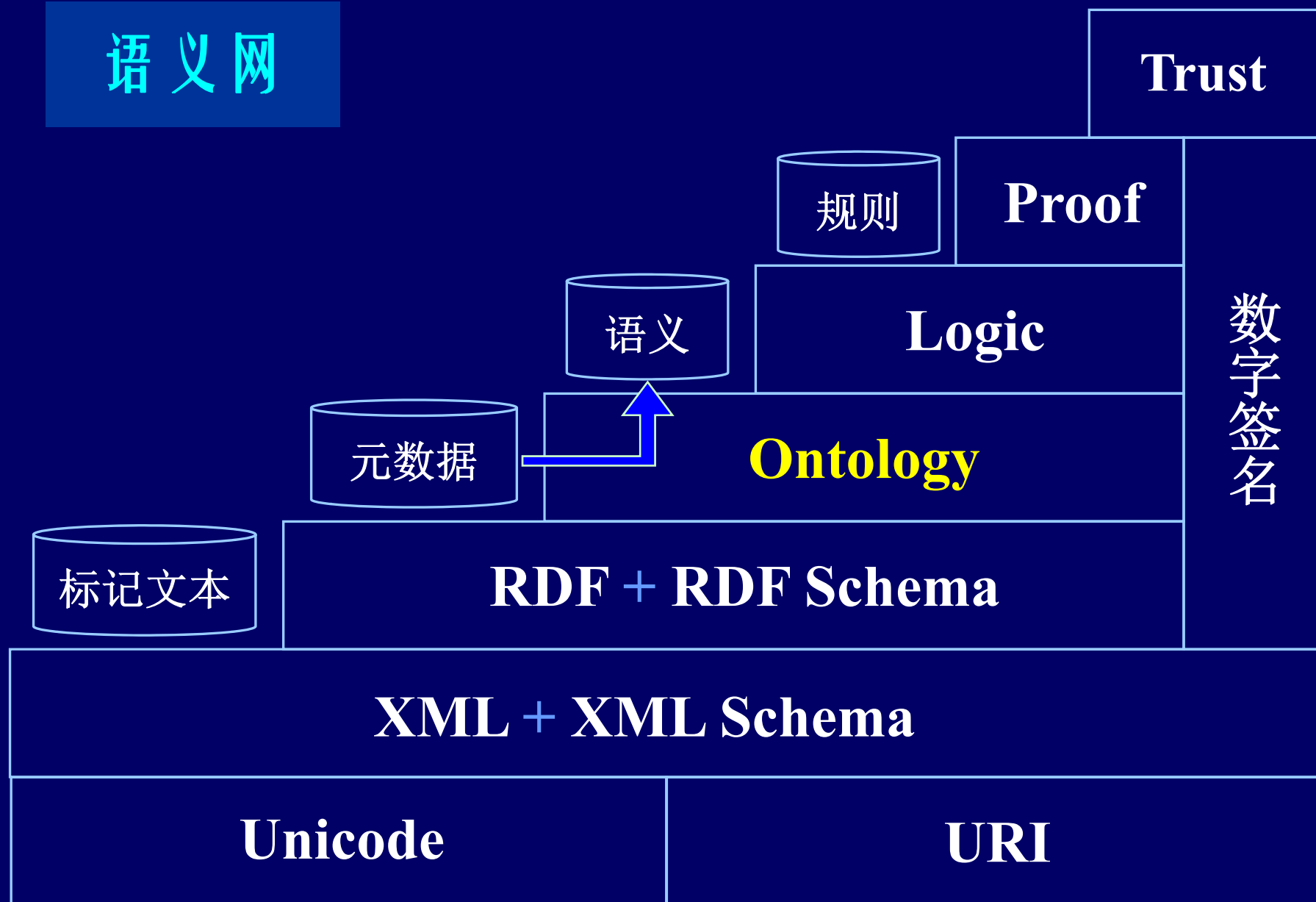


硕士研究生课程 《智能信息处理》

语义网基础理论

大连海事大学信息科学技术学院

语义网



XML + XML Schema

结构

<属性_n>属性值_n</属性_n>
.....
<属性₂>属性值₂</属性₂>
<属性₁>属性值₁</属性₁>

RDF + RDF Schema

资源 ← 属性 ← 属性值

信息资源

<元数据_n>值_n</元数据_n>

.....

<元数据₂>值₂</元数据₂>

<元数据₁>值₁</元数据₁>



XML

资源←元数据← 值



RDF

信息资源

<语义_n>内容_n</语义_n>

.....

<语义₂>内容₂</语义₂>

<语义₁>内容₁</语义₁>



XML

资源←语义←内容



RDF

信息资源

<语义_{A_n}>内容_{A_n}</语义_{A_n}>

.....

<语义_{A₂}>内容_{A₂}</语义_{A₂}>

<语义_{A₁}>内容_{A₁}</语义_{A₁}>



信息资源A

<语义_{B_m}>内容_{B_m}</语义_{B_m}>

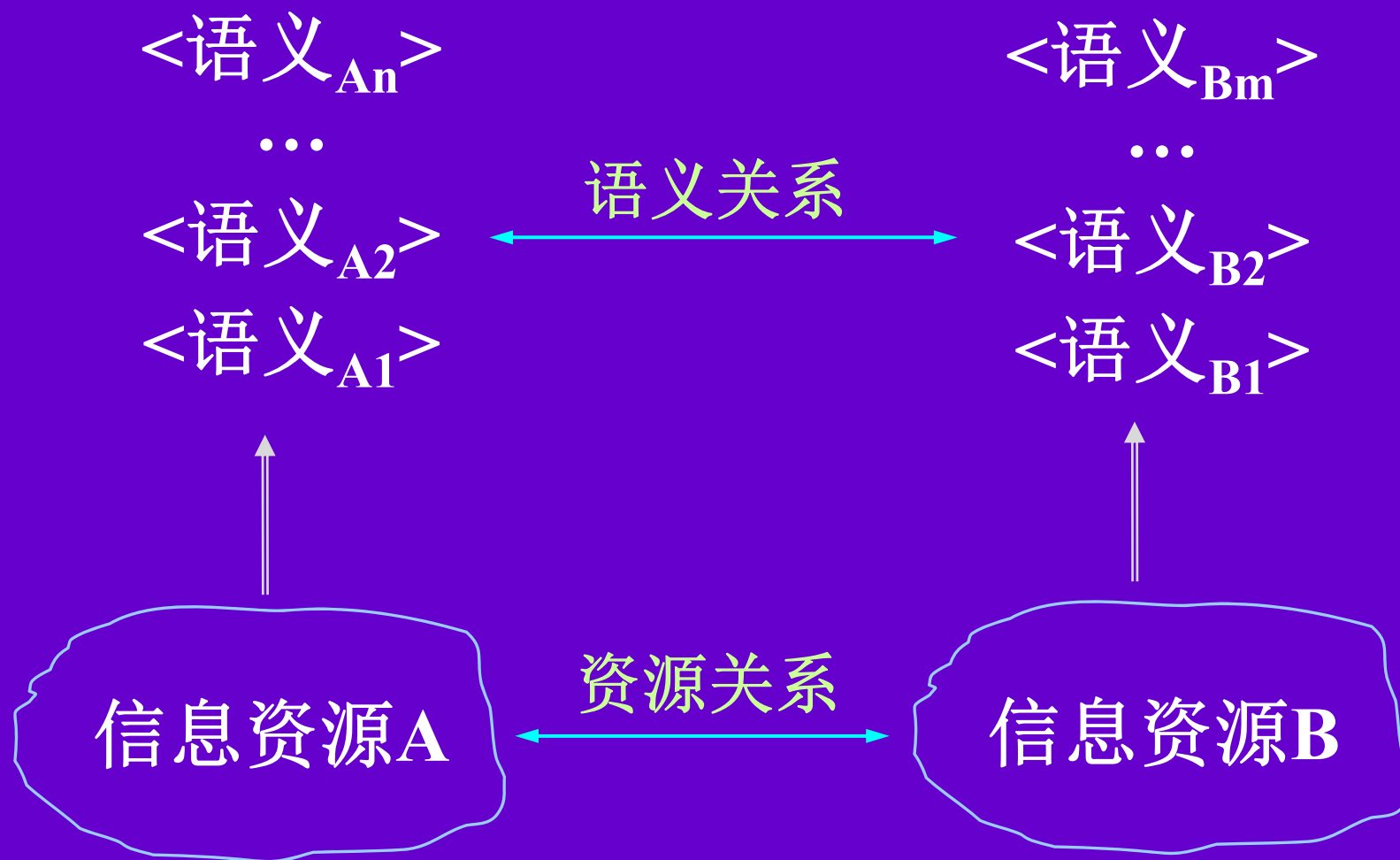
.....

<语义_{B₂}>内容_{B₂}</语义_{B₂}>

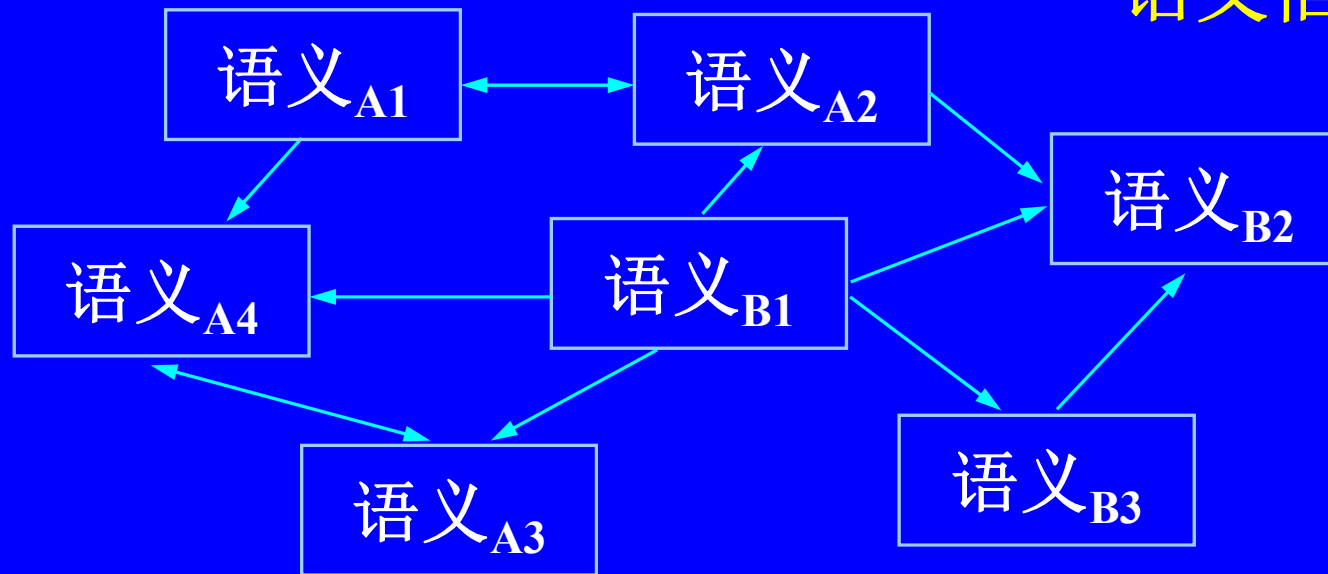
<语义_{B₁}>内容_{B₁}</语义_{B₁}>



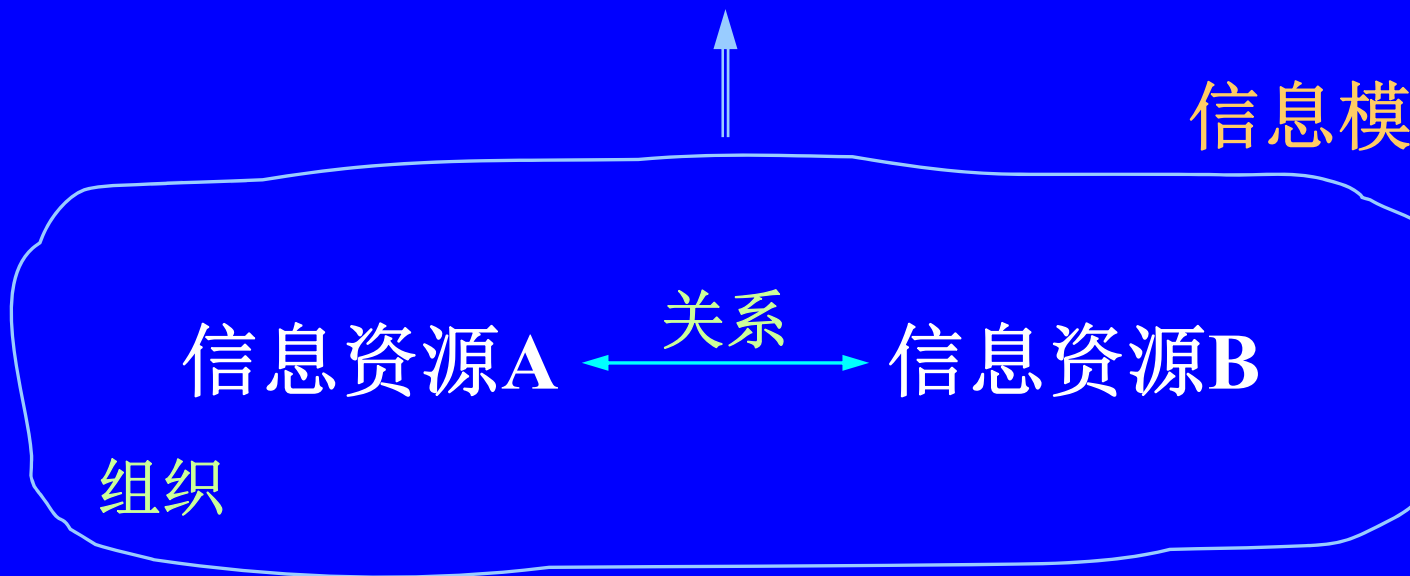
信息资源B



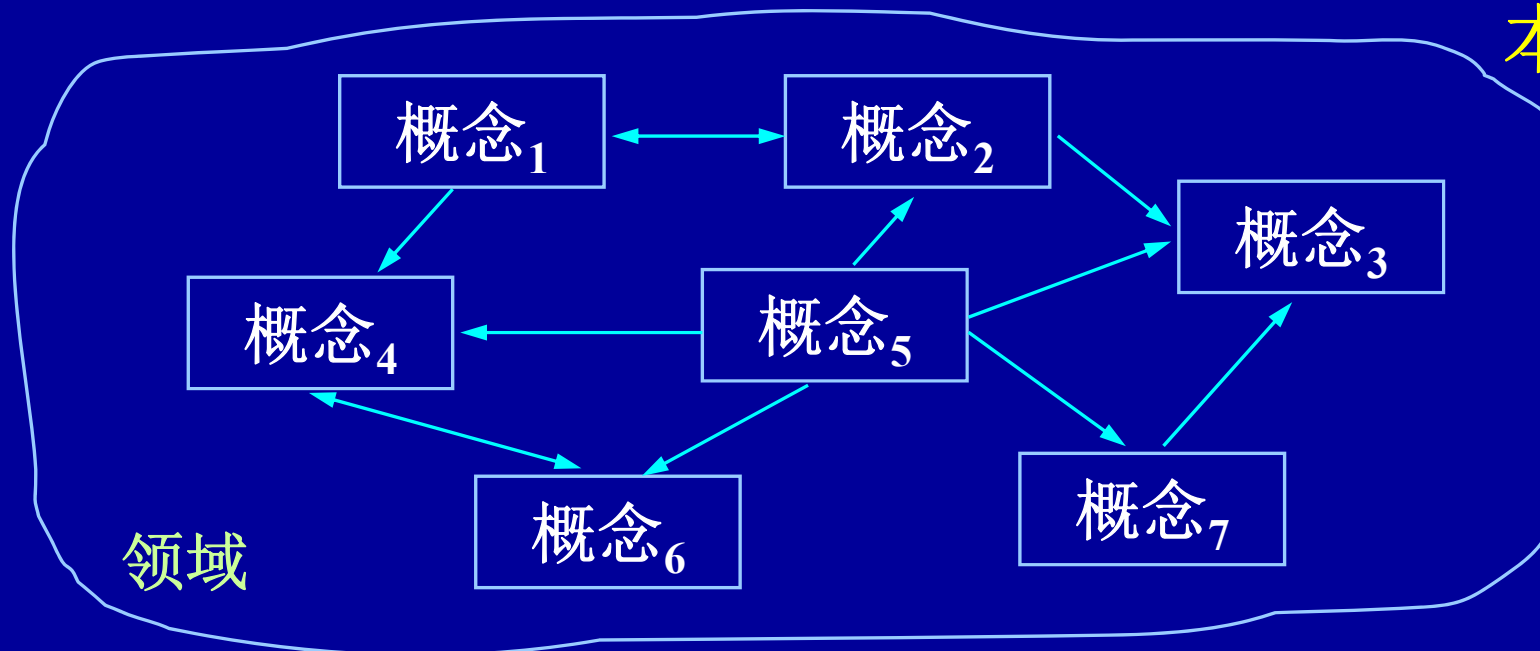
语义信息模型



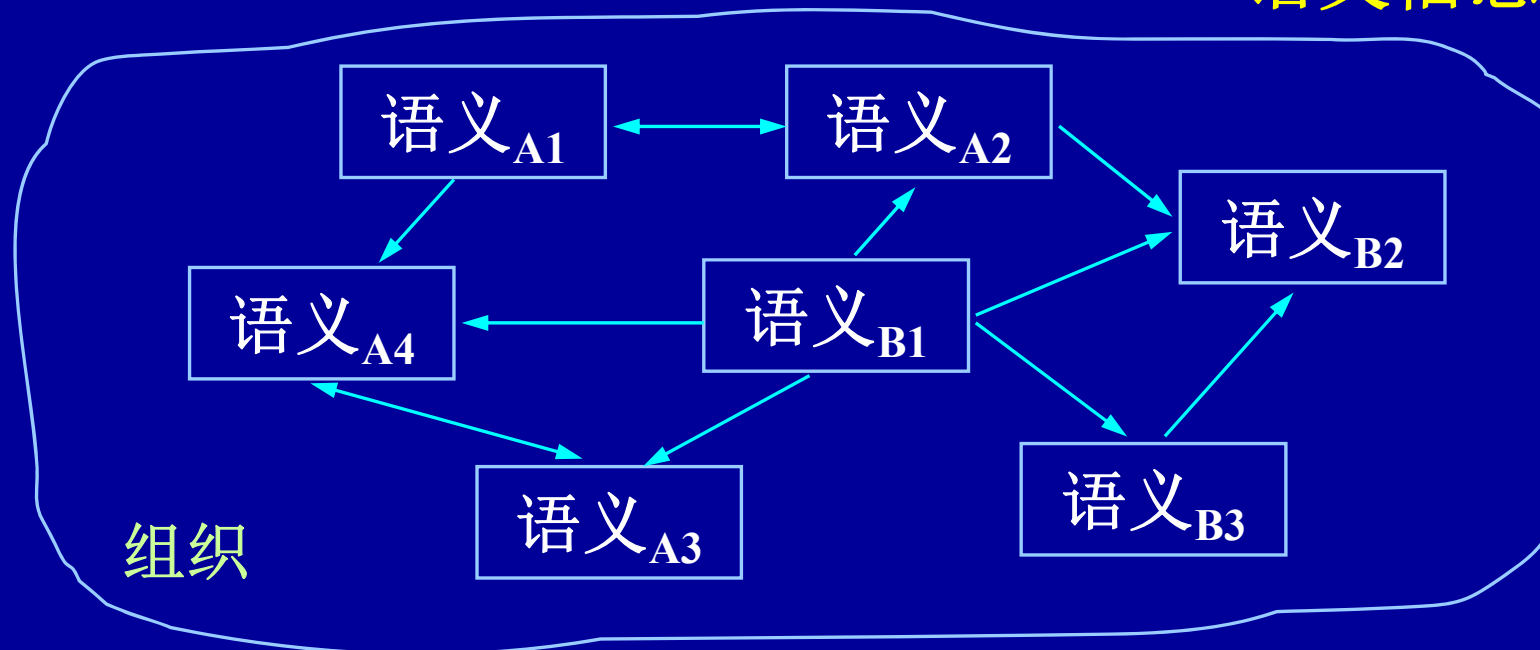
信息模型



本体



语义信息模型



第5章 本体(Ontology)

定义

评估标准

构成

分类

构造准则

典型实例

构造方法

应用



本体的定义

哲学领域

人工智能领域

计算机领域

西方哲学 本体论

关于存在及其本质和规律的学说。

中国哲学 本根论

关于万物产生、存在、发展、变化的根本原因的学说。

ontology = on + logic

ontology是关于**on**的学说

拉丁语

ontology是关于**being**的学说

英语

存在(本体) 有 是

汉语

本体论	关于“本体”的学说
存在论	关于“存在”的学说
万有论	关于“有”的学说
是论	关于“是”的学说

中国哲学界目前对**Ontology**的认识

本体论是关于**being**的学说。

存在派
是派

being ≈ 存在
being ≈ 是



树梢 木上曰末 末 用 使人或物发挥功能

树根 木下曰本 本 体 人的全身

底层的、起支撑作用的东西

人的生存与发展

被支撑的
do use
作用



本体
root body
支撑的

行为



知识

本体(Ontology) = 与任务独立的知识库

本体的定义(Neches)

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。



领域词汇及其相互关系

↑
基本术语
派生术语

↑
基本关系
派生关系

← 基本术语
基本关系
推导规则

本体的定义 (Gruber, 1993)

An ontology is an explicit specification of a conceptualization.

Gruber定义

本体是概念模型的明确的规范说明。

Borst定义

本体是共享概念模型的形式化规范说明。

Gruber定义

本体是概念模型的**明确的**规范说明。

Borst定义

本体是**共享**概念模型的**形式化**规范说明。

Studer定义

本体是**共享**概念模型的**明确的****形式化**规范说明。

概念模型(**conceptualization**)

明确的(**explicit**)

形式化(**formal**)

共享(**share**)

领域的主要概念

概念有明确定义

计算机可读

领域内共识

概念 Concept (辞海, 1989)

- 是反映对象的特有属性的思维形式
- 从对象的属性中抽出特有属性概括而成
- 表达的语言形式是词或词组
- 都有内涵和外延



对Gruber定义的理解

语词及其相互关系

本体是反映客观存在的概念模型，
是对概念模型的明确描述。

本体的目标(P111)

获取、描述和表示相关领域的知识，
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，
并从不同层次的形式化模式上给出这些
语词和语词之间相互关系的明确定义。

我们的理解

明确的、形式化的、可共享的

语词及其相互关系

规范说明为

本体论是反映客观存在的概念模型。

Ontology

本体论

On/Being

本体

Logic

论

本体论 = 概念模型 = 语词-关系图 = 词汇表
Ontology = 本体模型 = 术语-关系图 = 术语表

本体 = 概念 + 属性 + 公理 + 取值 + 名义
ontology concept property axiom value nominal

本体 = 概念 + 关系 + 函数 + 公理 + 实例
ontology concept relation function axiom instance

本体论 = 术语 + 术语之间关系

概念
名义
实例
属性
属性值

关系
函数
公理

我们的理解

明确的、形式化的、可共享的

语词及其相互关系

规范说明为

本体论是反映客观存在的概念模型。

Ontology

本体论

On/Being

本体

Logic

论

本体论 = 概念模型 = 语词-关系图 = 词汇表

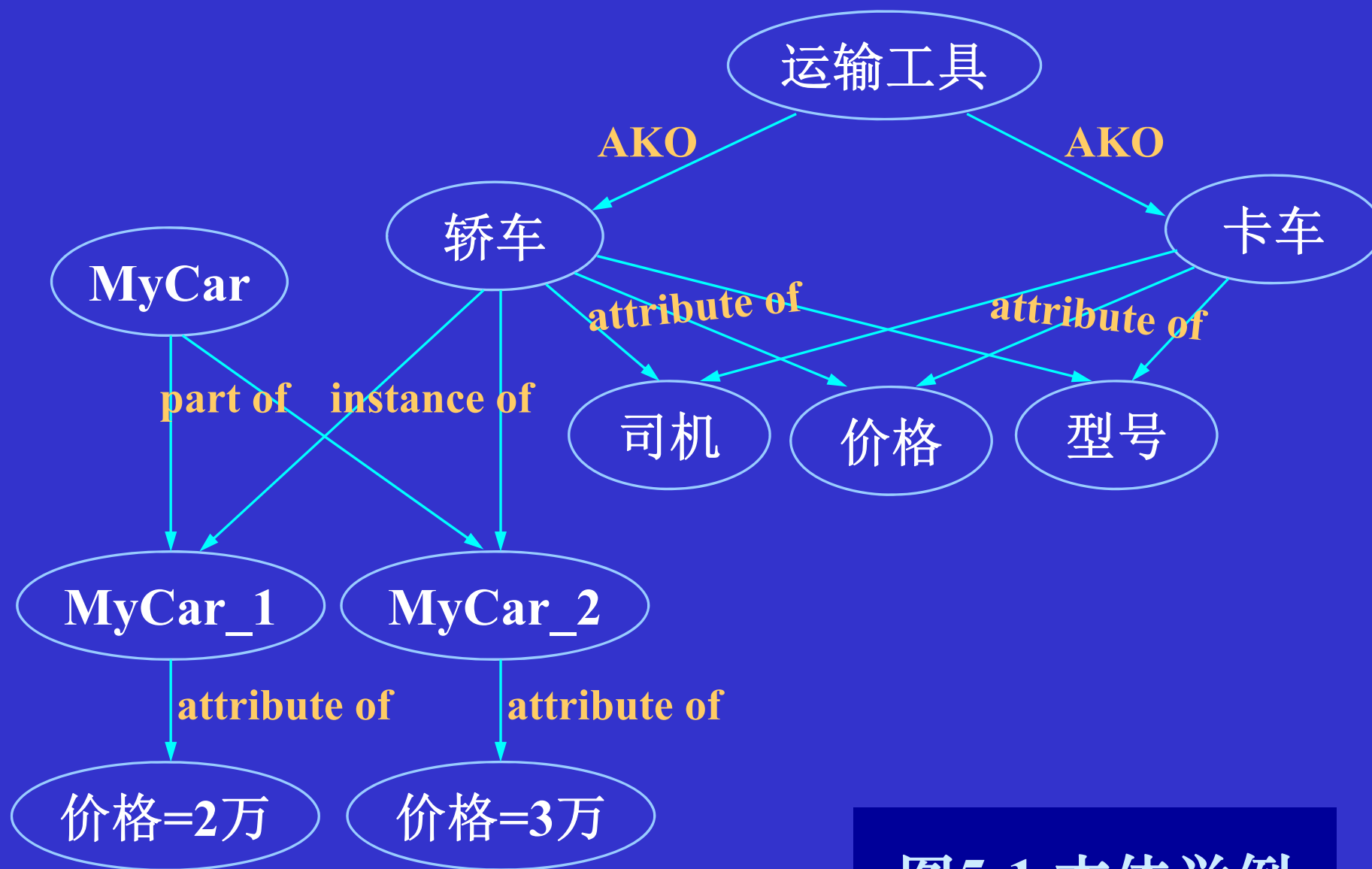


图5-1 本体举例

构造本体的准则 (Gruber, 1995)

1. 清晰性 (Clarity)
2. 完全性 (Completeness)
3. 一致性 (Coherence)
4. 可扩展性 (Extendibility)
5. 最小承诺 (Minimal ontological commitment)
最小编码偏好 (Minimal encoding bias)



最小承诺准则

只定义最必要的术语

只定义约束最弱的关系

最小编码偏好准则

不指定术语形式化用何编码

Gruber准则

1. 清晰性
2. 完全性
3. 一致性
4. 可扩展性
5. 最小承诺/最小编码偏好

Arpirez规则

- 概念命名标准化
- 概念层次多样化
- 语义距离最小化

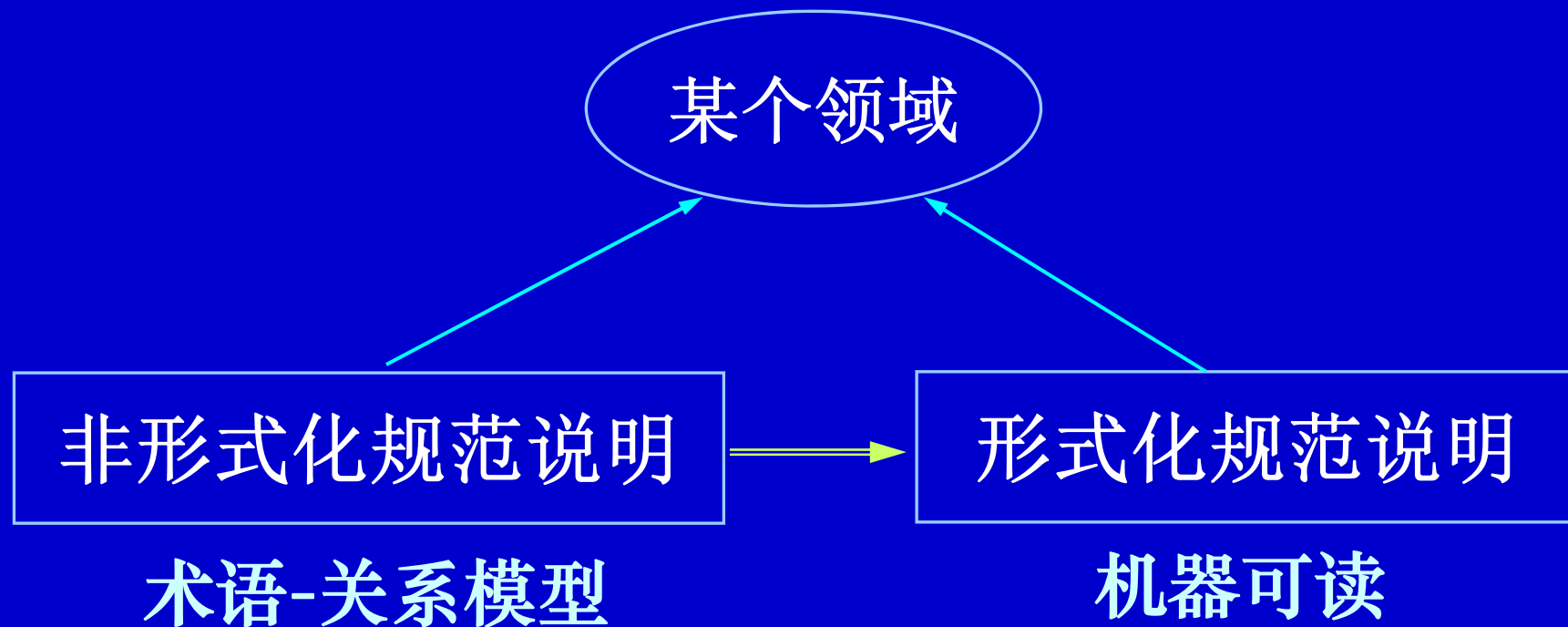
五种典型的构造本体的方法

- Uschold和King方法 ■
- Gruninger和Fox方法 ■
- Berneras方法 ■
- MethOntology方法 ■
- 基于SENSUS方法 ■

Uschold和King方法

- (1) 明确目标
- (2) 构造本体
- (3) 评价本体
- (4) 完善文档

Gruninger和Fox方法构造本体



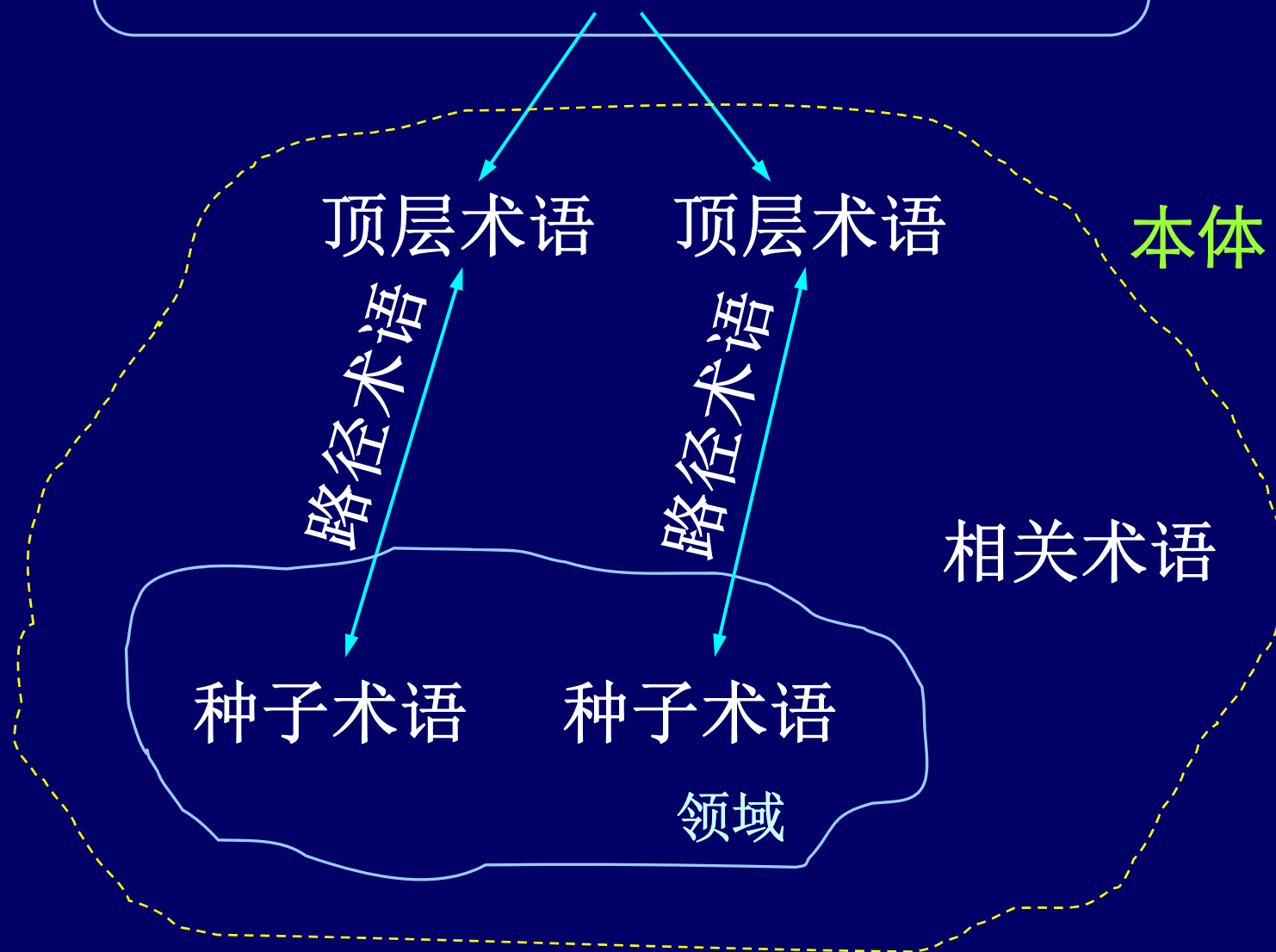
Berneras方法：面向知识共享与重用

- (1) 目标与需求说明
- (2) 基于高层本体的初步设计
- (3) 本体的提炼和构造

MethOntology方法构造本体

- (1) 明确构造本体时采用的行为
- (2) 规划构造本体的生命周期
- (3) 实施每种行为/每个阶段

自然语言处理本体SENSUS (70,000多个概念)



本体的评估标准(Mariano)

- 对知识工程的继承
- 详细程度
- 知识形式化工具
- 与具体应用的关联程度
- 自底向上或自顶向下
- 规划生命周期 开发软件寿命周期的处理
- 与IEEE-1074的差异
- 采用的技术
- 所构造的本体和所应用的系统
- 协作和分布性能

本体研究主题

拥有知识

通用知识
专用知识

通用本体
领域本体——{ 语言学本体
.....

解决问题

任务本体——{ 知识表示本体
.....

根据形式化程度的本体分类

非形式化 自然语言

半形式化 结构化的自然语言
程度低的形式语言

形式化 形式语言

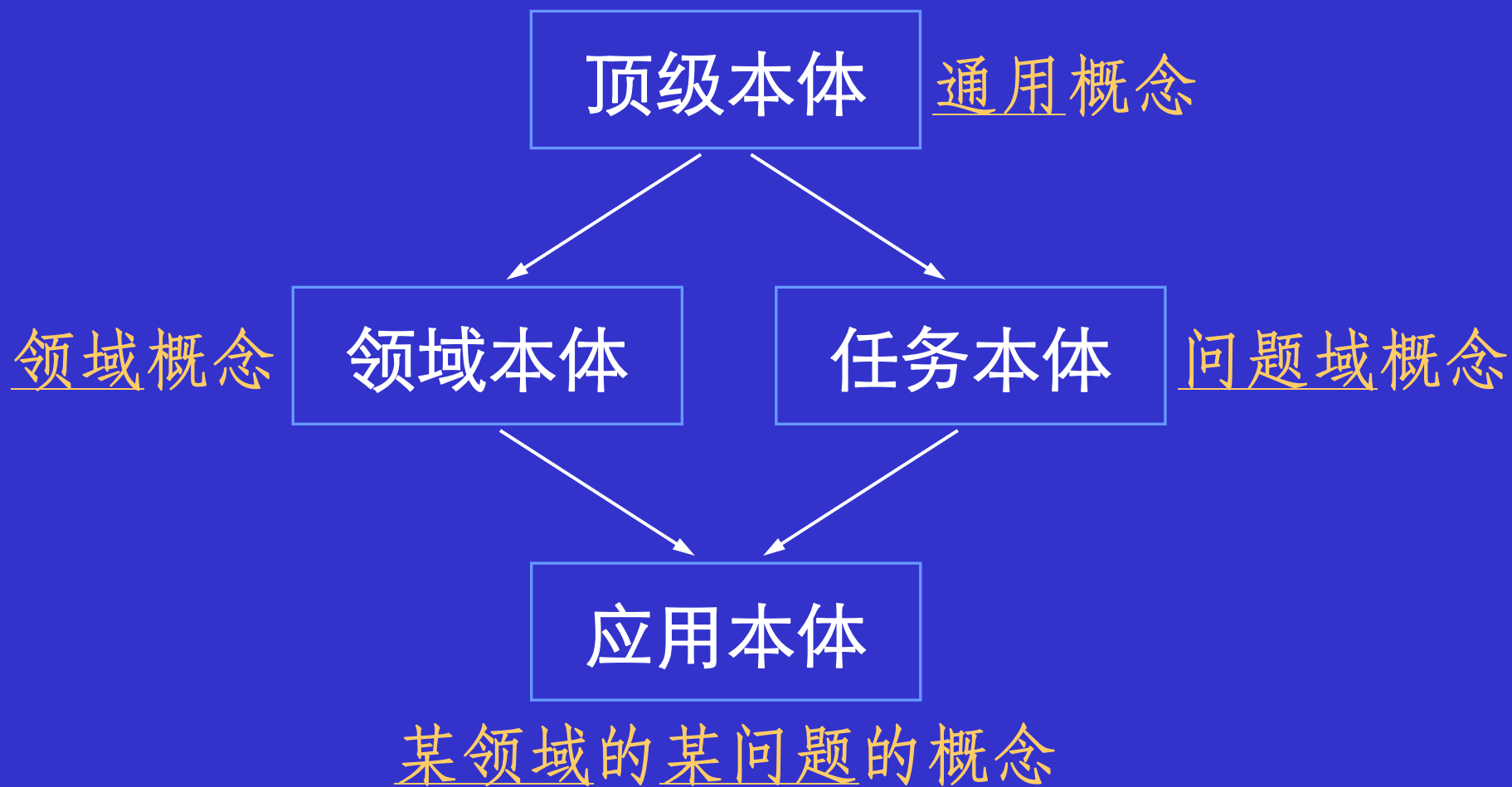


图5-2 本体的层次分类

典型本体

词典

主题映射

分类法

WordNet

词典中的语义关系



ASIS词表
(American Society for Information Science)

ATT词表
(Art & Architecture Thesaurus)

中国分类主题词表
(第一版, 1994; 第二版, 2005)

汉语分类主题词表片断

▲ 半导体存储器

TP333.5

D 大规模集成电路存储器

Z 存储器

C 局部存储器

C 随机存取存储器

C 只读存储器

C 主存储器

D 代替词 同义

Z 族首词 上位

C 参见词 相关

汉语分类主题词表的片断(P124)

▲ 程序设计

[O244]; TP311

- 层次设计法
- 程序分析
- 程序优化
- 程序正确性证明
- 程序转换
- 程序综合
- 宏汇编程序设计
- 解释性程序设计
- 模块化程序设计
- 冗余程序设计
- 顺序程序设计
- 微程序设计
- 系统化程序设计
- 结构化程序设计
- 逐步求精法
- 自动程序设计
- C 软件
- C 软件设计

主题映射的定义 (ISO/IEC 13250)

主题映射是一套用来组织信息的方法，使用这个方法可以提供最佳的信息资源导航。

相当于目录、术语表、主题索引
三个部分：主题、关联、参考位置

信息资源管理的分类语言

分类法  体系分类法 — 混合分类法
组配分类法

杜威十进制分类法(DDC)

国际十进制分类法(UDC)

国会图书馆分类法(LCC)

冒号分类法(CC)

中国图书馆分类法(中图法)

WordNet

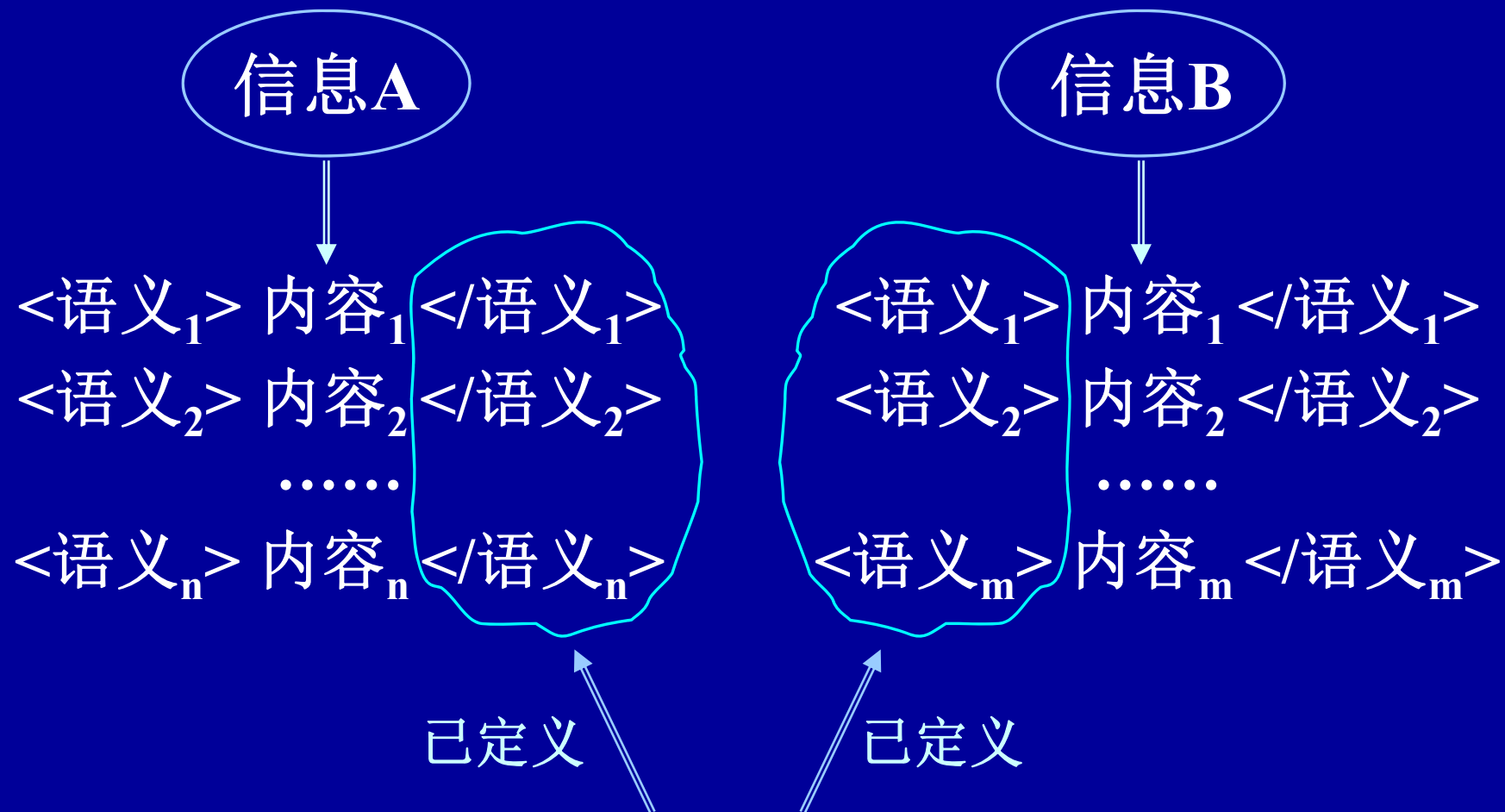
以同义词集合为基本建构单元的词库
同义词集合之间以关系类型相关联



本体的应用

在语义网中的应用

在信息检索中的应用



本体：概念及其相互关系

共享的 明确的 形式化的

基于本体的信息检索的原理(P130)

- (1) 建立相关的领域本体;
- (2) 参照本体将信息按语义格式存储;
- (3) 查询请求转换为语义格式, 检索;
- (4) 检索结果进行语义格式反处理。

课间休息

再会

