



文本挖掘技术(2009)

第十三章：

本体 (Ontology)

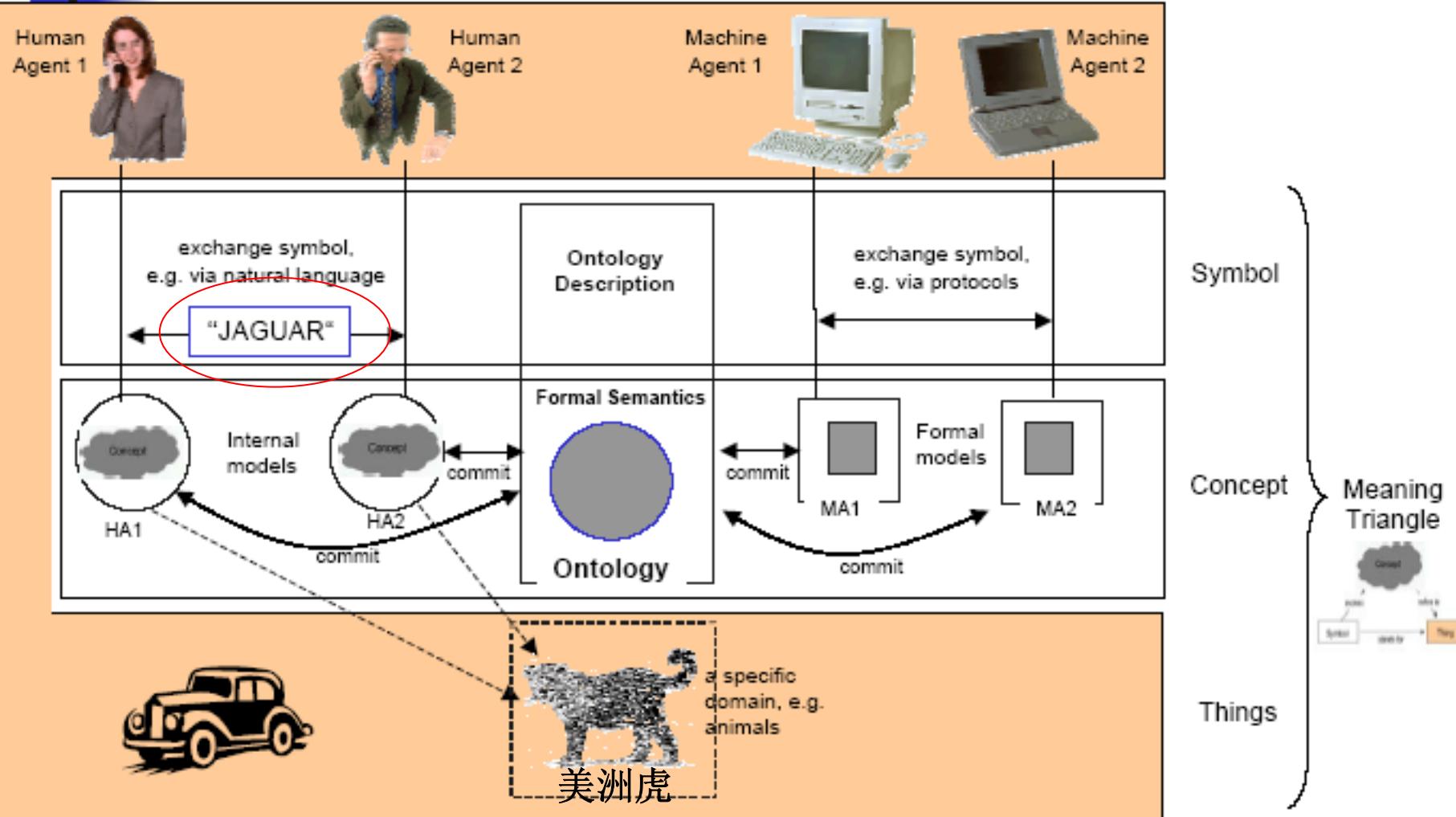
杨建武

北京大学计算机科学技术研究所

Email:yangjianwu@icst.pku.edu.cn

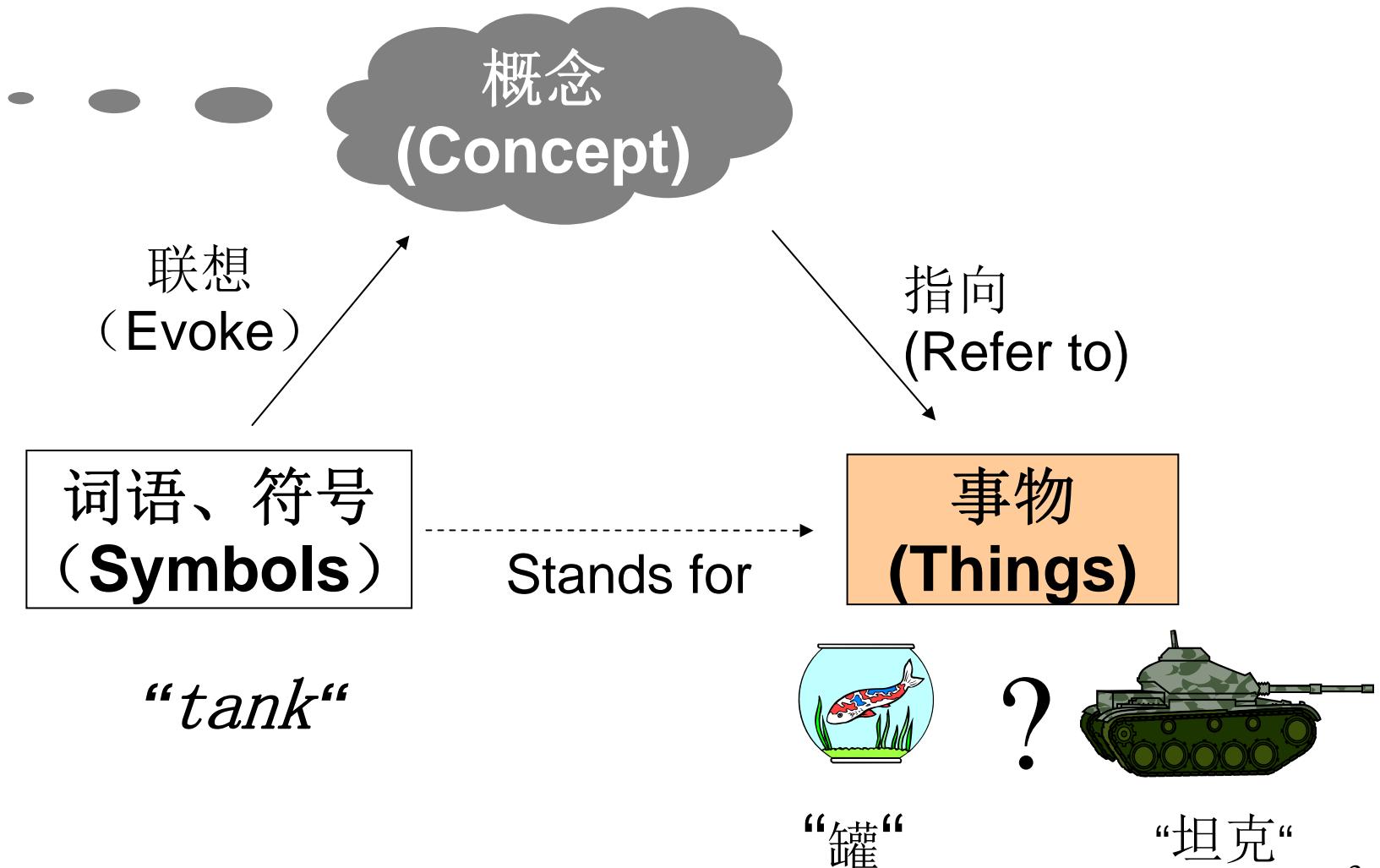


信息交流





对内容的理解





关于Ontology的译名和定义

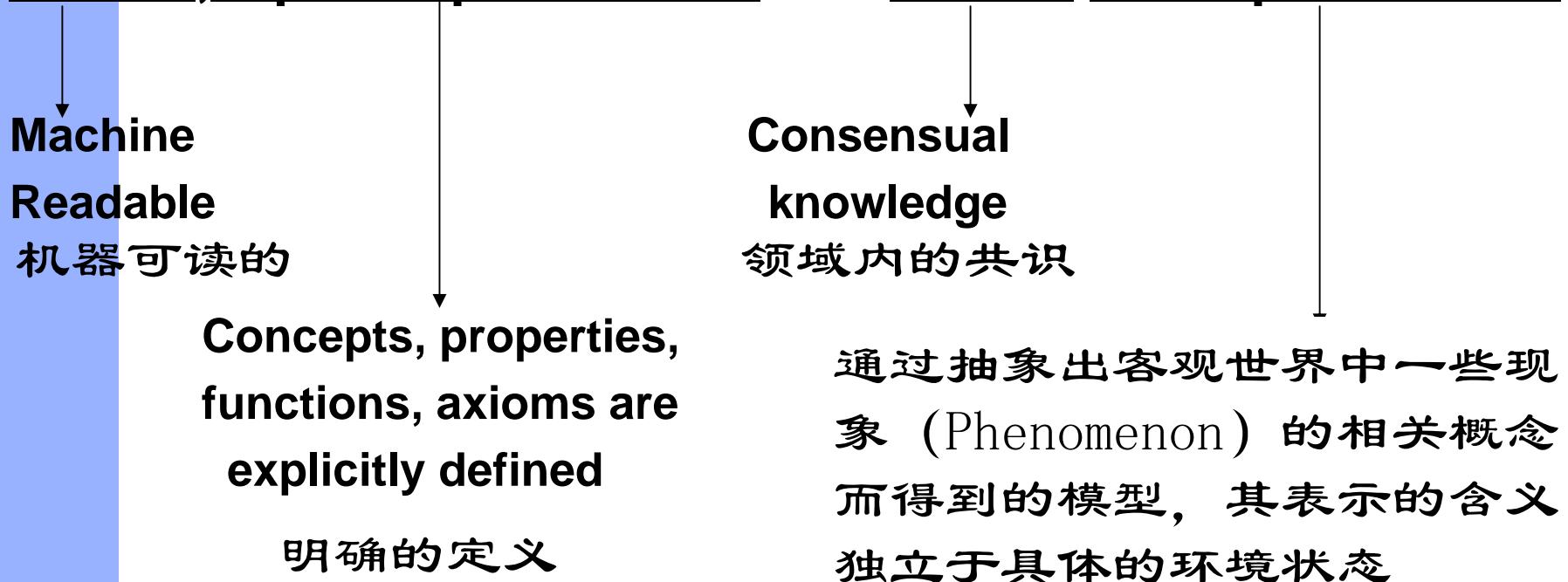
- 哲学领域
 - ❖ ontology是运用以**存在(Being)**为核心的各种范畴通过逻辑的方法去构造哲学原理的学问。
 - ❖ 它是在概念王国里的纯粹的**思辨哲学**。
 - ❖ **本体论** 存在论 是论
- 人工智能领域
 - ❖ 1991/Neches等：给出构成相关领域词汇的基本**术语**和**关系**，以及利用这些术语和关系构成的规定这些词汇外延的规则的**定义**；
 - ❖ 1993/Gruber：**概念模型**的明确的规范说明；
 - ❖ 1997/Borst：共享概念模型的形式化规范说明；
 - ❖ 1998/Studer：共享概念模型的明确的形式化规范说明



本体的定义

Studer (98) 共享概念模型的明确的形式化规范说明

Formal, explicit specification of a shared conceptualization



提供一套概念和术语来描述某一领域，并且获取该领域的本质的概念结构。



为什么需要Ontology?

- 知识**分享**的问题：不同的系统使用不同的概念和术语。很难从一个系统中提取知识运用到另一个系统中。
- 由于对跨系统的知识重用和分享的兴趣的增大，导致对Ontology的兴趣大增。
- 开发可**重用**的Ontology以推进共享和重用是Ontology研究的一个重要目标。
- 开发支持构建、合并和翻译Ontology的**工具**是研究的另一个目标。



构建Ontology的目标

➤ 构建Ontology的目标

- ❖ 捕获相关的领域的知识，
- ❖ 提供对该领域知识的共同理解，
- ❖ 确定该领域内共同认可的词汇，
- ❖ 从不同层次的形式化模式上给出这些词汇（术语）和词汇之间相互关系的明确定义。



明确的Ontology能够支持

- 在人们之间共享理解
- 工具间的互操作性
- 系统工程
- 可重用性
- 定义主题词表和概念
- 定义必要的充足的概念和“方法”
- 共同的知识理论
- 扮演数学在物理中所扮演的角色（内容理论）



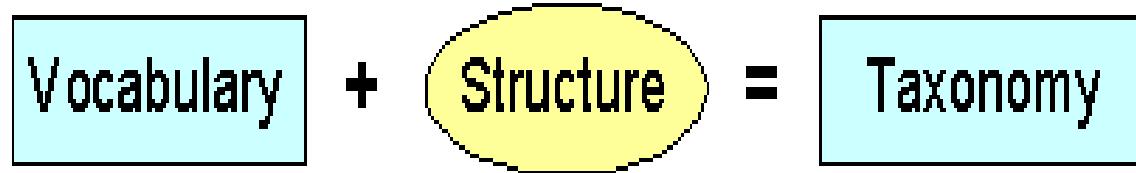
Ontology : 对现实世界的描述

- 世界存在对象 (Object)
- 对象可以抽象出类 (Class)
- 对象具有属性 (Property/Attribute) , 属性可以赋值 (Value)
- 对象之间存在着不同的关系 (Relation)
- 对象可以分解成部分 (Part)
- 对象具有不同的状态 (State)
- 属性和关系随着时间推移而改变
- 不同时刻会有不同的事件 (Event) 发生
- 事件能导致其他事件发生或状态改变
- 在一定的时间段上存在着过程 (Process) , 对象则参与到过程之中



本体的构成

- 本体 = 概念 + 属性 + 公理 + 取值 + 名义
- 本体 = 概念类 + 关系 + 函数 + 公理 + 实例





Ontology与知识库的关系

- Ontology为知识库的建立提供一个基本的结构；
- Ontology提供一套概念和术语来描述某一领域，并且获取该领域本质的概念结构；
- 知识库运用这些术语去表达现实或者虚拟世界中的正确知识。



本体的分类

- 顶级本体
 - ❖ 描述最普遍的概念及概念之间的关系，如空间、时间、事件、行为等等，
 - ❖ 与具体的应用无关，其他种类的本体都是该类本体的特例。
- 领域本体
 - ❖ 特定领域(如：医药)中的概念及概念之间的关系。
- 任务本体
 - ❖ 特定任务或行为中的概念及概念之间的关系。
- 应用本体
 - ❖ 依赖于特定领域和任务的概念及概念之间的关系。



Semantic Web 与 OWL



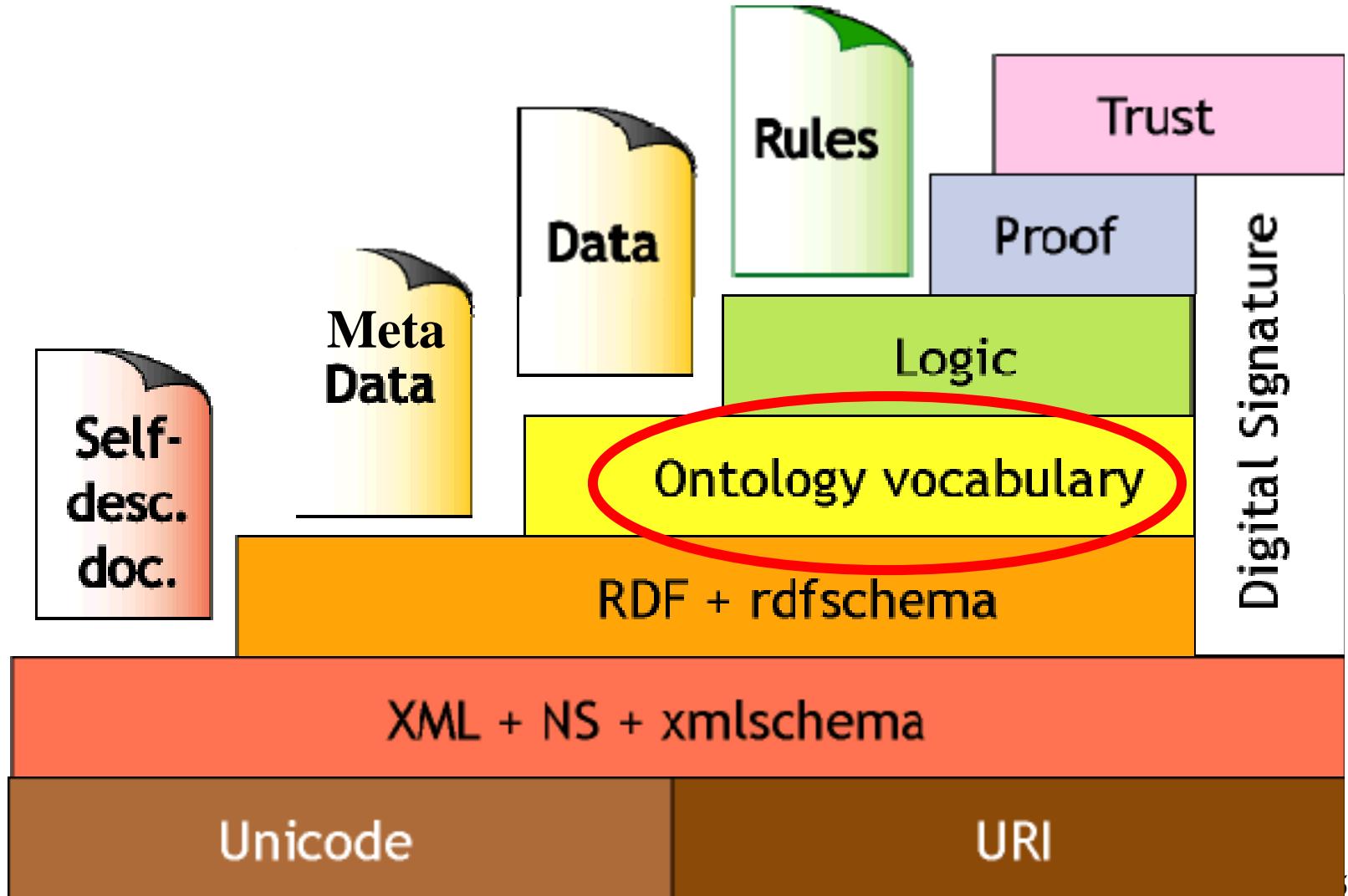
什么是语义网

- 不同于： Semantic Network（语义网络）
- Web现状：
 - ❖ 网上的信息表达一般是以
 - 自然语言(中文、英语、匈牙利语...)
 - 图表、多媒体、版面格式
 - ❖ 人类可以使用, 但机器难以使用
- The Semantic Web（语义网）
 - ❖ Tim Berners-Lee, www 2000会议(2000.12)
 - ❖ 语义网是当前万维网的扩展与延伸
 - ❖ 语义网的信息是机器可理解的，具有充分完备的语义定义，能促进人与计算机建立语义上的合作

机器可读→机器可理解



语义网体系结构



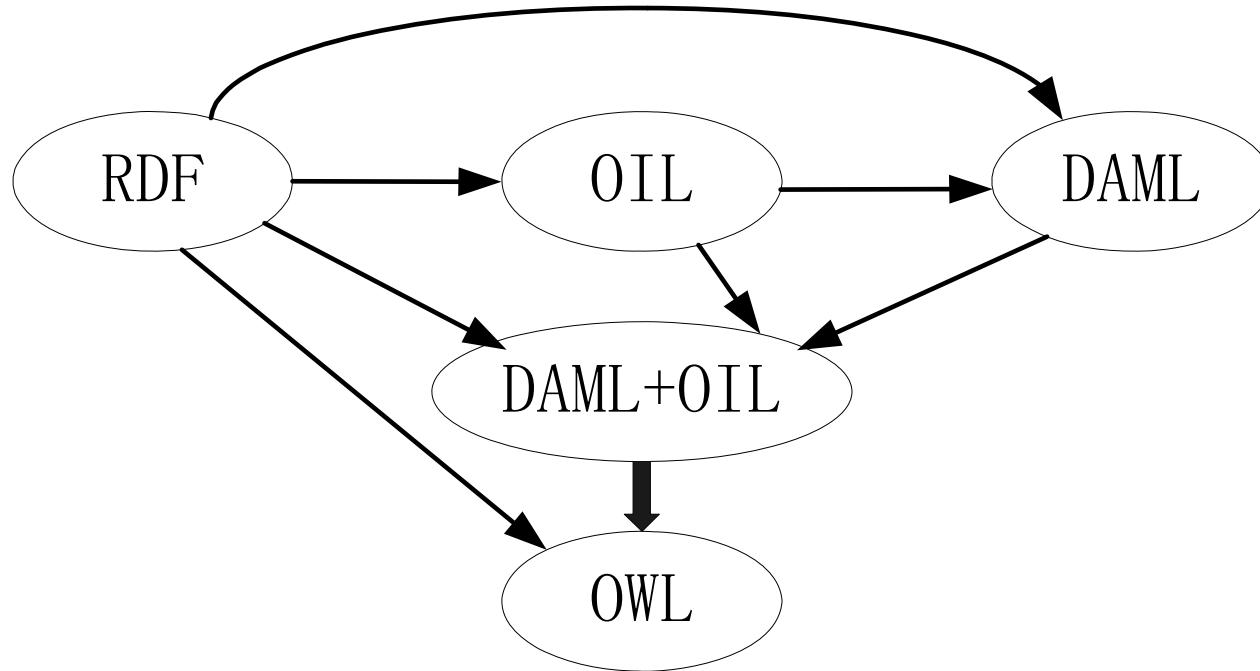


语义网体系结构

- 1. Unicode和URI: 是整个语义网的基础, Unicode处理资源的编码; URI负责标识资源。
- 2. XML + Name Space + XML Schema: 表示数据的内容和结构。
- 3. RDF+RDFSschema, 描述资源及其类型。
- 4. 本体层(OWL) : 描述资源之间的联系。
- 5. 逻辑层: 公理和推理规则
- 6. 证明层: 执行规则、评判信赖
- 7. 信任层: 信任机制



本体描述语言的演变



- OWL: Ontology Web Language
 - ❖ 2004年2月由W3C正式推出
 - ❖ A language for defining and instantiating Web ontologies



知识表达

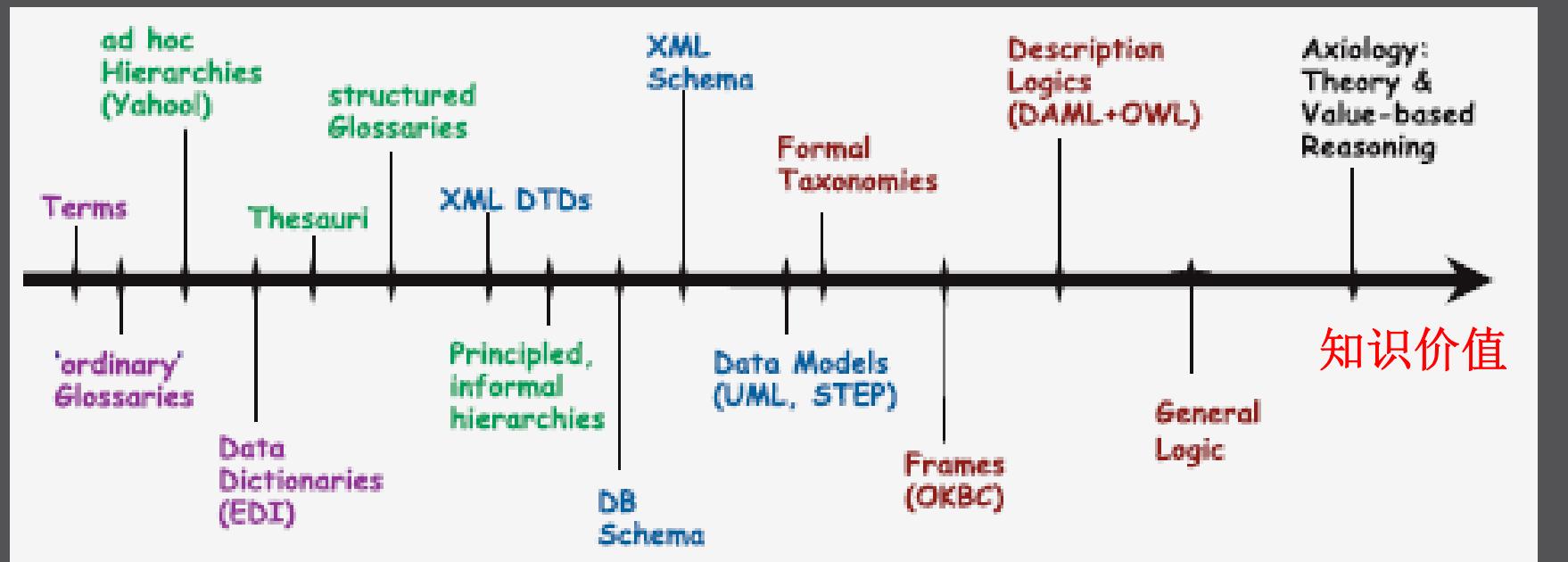
术语表和
数据字典

辞典与
分类法

元数据与
数据模型

形式化本
体与推理

语义形式的
陈述性知识

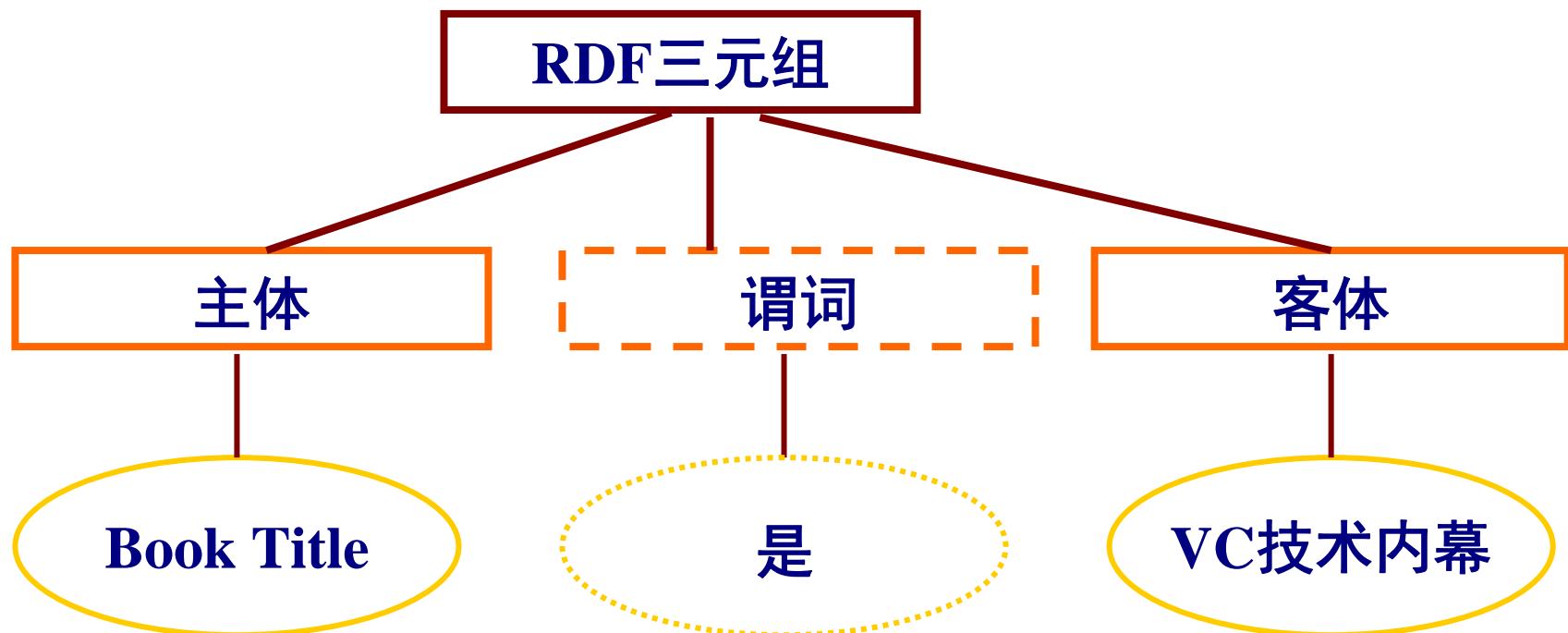


Source: Craig Schlenoff, NIST & Project10X



RDF 概念

- RDF：通过属性与值来描述资源以及资源之间的关系，三元组：“主体-谓词-客体”





RDF 概念

➤ 举例：

```
<rdf : Description about='http://www.my.com/book.html'>  
    <Book Title> VC技术内幕 </Book Title>  
    <Home-Page rdf:resource='http://www.my.com/'/>  
</rdf: Description>
```

被描述资源的URI

被描述资源中有一个属性叫Book Title（即书名），其取值是VC技术内幕

结束标志

被描述资源中有一个属性叫Home-Page（即主页），其值指向另一资源



从 RDF/RDFS到OWL

- RDFS：提出了类的概念，定义了类和性质并可以用来描述其他的类和性质
- RDFS 描述资源的能力还不够强大
- RDFS 难以提供推理方面的支持



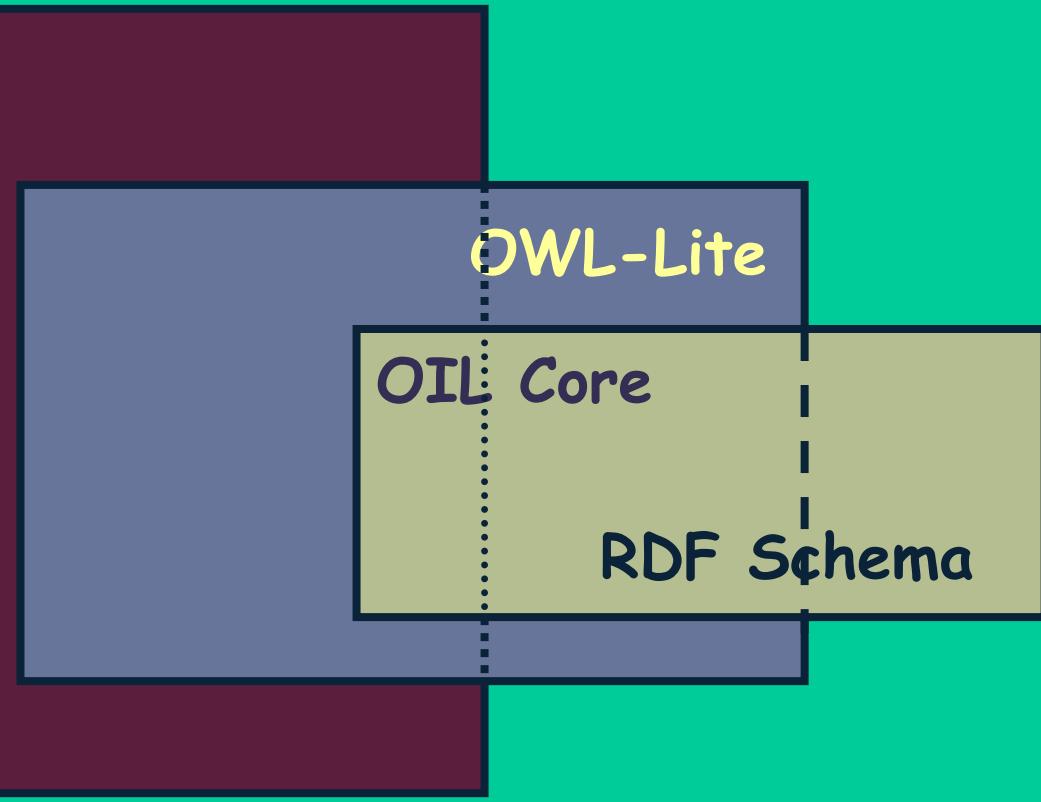
OWL的三个子集

- OWL full
 - ❖ OWL 语法与RDF的合集，完备的本体语言
- OWL DL
 - ❖ 倾向于逻辑描述，表达能力强，
 - ❖ 忽略了对RDFS的兼容性
- OWL Lite
 - ❖ 降低了OWL DL中的公理约束
 - ❖ 保证一个高效的推理过程
- <http://www.w3.org/TR/owl-features/>



OWL Full \approx DAML+OIL \approx Instance OIL

OWL - DL





OWL Lite 语言构造子

- RDFSchema特征：
 - ❖ Class; rdfs:subClassOf; rdf:Property;
rdfs:subPropertyOf; rdfs:domain; rdfs:range; Indivual;
- 等同性：
 - ❖ equivalentClass; equivalentProperty; sameAs;
differentFrom; AllDifferent; distinctMembers;
- 属性特征：
 - ❖ ObjectProperty; DatatypeProperty; inverseOf;
TransitiveProperty; SymmetricProperty;
FunctionalProperty; InverseFunctionalProperty
- 属性限制
 - ❖ Restriction; onProperty; allValuesFrom;
someValuesFrom



OWL Lite 语言构造子

- 基数限制：
 - ❖ minCardinality; maxCardinality; cardinality(0/1)
- 类相交：
 - ❖ intersectionOf
- 头信息：
 - ❖ Ontology; imports
- 版本信息：
 - ❖ versionInfo; priorVersion; backwardCompatibleWith; incompatibleWith; DeprecatedClass; DeprecatedProperty
- 评注性质：
 - ❖ rdfs:label; rdfs:comment; rdfs:seeAlso; rdfs:isDefinedBy; AnnotationProperty; OntologyProperty



OWL DL/FULL扩展的构造子

- 类公理：
 - ❖ oneOf; dataRange; disjointWith;
equivalentClass; rdfs:subClassOf
- 类表达式的布尔组合：
 - ❖ unionOf; complementOf; intersectionOf
- 任意基数：
 - ❖ minCardinality; maxCardinality; cardinality
- 填充信息：
 - ❖ hasValue



RDFS 相关特征

- Class, 类, 拥有共同性质的个体集合
 - ❖ OWL:Thing; OWL:Nothing
- rdfs:subClassOf, 子类, 类间层次关系
- Individual, 个体, 类的实例
- rdf:Property, 属性,
 - ❖ owl:ObjectProperty:个体间二元关系
 - ❖ owl:DatatypeProperty:个体到数值的二元关系
- rdfs:subPropertyOf, 子属性, 属性间层次关系
- Rdfs:domain; rdfs:range, 定义域、值域



等价与不等价

- equivalentClass, 两个类等价
- equivalentProperty, 两个属性等价
- sameAs, 两个个体相同
- differentFrom, 两个个体不同
- AllDifferent, 多个个体互不相同



属性特征（性质类型）

- TransitiveProperty, 性质具传递性
 - ❖ $P(x,y) \& P(y,z) \rightarrow P(x,z)$
- SymmetricProperty, 性质具对称性
 - ❖ $P(x,y) \rightarrow P(y,x)$
- FunctionalProperty, 性质具函数性
 - ❖ $P(x,y) \& P(x,z) \rightarrow y=z$
- inverseOf, 两性质互逆
 - ❖ $P1(x,y) \rightarrow P2(y,x)$
- InverseFunctionalProperty, 性质具反函数性
 - ❖ $P(y,x) \& P(z,x) \rightarrow y=z$



属性限制与基数限制

- allValueFrom, 属性取值来源于指定的类
 - someValueFrom, 至少一个来源于指定类
-
- Candinality, 基数, 取值个数
 - maxCandinality, 最多取值个数
 - minCandinality, 最少取值个数

OWL DL / OWL FULL 增值特征



- oneOf, 枚举, 枚举个体对类进行描述
- hasValue, 相关个体中至少一个的值等于指定的值
- disjointWith, 类与类之间互斥, 没有公共元素
- unionOf, complementOf, intersectionOf, 布尔组合, 形成新的类或属性
- Cardinality, maxCardinality, minCardinality,
 - ❖ OWL Lite中, 基数被限定为只能是0或1;
 - ❖ OWL DL和OWL FULL中基数允许为任意非负数



OWL语义小结

- 概念及其关系的分类化、层次化结构
 - ❖ subClassOf、subPropertyOf
- 概念间的同义、反义等语义关系
 - ❖ sameClassAs、samePropertyAs、inverseOf、equivalentTo
- 概念间的逻辑组合关系
 - ❖ intersectionOf、unionOf、complementOf、oneOf
- 对关系约束进行描述
 - ❖ domain、range、toClass、hasValue、candinality、minCandinality、maxCandinality
- 对概念及其关系的公理的定义
 - ❖ disjointWith、uniqueProperty、transitiveProperty



XML(S), RDF(S), OWL

	XML DTD	XML Schema	RDF(S)	OWL
有界列表				X
基数约束	X	X		X
类表达式			X	X
数据类型		X	X	X
已定义的类				X
枚举	X	X		X
等价				X
可扩展性			X	X
形式化语义			X	X
继承			X	X
推理				X
本地化约束				X
条件约束				X
实例化			X	X

OWL XML/RDF Example - EyeGlass



```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE owl [
    <!ENTITY eyeglass "file:/G:/Glasses#">
    <!ENTITY xsd="http://www.w3.org/2000/10/XMLSchema#">
]>
<rdf:RDF
    xmlns="file:/G:/myclasses#"
    xmlns:eyeglass="file:/G:/Glasses#"
    xmlns:owl="http://www.w3.org/2003/02/owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#">
    <owl:Ontology rdf:about="file:/G:/myclasses.owl#"> // Ontology Headers
        <rdfs:comment>An EyeGlass OWL Ontology designed by COMP691B group;  

        last modified at March 10, 2003</rdfs:comment>
        <owl:priorVersion rdf:resource="file:/G:/myclasses01.owl#" />
    </owl:Ontology>
```

```
<owl:Class rdf:ID="Glasses">
    <owl:equivalentClass rdf:resource="#eyeglass;Glasses"/>
    <rdfs:label xml:lang="en">Glasses</rdfs:label> ... ...
</owl:Class>

<owl:Class rdf:ID="FramedGlasses"> //框式眼镜
    <rdf:subClassOf rdf:resource="#Glasses">
</owl:Class>

<owl:ObjectProperty rdf:ID="has-len"> //透镜
    <rdfs:domain rdf:resource="#Glasses">
    <rdfs:range rdf:resource="#Lens">
</owl:ObjectProperty>

<owl:Class rdf:ID="ContactLenses"> //隐形眼镜
    <owl:disjointWith rdf:resource="#FramedGlasses" />
    <owl:equivalentClass>
        <rdfs:subClassOf rdf:resource="#Glasses" />
        <rdfs:subClassOf>
            <owl:Restriction>
                <owl:onProperty rdf:resource="#has-len"/>
                <owl:minCardinality rdf:datatype="&xsd;NonNegativeInteger">2</owl:minCardinality> ...
            </owl:Restriction>
        </rdfs:subClassOf>
    </owl:equivalentClass>
</owl:Class>
```



支持OWL的工具

➤ Jena

- ❖ HP公司开发的，Java开发工具包
- ❖ Jena 2.1支持OWL

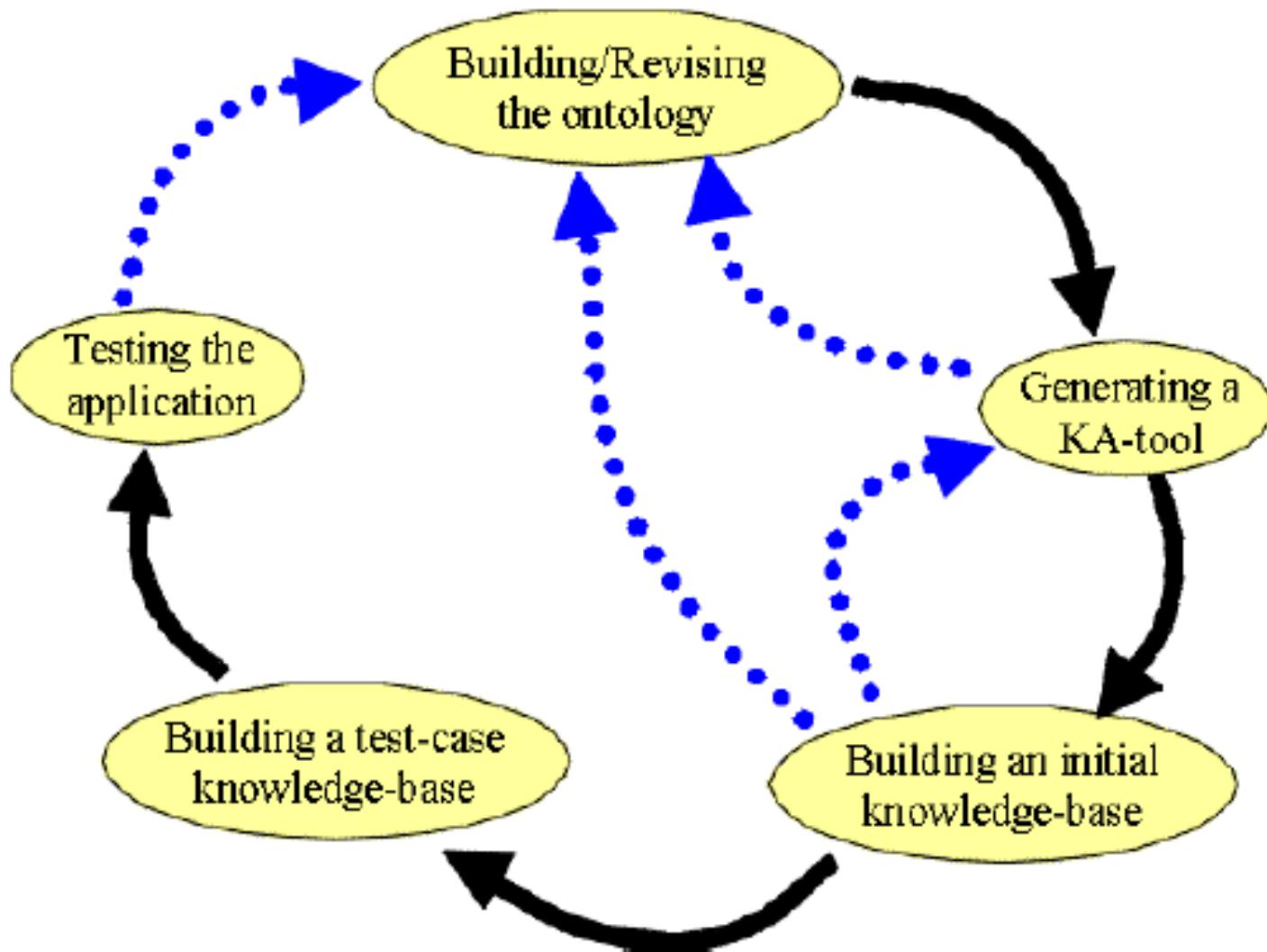
➤ Protégé-2000

- ❖ 美国斯坦福大学医学院采用Jena开发
- ❖ 使用Java和Open Source作为操作平台
- ❖ 可用于编制实用分类系统和知识库
- ❖ 图形化界面系统
- ❖ Protege 2.1支持OWL
- ❖ <http://protege.stanford.edu/>





Protégé-2000





Protégé 例子 newspaper

newspaper Protégé 2.1 (file:\E\Research\semantic\protege\samples\newspaper, OWL Files)

Project Edit Window OWL Wizards Code Help

OWLClasses Properties Forms Individuals Metadata

Subclass Relationship Article (type=owl:Class)

Asserted Hit Article

owl:Thing
Newspaper
Library
Content
Advertisement
Article
Layout_info
Billing_Chart
Content_Layout
Prototype_Newspaper
Rectangle
Section
Person
Employee
Columnist
Editor
Manager
Reporter
Salesperson
Author
Organization

Name Article

rdfs:comment
gardening could be submitted on Monday and
not be published until Thursday (when the
gardening section is included in the paper).

Annotations

D rdfs:comment Articles are included...

Asserted Inferred

NECESSARY & SUFFICIENT NECESSARY

Content

Properties

article_type
author
headline
keywords
reading_level
text
containing_section
expiration_date

Disjoints

Logic View Properties View



Protégé 例子 newspaper

newspaper Protégé 2.1 (file:\E\Research\semantic\protege\samples\newspaper, OWL Files)

Project Edit Window OWL Wizards Code Help

OWLClasses Properties Forms Individuals Metadata

Classes Display Slot

owl:Thing
Newspaper (8)
Library (1)
Content
Advertiseme
Article (1)
Layout_info
Person
Author
Columnist (1)
Editor (4)
News_Service
Reporter (4)
Organization (1)

S :NAME

Direct Inst:

OWL标准发布

owl:versionInfo
1.00

Annotations

Property	Value
owl:versionInfo	1.00
rdfs:label	N1

Expiration Date Page Number Article Type Layout

12/02/2004 2 News newspaper_

Headline Reading Level Published In

OWL标准发布 College newspaper_

Text Author Keywords

标准，各研究机构纷纷表示祝贺！ Jeff www标准语义网

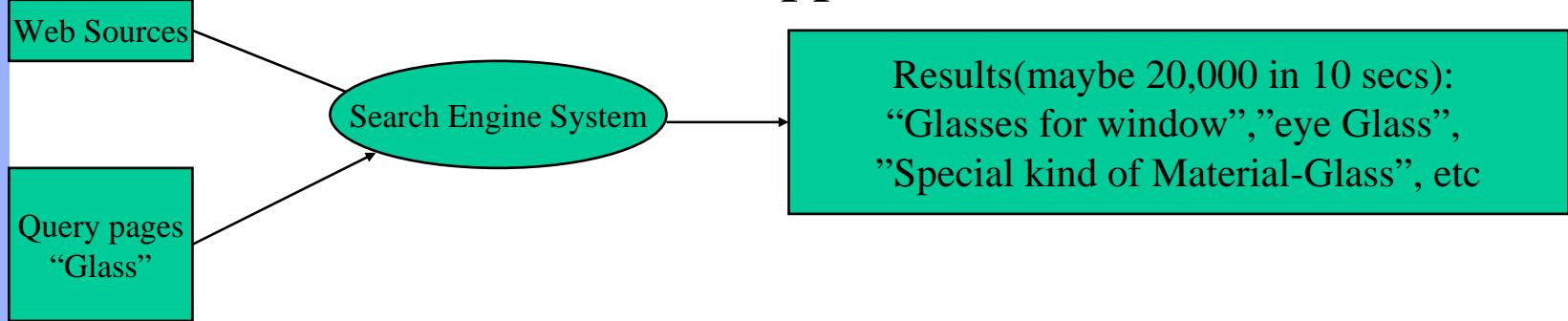
Urgent Containing Section

instance_00019

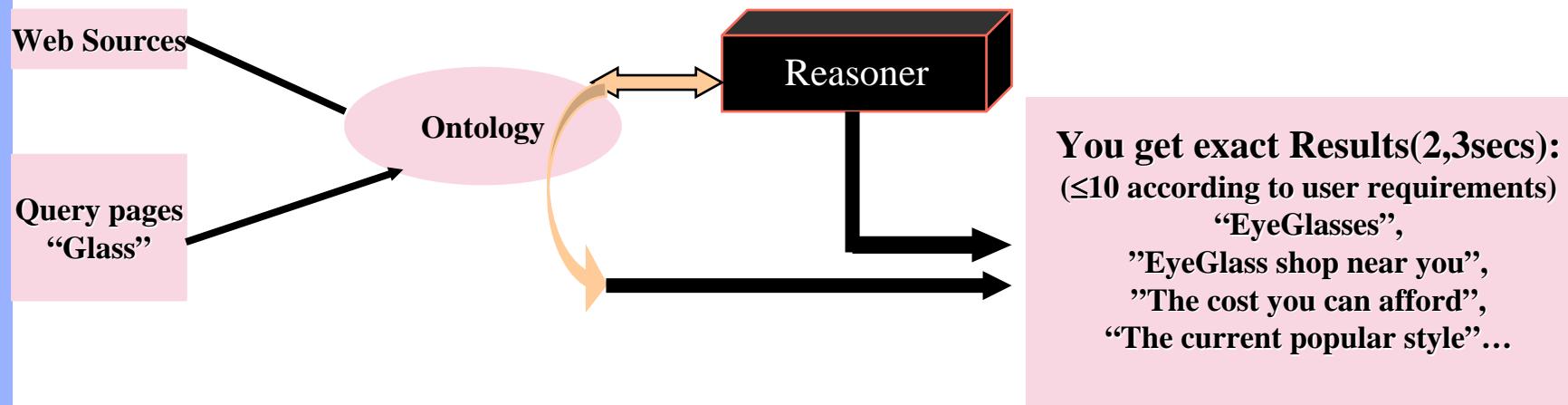
应用：WEB 检索



➤ Traditional Web Search application



➤ The Semantic Web Search Application





词网 WordNet



WordNet

- 网址：
 - ❖ <http://wordnet.princeton.edu/>
- 开发单位：
 - ❖ 1985年
 - ❖ 普林斯顿大学认知科学实验室
 - ❖ 米勒(G. A. Miller)和贝克威斯(R. Beckwith)等人，
 - ❖ 根据词义而不是根据词形来组织词汇信息。
- 免费的在线词汇数据库
- 在自然语言处理中得到广泛的应用
- 世界很多语种都开发了相应的版本
 - ❖ 各种欧洲语言： EuroNet
 - ❖ 汉语： CCD (Chinese Concept Dictionary)



WordNet规模

➤ 规模

- ❖ 名词: 57,000 words, 48,800 synsets
- ❖ 形容词: 19,500 words, 10,000 synsets
- ❖ 动词: 21,000 words, 8,400 synsets
- ❖ 还在不断发展之中

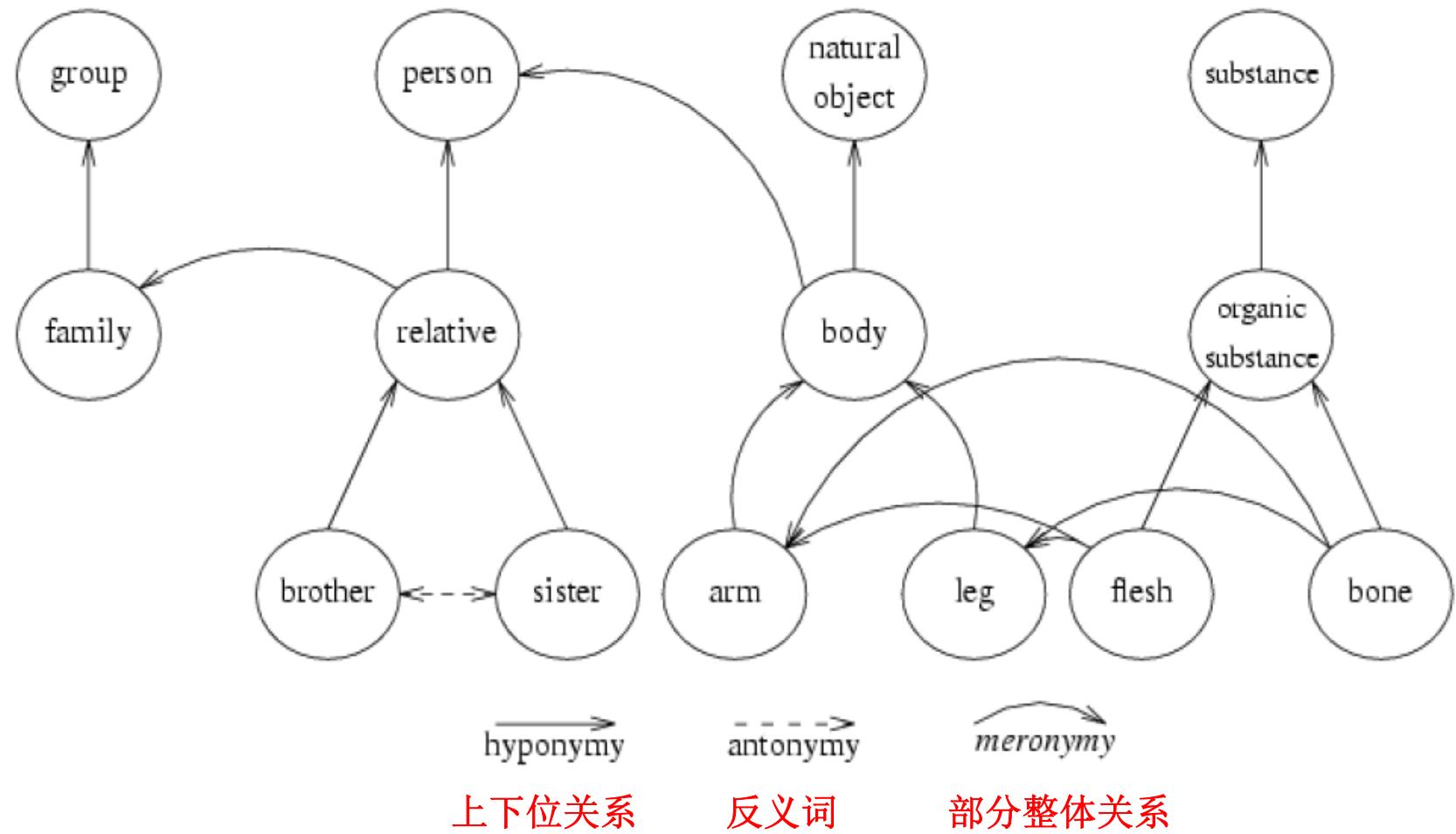


名词

- 名词约有57,000个词形（48,800个词义）。
- 它们被组织成同义词词集：
 - ❖ 用一组同义词的集合Synset来表示一个概念
 - ❖ 每一个概念有一段描述性的说明
- 每一个词汇概念结点又都具有若干个指针，分别指向它的表示上位、下位、反义、部分等关系的词：
 - ❖ 上下位关系 (hyponymy, troponymy)
 - ❖ 同义反义关系 (synonymy, antonymy)
 - ❖ 部分整体关系 (entailment, meronymy)
 - ❖
- 构成一个完整的词汇语义网络体系。



名词概念的组织



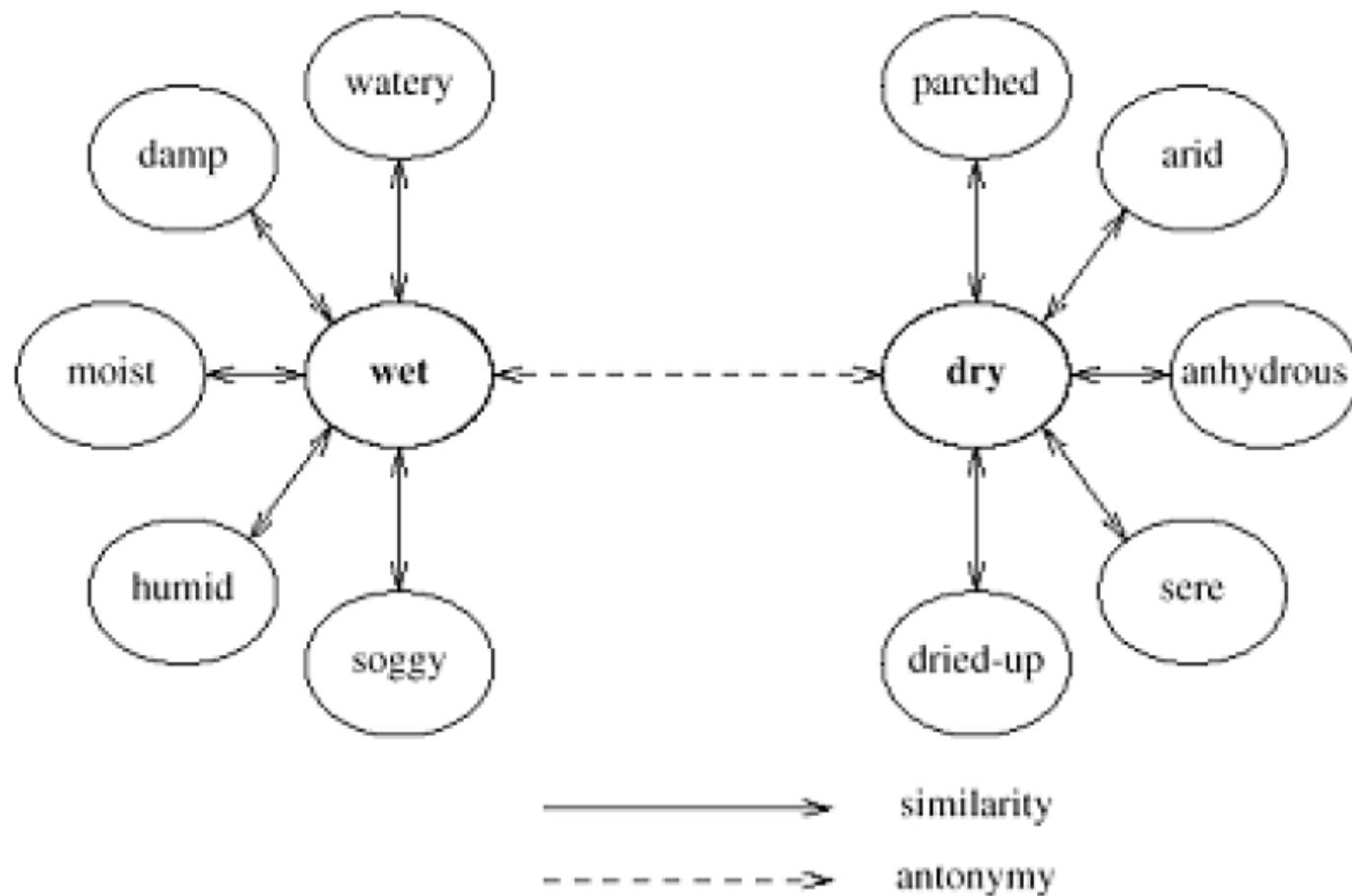


形容词

- Wordnet 中有 19,500 个形容词词形
(包含大约 10,000 个词义)
- 形容词可分为两种：
 - ❖ 描述性形容词
 - ❖ 关系性形容词
- Wordnet 中的形容词不是用分层次的树来组织的，而是按 N 维超空间的方式 来组织的



形容词概念的组织





动词

- 词网中的动词约有21,000个（包含约8,400个词义）。
- 动词的义项比名词丰富，在COLLINS词典中，动词平均有2.11个义项，而名词只有1.74个义项。往往越是常用的动词义项越是丰富。
- 动词的语义关系用“推演”来表示。

逻辑学中的推演是这样定义的：

命题P推演出命题Q当且仅当不存在使得P为真而Q为假的可能性。



动词的推演关系

- 把推演的定义引入语言研究中，当句子“某人V1”合乎逻辑地推演出句子“某人V2”时，两个动词V1和V2之间的关系，可以叫做词汇上的“推演”，即动词V1推演出动词V2。
 - ❖ 例如，从句子“约翰打酣”合乎逻辑地推演出句子“约翰睡觉”，那么，就说在词汇上动词“打酣”推演出动词“睡觉”。也就是说，如果动词“打酣”为真，则动词“睡觉”也一定为真，不存在动词“打酣”为真而动词“睡觉”为假的情形。
 - ❖ 反之不然，如果动词“睡觉”为真，动词“打酣”不一定也为真，因为存在着“睡觉而不打酣”的情况，因此，我们只能说动词“打酣”推演出动词“睡觉”，不能说动词“睡觉”推演出动词“打酣”。
 - ❖ 由动词“跛行”可以推演出动词“走路”。



WordNet小结

- 名词：词汇语义网络体系
 - ❖ 上位、下位、反义、部分
- 形容词：N维超空间
- 动词：推演关系



知网 HowNet



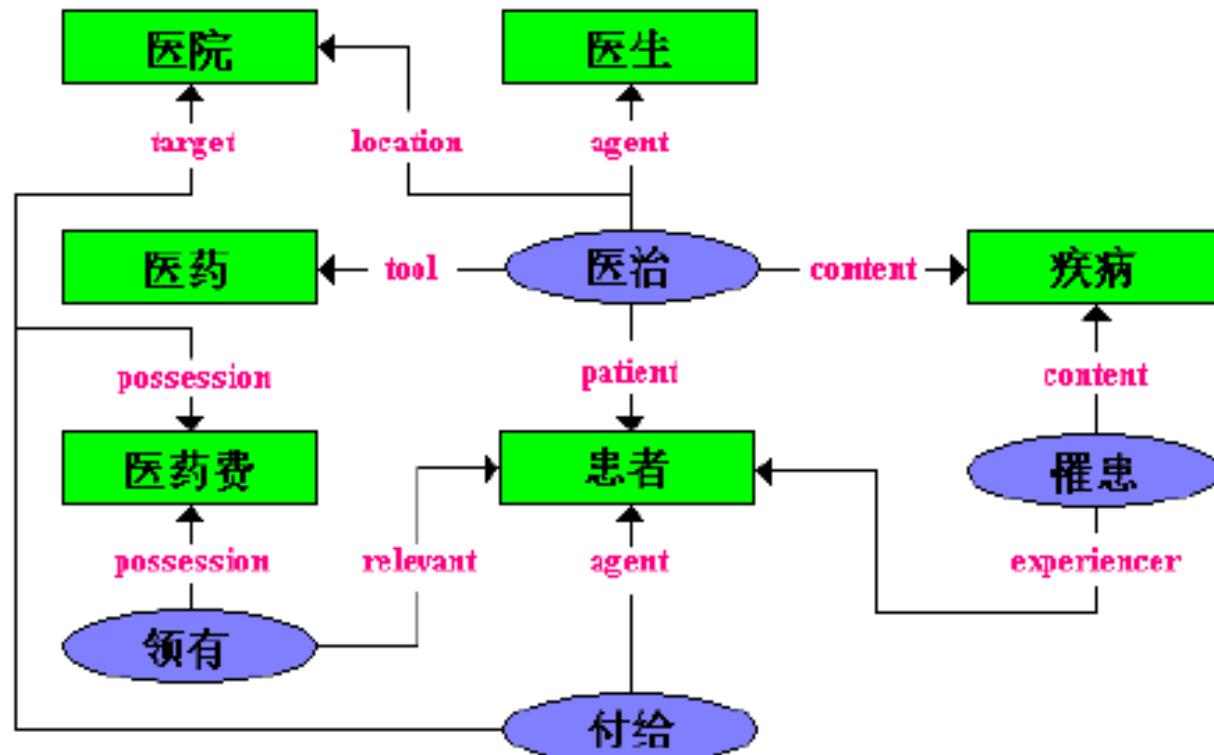
知网

- **www.keenage.com**
- 1999年初，中国中文信息学会常务理事董振东先生在因特网上公布了自己的研究成果----知网（How-Net）。
- 知网是一个以英汉双语所代表的概念以及概念的特征为基础的，以揭示概念与概念之间以及概念所具有的特性之间的关系为基本内容的常识知识库。



知网的特色

- 知网作为一个知识系统
 - ❖ 反映概念的共性和个性
 - ❖ 反映概念之间和概念的属性之间的各种关系





知网描述的各种关系

- (a) 上下位关系（由概念的主要特征体现）
- (b) 同义关系（可通过《同义、反义以及对义组的形成》获得）
- (c) 反义关系（可通过《同义、反义以及对义组的形成》获得）
- (d) 对义关系（可通过《同义、反义以及对义组的形成》获得）
- (e) 部件-整体关系（如“心”，“CPU”等）
- (f) 属性-宿主关系（如“颜色”，“速度”等）
- (g) 材料-成品关系（如“布”，“面粉”等）



知网描述的各种关系

- (h) 施事/经验者/关系主体-事件关系（如“医生”，“雇主”等）
- (i) 受事/内容/领属物等-事件关系（如“患者”，“雇员”等）
- (j) 工具-事件关系（如“手表”，“计算机”等）
- (k) 场所-事件关系（如“银行”，“医院”等）
- (l) 时间-事件关系（如“假日”，“孕期”等）
- (m) 值-属性关系（如“蓝”，“慢”等）
- (n) 实体-值关系（如“矮子”，“傻瓜”等）
- (o) 事件-角色关系（如“购物”，“盗墓”等）
- (p) 相关关系（如“谷物”，“煤田”等）



上位概念

“火车”

上位概念: entity|实体

==> thing|万物

==> physical|物质

==> inanimate|无生物

==> artifact|人工物

==> implement|器具

==> vehicle|交通工具

==> LandVehicle|车



下位词语

“火车”

下位词语：货车 军列 邮车

客车 旅客列车 特等豪华铁路客车

通勤列车 与船期衔接的旅客列车

快车 流线型火车 特别快车 直达车

特快 直达快车 直快 子弹头 子弹头列车

慢车

晚班车 晚车 夜车

早班车 早车



上下位关系

- 义原的上下位关系构成树结构

- entity|实体
 - └ thing|万物
 - ... └ physical|物质
 - ... └ animate|生物
 - ... └ AnimalHuman|动物
 - ... └ human|人
 - └ humanized|拟人
 - └ animal|兽
 - └ beast|走兽

...



部件

“火车”

部件： 餐车

车厢 豪华车厢 客车车厢 旅客车厢

寝车 卧车

火车头 蒸汽机车 机车



知网的建设

- 义原是最基本的、不易再分割的意义的最小单位
- 设想
 - ❖ 所有的概念都可以分解成各种各样的义原；
 - ❖ 应该有一个有限的义原集合，其中的义原组合成一个无限的概念集合。
- 利用中文来寻求这个有限的集合
 - 例： 治： 医治 管理 处罚
 - 处： 处在 处罚 处理
 - 理： 处理 整理 理睬
- 自下而上的归纳的方法：
 - ❖ 通过对全部基本义原进行观察分析并形成义原的标注集
 - ❖ 然后用更多的概念对标注集进行考核，据此建立完善的标注集。



知网的义原、语义特征、语义关系

义原 Sememes	2099
Entity	151
<u>thing</u> (physical, mental, fact)	
<u>component</u> (part, fitting)	
<u>time</u>	
<u>space</u> (direction, location)	
Event (relation, state; action)	812
Attribute	247
AttributeValue	889
次要特征 Secondary features	113
语义角色 Semantic roles	90



知网中的概念

-
- 物事: 57488
 - 部件: 6904
 - 时间: 2209
 - 空间: 1062
 - 属性: 3749
 - 属性值: 9044
 - 事件: 12251



知网--概念的词语表达

{famous|著名}:

大名鼎鼎	鼎鼎大名	妇孺皆知	驰名	驰誉	
赫赫有名	家喻户晓	尽人皆知	红	当红	老牌
流芳百世	流芳千古	名垂千古	名垂青史		名
名声大振	名特优	名特优新	名噪一时		名牌
名震一时	如雷贯耳	声名大噪	声名大振		头面
闻达	闻名	显赫	小有名气	已知	永垂不朽
有名气	知名		炙手可热		有名



简单概念的描述方法

- 直接标注该概念的意义。
 - ❖ 通常情况下的**简单概念**是指一个明确的事件，实体，属性或属性值，在概念中不包含任何的其它成分。
- 例如：
 - ❖ 遗失 lose {lose|失去}
 - ❖ 假装 pretend {pretend|假装}
 - ❖ 料理 manage {handle|处理}
 - ❖ 资助 subsidize {grant|赐}
 - ❖ 滑翔 glide {fly|飞}
 - ❖ 教导 teach {teach|教}



复杂概念的描述方法

- 利用动态角色与特征来标注复杂概念。
所谓的复杂概念是以事件为中心，除了事件中心本身以外还有一个或一个以上的动态角色
- 例如：
 - ❖ 严 禁 包含动态角色——方式 (manner)
 - ❖ 贷 款 包含动态角色——所有物 (possession)
 - ❖ 呼 救 包含动态角色——目的 (purpose)
 - ❖ 复 原 包含动态角色——原状态 (StateIni)



四类概念的描述

➤ 事件类概念

救灾: DEF=rescue|救助, StateIni= unfortunate |不幸

➤ 属性值和数量值类概念

美味: DEF=aValue|属性值, taste|味道, good|好

➤ 属性和数量类概念

味道: DEF=attribute|属性, taste|味道, &edible|食物

➤ 单位类概念

公里: DEF=unit|单位, &length|长度

➤



知网系统

- 中英双语知识词典
- 知网管理工具
- 知网说明文件
 - ❖ 动态角色与属性
 - ❖ 词类表
 - ❖ 同义、反义以及对义组的形成
 - ❖ 事件关系和角色转换
 - ❖ 标识符号及其说明



知识词典

-
- NO.=000001
 - W_C=打
 - G_C=V //词性
 - E_C=~酱油, ~张票, ~饭, 去~瓶酒, 醋
~来了 //词语例子
 - W_E=buy
 - G_E=V
 - E_E=
 - DEF=buy|买 //概念定义



知识词典

- NO.=015492
- W_C=打
- G_C=V
- E_C=~毛衣, ~毛裤, ~双毛袜子, ~草鞋, ~一条围巾, ~麻绳, ~条辫子
- W_E=knit
- G_E=V
- E_E=
- DEF=weave|辫编



知网的特色

- 是计算机化的，不是纸面的
- 是知识系统，不是语义词典
- 意义关系是网状的，不是树状的
- 义原作为特性标注概念，不是意义分类
- 面向开发者的，不是面向最终用户的
- 词语的选择考虑是真实语料



小结

-
- 本体（Ontology）的概念
 - Semantic Web 与 OWL
 - 词网 WordNet
 - 知网 HowNet



Any Question?