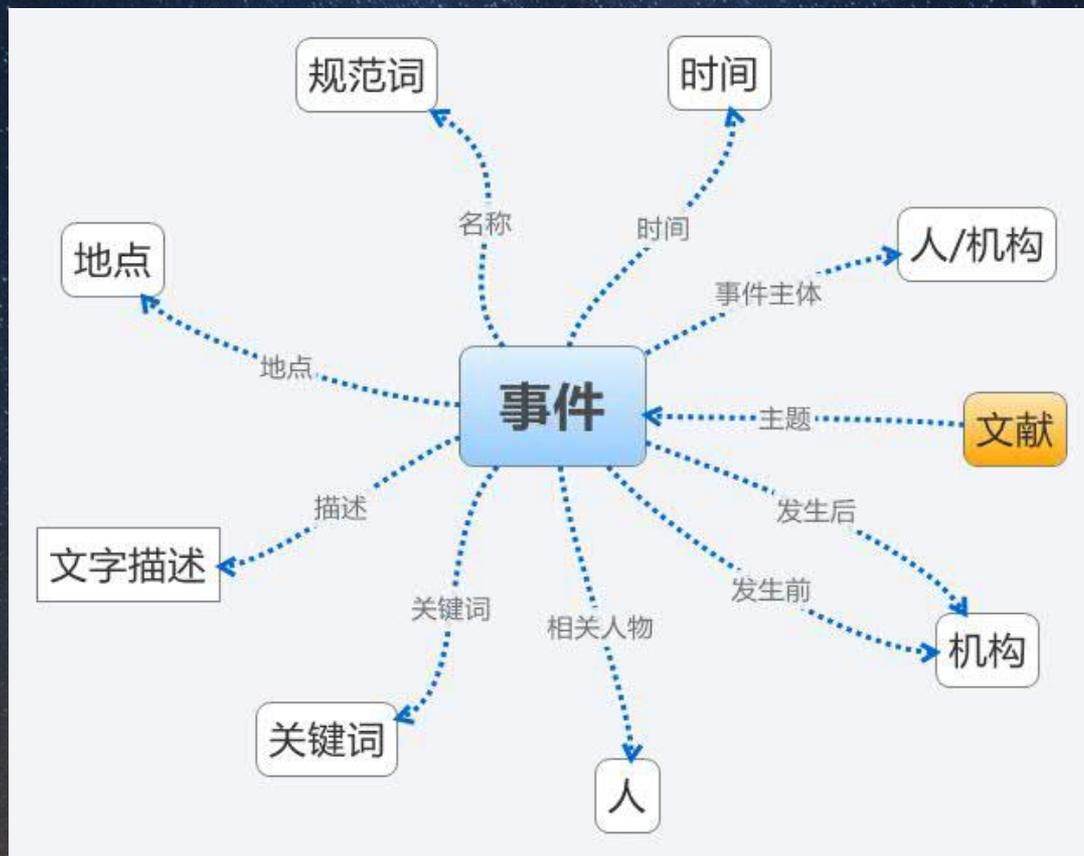
图档博抽象模型



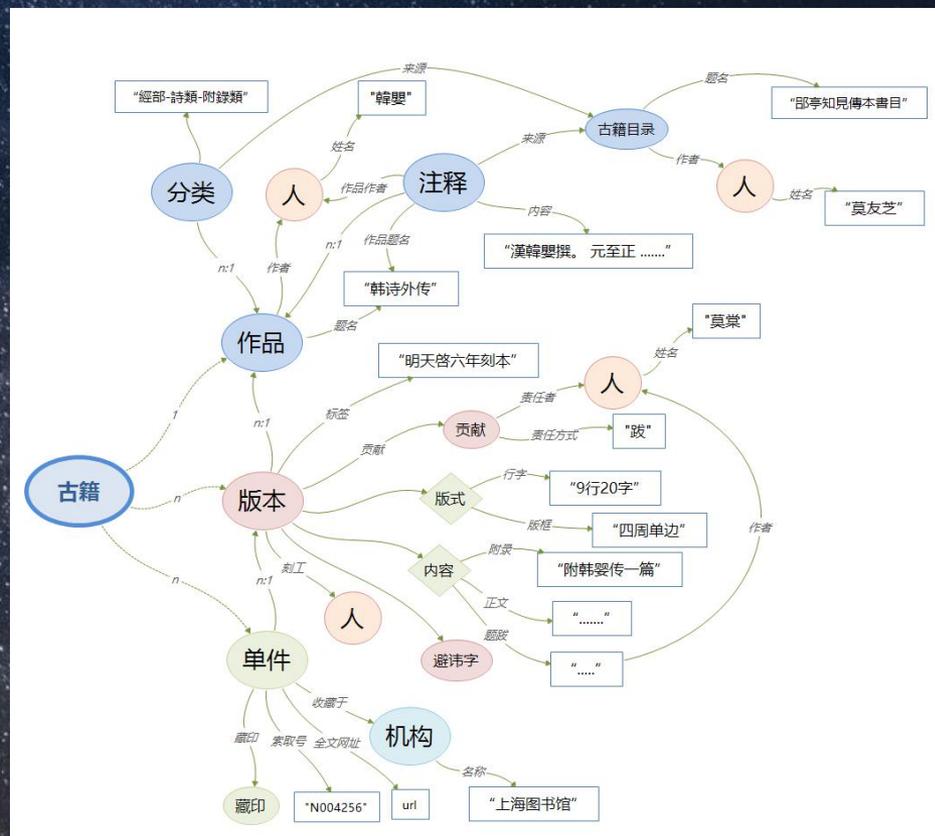
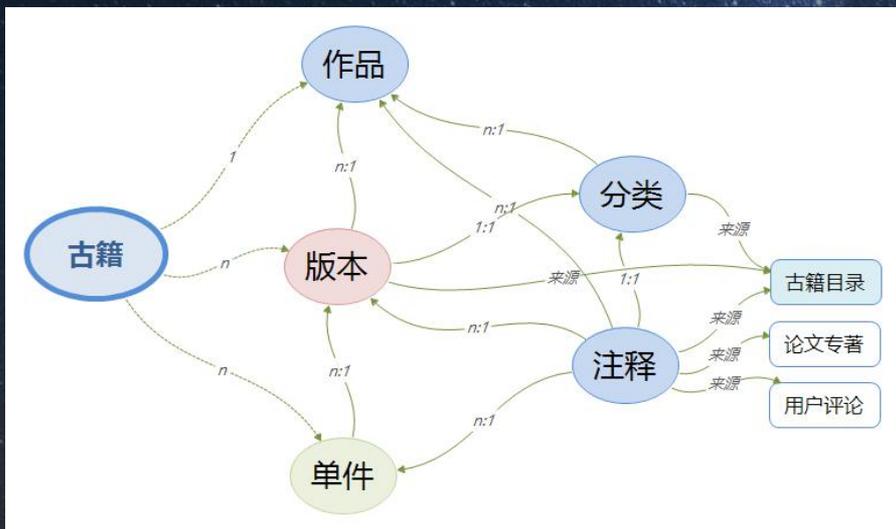
时间的本体



事件的本体



文献的本体



本体词表的编码和发布

本体 (属性)				
property	label	subPropertyOf	domain	range
bf:identifiedBy	标识符		bf:Resource	bf:Identifier
rdfs:label	标签		rdfs:Resource	Literal
dc:title	题名		shl:Resource	Literal
bf:title	题名		bf:Work	bf>Title
dc:creator	责任者		shl:Resource	Literal
bf:creator	责任者		bf:Work	bf:Agent
dc:contributor	其他责任者		shl:Resource	shl:Agent
bf:contributor	其他责任者		bf:Work	bf:Agent
bf:subject	主题词		bf:Work	bf:Topic
bf:classification	分类		bf:Work, bf:Instance, bf:Item	bf:Classification
pmb:classification	分类		pmb:Work, pmb:Instance, pmb:Item	Literal
shl:temporal	时间		shl:Resource	shl:Temporal
shl:temporalValue	时间		shl:Resource	Literal
shl:place	地点		shl:Resource	shl:Place
shl:placeValue	地点		shl:Resource	Literal
shl:description	描述		shl:Resource	Literal
pmb:hasAnnotation	注释		pmb:Work	pmb:Annotation
oa:annotatedAt	注释时间		oa:Annotation	shl:Temporal
oa:annotatedBy	注释责任者		oa:Annotation	foaf:Agent
oa:hasBody	注释内容		oa:Annotation	Literal
bf:hasInstance	实例		bf:Work	bf:Instance
bf:instanceOf	作品		bf:Instance	bf:Work
bf:extent	大小		bf:Instance	Literal
bf:edition	版本类型		bf:Instance	bf:Category
bf:dimensions	容量		bf:Instance	Literal
bf:hasItem	单件		bf:Instance	bf:Item
bf:itemOf	实例		bf:Item	bf:Instance
dc:abstract	摘要		shl:Resource	Literal
pmb:font	字体		pmb:Instance	bf:Category

```

@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>.
@prefix geovocab: <http://geovocab.org/spatial#>.
@prefix shl: <http://www.library.sh.cn/ontology/>.
@prefix rel: <http://purl.org/vocab/relationship/>.
@prefix org: <http://www.w3.org/ns/org#>.
@prefix edm: <http://www.europeana.eu/schemas/edm/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix prov: <http://www.w3.org/ns/prov#>.
@prefix oa: <http://www.w3.org/ns/oa#>.
@prefix pmb: <http://pmb.library.sh.cn/ontology/>.

```

```

bf:Contribution a rdfs:Class ;
  rdfs:label "贡献" ;
  rdfs:subClassOf bf:Resource ;
  rdfs:comment "人或机构对所描述的资源所作的贡献。" .

```

```

shl:BookPlate a rdfs:Class ;
  rdfs:label "藏书票" ;
  rdfs:subClassOf shl:Resource ;
  rdfs:comment "藏书票相关信息。" .

```

```

shl:Envelope a rdfs:Class ;
  rdfs:label "信封" ;
  rdfs:subClassOf shl:Resource ;
  rdfs:comment "信封相关信息。" .

```

```

shl:Inscription a rdfs:Class ;
  rdfs:label "题记" ;
  rdfs:subClassOf shl:Resource ;
  rdfs:comment "题记相关信息。" .

```

```

shl:LetterPaper a rdfs:Class ;
  rdfs:label "信纸" ;
  rdfs:subClassOf shl:Resource ;
  rdfs:comment "信纸相关信息。" .

```

```

shl:MediaCharacteristic a rdfs:Class ;
  rdfs:label "多媒体特征" ;
  rdfs:subClassOf shl:Resource ;
  rdfs:comment "多媒体特征相关信息。" .

```

基于本体的知识组织

知识编码的形式化——机器可读
知识单元的细粒度化——机器可计算
知识表示的语义化——机器可理解
知识组织的关联化——机器可推理
知识增长的自动化——机器可自学习



知识呈现的可视化——发现、探索、理解、质疑、交流、修正、成长……

如何设计一个本体？

05 本体形式化

以机器可读的格式编码。并在网上公开发布。

03 术语的复用

注意属性的Domain和Range，在复用时不要改变原有术语的外延和内涵。

01 实体关系分析

词表设计

类、属性、关系的定义。尽量复用，必要时自定义。关键在于定义好属性的Domain和Range，一个类同时继承不同的类。

04

02 参考模型/词表调研

调研与需求领域相关的模型和词表有哪些？





谢谢THANKS!

xtykc@yeah.net