

人工智能

- 知识表达与自动推理

马 超
武 汉 大 学

`chaoma@whu.edu.cn`

提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

自动推理证明发展史：引言

- 出色的老式人工智能（Good Old Fashioned AI, GOFAI）——哲学家约翰.豪格兰德
- 一个用规则和事实来程序化的高速数字计算机可能表现出智力行为

——图灵

- 人类是借助事实与规则来产生智力行为的
- 经典人工智能技术主要以符号表示、符号处理为实现智能的主要手段，推理和搜索是其中的核心技术

自动推理证明发展史：示例

■ 5个房间问题

- 有5间不同颜色的房间，每间住个不同国籍的人，每人有自己喜欢的饮料、香烟和宠物。已知信息：

1. 英国人在红房间中
2. 西班牙人有一条狗
3. 挪威人住在左边第一间房里
4. 黄房间中的人在抽库尔斯牌香烟
5. 抽切斯菲尔德牌香烟的人是养了一只狐狸的人的邻居
6. 挪威人住在蓝房间隔壁
7. 抽温斯顿牌香烟的人有一只蜗牛
8. 抽幸运牌香烟的人喝橘子汁
9. 乌克兰人喝茶
10. 日本人抽国会牌香烟
11. 抽库尔斯牌烟的房间在有匹马的房间隔壁
12. 绿房间中的人喝咖啡
13. 绿房间在白房间的左边
14. 中间房间的人喝牛奶

**问题：斑马在哪个房间中？
哪个房间中的人喝水？**

自动推理证明发展史

- 1. 某房间中的人左抽库尔斯
- 2. 西班牙人右一条狗
- 3. 抽切斯菲尔德牌香烟的人的邻居养了一只狐狸
- 4. 抽切斯菲尔德牌香烟的人的邻居养了一只狐狸
- 5. 抽切斯菲尔德牌香烟的人的邻居养了一只狐狸

房间号	1	2			
颜色					
国籍	挪威人				西班牙人
香烟	库尔斯				切斯菲尔德
饮料	水			咖啡	橘子汁
宠物	狐狸	马	蜗牛	斑马	狗

机器真的能自动完成这样的推理吗？



自动推理证明发展史：机器求解

```
domains
  ID= symbol
  HOUSE = h(ID,NO)
  HLIST = reference HOUSE*
  NO = integer
  NOLIST = NO*
  CHARLIST = CHAR*
  CHARLISTS = CHARLIST*

predicates
  nondeterm solve
  nondeterm
  candidate(HLIST,HLIST,HLIST,HLIST,HLIST)
  nondeterm perm(HLIST)
  nondeterm
  constraints(HLIST,HLIST,HLIST,HLIST,HLIST)
  nondeterm
  permutation(NOLIST,NOLIST)
  nondeterm
  delete(NO,NOLIST,NOLIST)
  member(HOUSE,HLIST)
  nondeterm next(NO,NO)
  nondeterm lleft(NO,NO)

clauses
  solve():-
    constraints(Colours,Drinks,Nationalities,Cigarettes,Pets),
    candidate(Colours,Drinks,Nationalities,Cigarettes,Pets),
    member(h(water,WaterHouse),Drinks),
    member(h(WaterColour,WaterHouse),Colours),
    member(h(zebra,ZebraHouse),Pets),
    member(h(ZebraColour,ZebraHouse),Colours),
    write("They drink water in the ",WaterColour," house\n"),
    write("The zebra live in the ",ZebraColour," house\n").
  candidate(L1,L2,L3,L4,L5):-
    perm(L1),
    perm(L2),
    perm(L3),
```

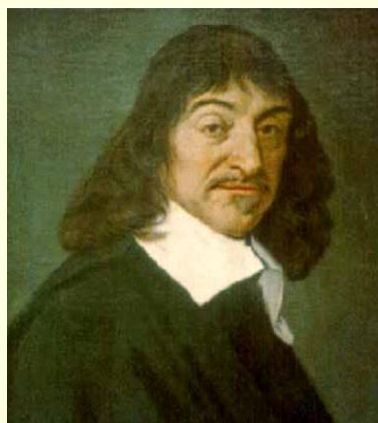
```
    perm(L4),
    perm(L5).

perm([h(_,A),h(_,B),h(_,C),h(_,D),h(_,E)]):-
  permutation([A,B,C,D,E],[1,2,3,4,5]).

constraints(Colours,Drinks,Nationalities,Cigarettes,Pets):-
  member(h(englishman,H1),Nationalities),
  member(h(red,H1),Colours),
  member(h(spaniard,H2),Nationalities),
  member(h(dog,H2),Pets),
  member(h(norwegian,1),Nationalities),
  member(h(kools,H3),Cigarettes),
  member(h(yellow,H3),Colours),
  member(h(chesterfields,H4),Cigarettes),
  next(H4,H5),
  member(h(fox,H5),Pets),
  member(h(norwegian,H6),Nationalities),
  next(H6,H7),
  member(h(blue,H7),Colours),
  member(h(winston,H8),Cigarettes),
  member(h(snails,H8),Pets),
  member(h(lucky_strike,H9),Cigarettes),
  member(h(orange_juice,H9),Drinks),
  member(h(ukrainian,H10),Nationalities),
  member(h(tea,H10),Drinks),
  member(h(japanese,H11),Nationalities),
  member(h(parliaments,H11),Cigarettes),
  member(h(kools,H12),Cigarettes),
```

```
    next(H12,H13),
    member(h(horse,H13),Pets),
    member(h(coffee,H14),Drinks),
    member(h(green,H14),Colours),
    member(h(green,H15),Colours),
    lleft(H16,H15),
    member(h(ivy,H16),Colours),
    member(h(milk,3),Drinks).
  permutation([],[]).
  permutation([A|X],Y):-
    delete(A,Y,Y1),
    permutation(X,Y1).
  delete(A,[A|X],X).
  delete(A,[B|X],[B|Y]):-
    delete(A,X,Y).
  member(A,[A|_]):-!.
  member(A,[_|X]):-
    member(A,X).
  next(X,Y):-
    lleft(X,Y).
  next(X,Y):-
    lleft(Y,X).
  lleft(1,2).
  lleft(2,3).
  lleft(3,4).
  lleft(4,5).
goal
  solve.
```

自动推理证明发展史：起源



笛卡尔



莱布尼茨

自动证明的提出

- 笛卡尔、莱布尼茨（17世纪）
 - 萌发了用机械系统实现定理证明的想法
 - 把一类数学问题当作一个整体，建立统一的证明过程，按照规定的程序步骤机械地进行下去，在有限步骤之后判断出定理的正确性

自动推理证明发展史：发展

- 由于传统的兴趣和应用的價值，初等几何问题的自动求解成为数学机械化的研究焦点。

- 希尔伯特

- 20世纪初，在他的名著《几何基础》中给出了一条可以对一类几何命题进行判定的定理。
- 希尔伯特对命题的要求太高，当时仅能解决很少的一类几何定理的机器证明，却是历史上第一个关于某类几何命题的机械化检验方法的定理。



希尔伯特

自动推理证明发展史：发展

■ 塔斯基（波兰）

- 1950年，证明了：“一切初等几何和初等代数范围的命题都可以用机械方法判定”
- 为几何定理的机器证明开拓了一条利用代数方法的途径
- 方法太复杂，即使用高速计算机也证明不了稍难的几何定理

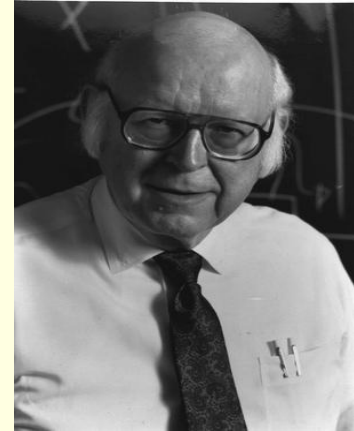


塔斯基

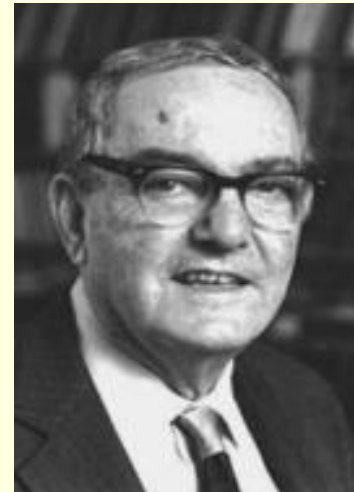
自动推理证明发展史：发展

■ 纽厄尔, 西蒙

- 1956年, 发表了论文《逻辑理论机》(LTM)
- 认为LTM不仅是计算机智力的有力证明, 也是人类认知本质的证明
- 1957年开发了最早的AI程序设计语言IPL语言
- 1960年, 成功地合作开发了“通用问题求解系统” GPS (General Problem Solver)



艾伦.纽厄尔



赫伯特.西蒙

自动推理证明发展史：发展

- 美籍华裔王浩（1921—1995）
 - 数学家、逻辑学家、计算机科学家、哲学家。
 - 简历
 - 生于中国山东省济南市
 - 1943年毕业于西南联合大学数学系
 - 1945年毕业于清华大学研究生院哲学系
 - 1948年获哈佛大学哲学博士学位
 - 1954-1956年在牛津大学任高级教职
 - 1961-1967年任哈佛大学教授
 - 1967-1991年任洛克菲勒大学逻辑学教授
 - 20世纪50年代初当选美国科学院院士及不列颠科学院外籍院士。



王浩

自动推理证明发展史：发展

- 美籍华裔王浩（1921—1995）
 - 1953年起开始计算机理论与机器证明的研究。他敏锐地感觉到被认为过分讲究形式的精确又十分繁琐而无任何实际用处的数理逻辑，可以在计算机领域发挥极好的作用。
 - 1959年，采用“王浩算法”用计算机证明了罗素、怀海德的巨著《数学原理》中的几百条有关命题逻辑的定理，仅用了9分钟（vs 10年），宣告了用计算机进行定理证明的可能性，第一次明确提出“**走向数学的机械化**”。



王浩

自动推理证明发展史：发展

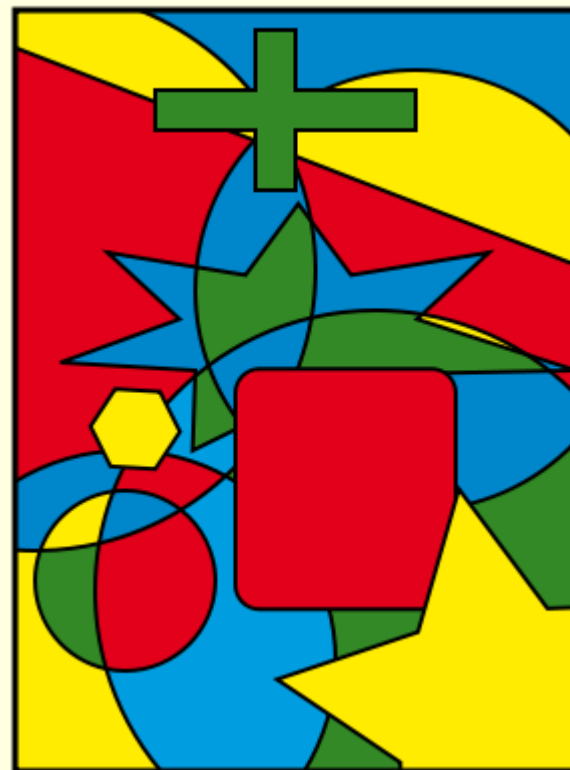
■ 美籍华裔王浩



- 1983 年，获国际人工智能联合会“数学定理机械证明里程碑奖”，表彰他在数学定理机械证明研究领域的开创性贡献。
- 1972年以后，王浩数次回国讲学。1985年兼任北京大学教授；1986年兼任清华大学教授。
- 新中国成立之初，公开发表演说表示对新中国的支持。1972年回国时曾受周恩来总理的接见。1973年撰写了《访问中国的沉思》，赞美新中国，被报纸与杂志广泛刊载。为此，他受到了许多攻击。他热爱祖国和中华民族的精神值得人们学习与称道。

自动推理证明发展史：发展

- 1977年，美国年轻的数学家阿佩尔等在高速电子计算机上耗费 1200 小时的计算时间，证明了著名的“四色定理”，人类百年悬而未决的疑问最终被圆满解决了。这一成就轰动一时，成为机器定理证明的一个典范。



自动推理证明发展史：发展

■ 著名数学家吴文俊



- 中国科学院数学与系统科学研究院研究员、中国科学院院士、第三世界科学院院士
- 1919年出生于上海
- 1940年毕业于交通大学数学系
- 1949年获法国国家博士学位
- 1951年回到祖国，任北京大学数学系的教授
- 1956年与华罗庚、钱学森同台领取国家自然科学奖一等奖；38岁时当选为中国科学院学部委员
- 1993年获得陈嘉庚数理科学奖
- 1994年获首届求是科技基金会杰出科学家奖
- 1997年获Herbrand自动推理杰出成就奖
- 2001年获国家最高科学技术奖



自动推理证明发展史：发展

■ 吴文俊



- 1984 年出版专著《几何定理机器证明的基本原理》，利用中国古典数学的成就，提出具有中国特色的定理自动证明方法，被国际上誉为“**吴氏方法**”。
- 1985 年发表论文“关于代数方程组的零点”——吴文俊消元法，即“**吴氏公式**”。
- 2010年5月4日，国际小行星中心发布公报通知国际社会，将国际永久编号为第**7683**号的小行星永久命名为“吴文俊星”。



自动推理证明： 关键要素

- 逻辑方法是自动证明中常用的方法
- 如何进行逻辑推理？
- 推理的过程怎样？
- 怎么实现自动推理？

自动推理证明：推理示例

- 阿加莎.克里斯蒂侦探小说改编的电视剧
 - “马普尔小姐探案”
- 马克和约尔是孪生兄弟
- 谁是作案者：马克或约尔？
- 马普尔小姐的结论

自动推理证明：推理示例



谁是马克谁是约尔？



自动推理证明：推理示例



表
表



■ 只是穿着不同衣服的同一个人——约尔

人类的推理可以理解语义

机器如何进行这样类似的推理？

需要将推理的过程与理解分割开，将其形式化

自动推理证明：一般形式

已知：事实1，事实2， ...

如果 事实1 那么 结论1

如果 事实2 那么 结论2

....

得到：结论1，结论2， ...

- 将事实与规则等抽象出来，不涉及具体内容，借助一些符号来表示，推理过程可以被形式化

P：某已知事实

$P \rightarrow Q$ ：如果 P 那么 Q

结论： Q

这个过程不需要直觉和解释

自动推理证明：符号与形式语言

- 自然语言不适合计算机处理
 - 例：小王不方便接电话，他方便去了
- 需要一种无歧义，方便存储和表达的形式化符号表征体系
 - 数理逻辑
 - 命题逻辑
 - 谓词逻辑

提 纲

■ 自动推理证明

- 自动推理证明发展史
- 谓词逻辑
- 消解原理

■ 基于本体的知识表达及应用

- 本体定义
- 本体构建方法
- 本体应用

■ 基于知识图谱的知识表达及应用

- 知识图谱基础
- 知识图谱构建方法
- 知识图谱应用

谓词逻辑

- 什么是谓词？
 - 原子命题中刻画个体的性质或个体间关系的成分
- 谓词逻辑是一种形式语言
- 是目前为止能够表达人类思维活动规律的一种最精确的语言
- 接近自然语言，又方便存入计算机处理
- 最早应用于人工智能中表示知识
- 适合于表示事物的状态、属性、概念等，也可用来表示事物间确定的因果关系

谓词逻辑

■ Terms (项)

- a. 一个常量是项
- b. 一个变量是项
- c. 如果 f 是一个 n 元函数符号, t_1, t_2, \dots, t_n 是项, 则 $f(t_1, t_2, \dots, t_n)$ 也是项
- d. 所有项都是由规则(a)(b)(c)产生的

■ Atoms (原子公式)

如果 P 是一个 n 元谓词符号, t_1, t_2, \dots, t_n 是项, 则 $P(t_1, t_2, \dots, t_n)$ 是一个原子公式, 其他任何表达式都不是原子公式

谓词逻辑

- WFF（合适公式）
 - a. An atom is WFF
 - b. 如果F和G是WFF，则 $\sim F$ ， $F \wedge G$ ， $F \vee G$ ， $F \rightarrow G$ ， $F \equiv G$ 都是WFF
 - c. 如果F是WFF， x 是自由变量则 $(\forall x)F$ ， $(\exists x)F$ 都是WFF
 - d. WFF仅由有限次使用规则(a)(b)(c)产生。

谓词逻辑： 示例

$\text{man}(\text{smith})$

smith是人

$\text{between}(\text{albert}, \text{susan}, \text{david})$

albert在susan与david之间

$(\forall x)(\text{man}(x) \rightarrow \text{mortal}(x))$

凡是人都会死

$(\exists x)(\text{man}(x) \wedge \text{clever}(x))$

有的人聪明

谓词逻辑：推理过程

- 推理过程多种多样
- 例1：
 - 如果今天不下雨，我就去你家
 - 今天没有下雨
- 例2：
 - 小王说他下午或者去图书馆或者在家休息
 - 小王没去图书馆
- 计算机如何选择？

提 纲

■ 自动推理证明

- 自动推理证明发展史
- 谓词逻辑
- 消解原理

■ 基于本体的知识表达及应用

- 本体定义
- 本体构建方法
- 本体应用

■ 基于知识图谱的知识表达及应用

- 知识图谱基础
- 知识图谱构建方法
- 知识图谱应用

消解原理（归结原理）

- 美国数学家鲁滨逊
提出消解原理（1965年）
基本的出发点：要证明一个命题
为真都可以通过证明其否命题为
假来得到
将多样的推理规则简化为一个—
消解



鲁滨逊

消解原理：基本概念



- 消解只能在仅含否定和析取联接词的公式（子句）间进行
- 必须先把公式化成规范的形式（范式，子句集）

消解原理：示例

■ 例1:

■ 小王说他下午或者去图书馆或者在家休息

■ 小王下午没去图书馆

R—小王下午去图书馆

S—小王下午在家休息

$$\left. \begin{array}{l} R \vee S \\ \neg R \end{array} \right\} \Rightarrow S$$

■ 例2:

■ 如果今天不下雨，我就去你家

■ 今天没有下雨

$$\neg P \rightarrow Q \Leftrightarrow P \vee Q$$

$$\neg P$$

消解原理：含变量的消解

■ 例：苏格拉底论断

凡是人都会死. $\forall x (\text{Man}(x) \rightarrow \text{Mortal}(x))$

苏格拉底是人. $\text{Man}(\text{Socrates})$

如何得到结论：苏格拉底会死.

$\text{Mortal}(\text{Socrates})$

■ 要完成消解还面临几个问题

■ “ \forall ” 和 “ \rightarrow ” 必须消去

■ $\text{Man}(x) \rightarrow \text{Mortal}(x) \Leftrightarrow \neg \text{Man}(x) \vee \text{Mortal}(x)$

■ “ \forall ” 怎么办？

} 化为子句集

■ 如果能消去 “ \forall ” , $\neg \text{Man}(x)$ 和 $\text{Man}(\text{Socrates})$ 也不能构成互补对, 形式不一样, 怎么办?

} 置换与合一

消解原理：置换与合一

- 要把消解推理规则推广到含有变量的子句，必须找到一个作用于亲本子句的置换，使亲本子句含有互补文字

- 置换（**Substitution**）

- 置换是形为 $\{t_1/x_1, t_2/x_2, \dots, t_n/x_n\}$ 的一个有限集
- x_i 是互不相同的变元， t_i 是项
- 用 t_i 代换 x_i ，不允许 t_i 与 x_i 相同，也不允许变元 x_i 循环出现在另一个 t_j 中

$\{a/x, f(b)/y, w/z\}$ $\{f(a)/x, b/y, t/x\}$ $\{g(y)/x, f(x)/y\}$

$s = \{z/x, w/y\}$ $\omega = P(x, f(y), B)$ $\omega s = P(z, f(w), B)$

消解原理：置换与合一

■ 合一 (Unification)

- 寻找项对变量的置换，以使两表达式一致的过程。
- 如果一个置换 s 作用于表达式集 $\{E_i\}$ 的每个元素，则我们用 $\{E_i\} s$ 来表示置换例的集。我们称表达式集 $\{E_i\}$ 是可合一的 (unifiable), s 为合一者 (unifier)
- mgu (most general unifier, 最一般合一者)
 - 若 s 为 $\{F_i\}$ 的任一合一者，又存在某个置换 s' ，使得 $s=gs'$ ，则称 g 为 $\{F_i\}$ 的最一般(最简单的)合一者，记作mgu。
 - mgu is unique

$$F=\{P(x,f(y),B)\}$$

$$s=\{A/x,B/y\} \quad g=\{B/y\}$$

$$s'=\{A/x\} \quad gs'=s$$

消解原理：化为子句集

■ 相关概念

- 文字：原子公式及其否定统称为文字

- 子句集

(1)子句定义

任何文字的析取式称为子句

不包含任何文字的子句称为空子句（空子句是永假的）

由子句构成的集合称为子句集

例： $\{P(x) \vee Q(x), \sim P(x, f(x)) \vee Q(x, g(x))\}$

消解原理：化为子句集

(2)谓词演算公式化为子句式

任何一个谓词演算公式可以化为一个子句集合

步骤：

1)消去蕴涵符号

用 $\sim A \vee B$ 代换 $A \rightarrow B$

2)把非号 \sim 移入内层

$$\sim (P \wedge Q) = \sim P \vee \sim Q$$

$$\sim (P \vee Q) = \sim P \wedge \sim Q$$

$$\sim (\forall x)P = (\exists x) \sim P$$

$$\sim (\exists x)P = (\forall x) \sim P$$

消解原理：化为子句集

3)对变量标准化

改变变量名，使不同的变量不同名

$$(\forall x)P(x) \vee (\exists x)Q(x)$$

$$(\forall x)P(x) \vee (\exists y)Q(y)$$

4)消去存在量词(具体化 Skolemizing), 两种情况:

1. 存在量词不在全称量词的辖域内——用新的个体常量替换受存在量词约束的变元
2. 存在量词在全称量词的辖域内

Skolem函数，即具体化函数

$$(\forall x)P(x) \vee (\exists y)Q(y) \Rightarrow (\forall x)P(x) \vee Q(a)$$

$$(\forall x_1)(\forall x_2)\dots(\forall x_n)(\exists y)P(x_1, x_2, \dots, x_n, y)$$

$$\Rightarrow (\forall x_1)(\forall x_2)\dots(\forall x_n)P(x_1, x_2, \dots, x_n, f(x_1, x_2, \dots, x_n))$$

消解原理：化为子句集

5)化为前束形式

把全称量词提到最外层

前束形:= (前缀) {母式}

↑ ↑

全称量词串 无量词公式

6)把母式化为合取范式

7)消去全称量词

8)消去连词符号 \wedge ，写成子句集

9)变量分离标准化

改变变量名称，使一个变量符号不出现在一个以上的子句中

消解原理：消解式

■ 命题逻辑的消解式

- 设C1与C2是子句集中的任意两个子句，如果C1中的文字L1与C2中的文字L2互补，那么从C1和C2中分别消去L1和L2，并将两个子句中余下的部分析取，构成一个新子句C12，则称这一过程为消解，称C12为C1和C2的消解式，C1，C2为C12的亲本子句

例：子句 $C1 = P \vee C1'$ $C2 = \sim P \vee C2'$

消解式 $C12 = C1' \vee C2'$

消解原理：利用消解式进行证明

- 消解反演
- 通俗的说就是“反证法”
 - 要证命题 A 恒为真，等价于证 $\neg A$ 恒为假
- 证明过程
 - 否定结论 R ，得 $\neg R$ ；
 - 把 $\neg R$ 添加到已知前提集合 F 中去；
 - 把新产生的集合 $\{ \neg R, F \}$ 化成范式；
 - 应用消解原理，不断求消解式，直到得到一个表示矛盾的空子句

消解原理：消解反演示例

假设：所有不贫穷并且聪明的人都是快乐的，那些看书的人是聪明的。李明能看书且不贫穷，快乐的人过着激动人心的生活。

求证：李明过着激动人心的生活。

解：先定义谓词：

Poor(x) x是贫穷的

Smart(x) x是聪明的

Happy(x) x是快乐的

Read(x) x能看书

Exciting(x) x过着激动人心的生活

消解原理：消解反演示例

■ 问题谓词表示：

- “所有不贫穷并且聪明的人都是快乐的”

$$(\forall x)((\sim \text{Poor}(x) \wedge \text{Smart}(x)) \rightarrow \text{Happy}(x))$$

- “那些看书的人是聪明的”

$$(\forall y)(\text{Read}(y) \rightarrow \text{Smart}(y))$$

- “李明能看书且不贫穷”

$$\text{Read}(\text{Liming}) \wedge \sim \text{Poor}(\text{Liming})$$

- “快乐的人过着激动人心的生活”

$$(\forall z)(\text{Happy}(z) \rightarrow \text{Exciting}(z))$$

- 目标 “李明过着激动人心的生活” 的否定

$$\sim \text{Exciting}(\text{Liming})$$

消解原理：消解反演示例

- 将上述谓词公式转化为子句集如下：

(1) $\text{Poor}(x) \vee \sim \text{Smart}(x) \vee \text{Happy}(x)$

(2) $\sim \text{Read}(y) \vee \text{Smart}(y)$

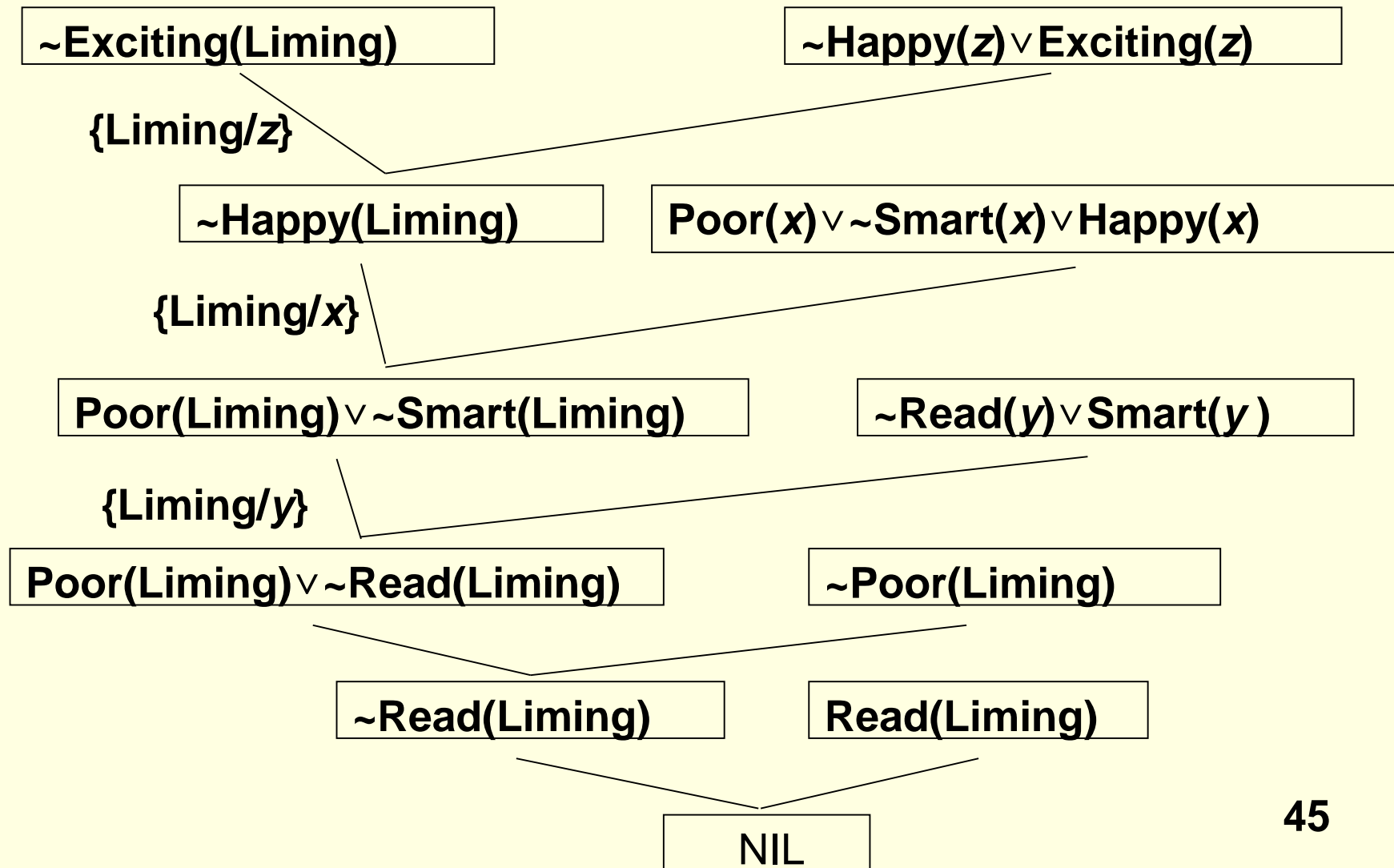
(3) $\text{Read}(\text{Liming})$

(4) $\sim \text{Poor}(\text{Liming})$

(5) $\sim \text{Happy}(z) \vee \text{Exciting}(z)$

(6) $\sim \text{Exciting}(\text{Liming})$ (结论的否定)

消解原理：消解反演示例



消解原理：局限性

- 消解原理推进了用逻辑方法进行机器证明的研究，使得自动定理证明领域发生了质的变化。

- 但是

- 要求把逻辑公式转化为某种范式，丧失了其固有的逻辑蕴含语义。例如：

如果一个人发烧、肚子痛，那么很可能是感染了。

$\forall x (\text{has_fever}(x) \wedge \text{tummy_pain}(x) \rightarrow \text{has_an_infection}(x))$
 $\neg \text{has_fever}(x) \vee \neg \text{tummy_pain}(x) \vee \text{has_an_infection}(x)$

- 表达能力的局限性, 限制了应用范围
 - 后来有许多重要改进：语义归结（消解）、锁归结（消解）、线性归结（消解）等。

提 纲

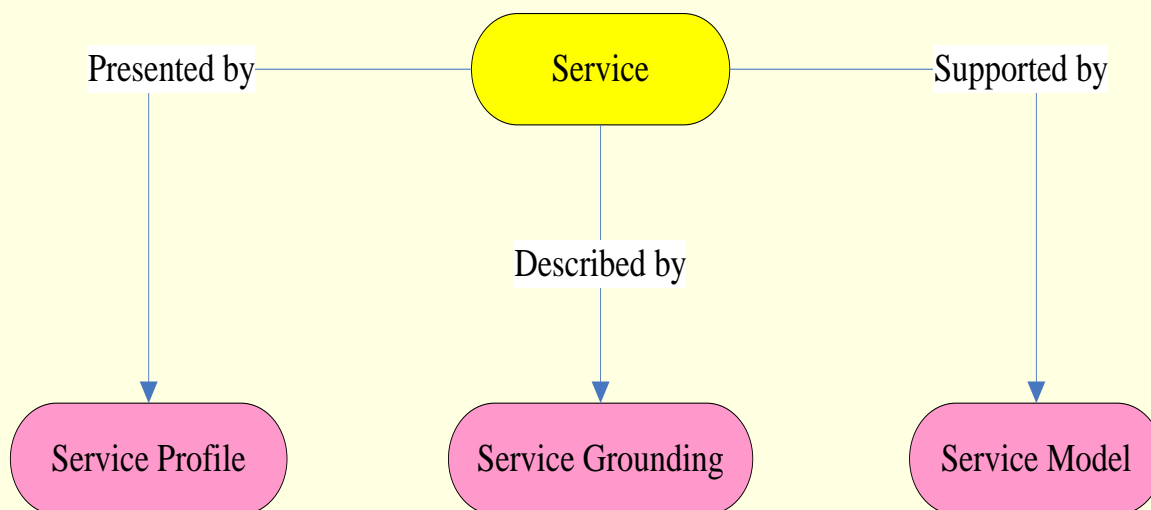
- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

本体定义:

- ◆ 语义Web服务具影响力的代表性工作包括:
 - ◆ WSMO[59, 60](Web Service Modeling Ontology): 基于框架逻辑, 从四个方面描述语义Web服务: Ontology, Web Service, Goal和Mediator;
 - ◆ WSDL-S[61]: 独立于语义表达语言, 允许开发人员选择本体语言, 如OWL或者UML;
 - ◆ OWL-S[26, 58] (Web Ontology Language for Services): 前身为DAML-S, 基于可判定的描述逻辑, 将Web服务的本体分成三个上层本体。

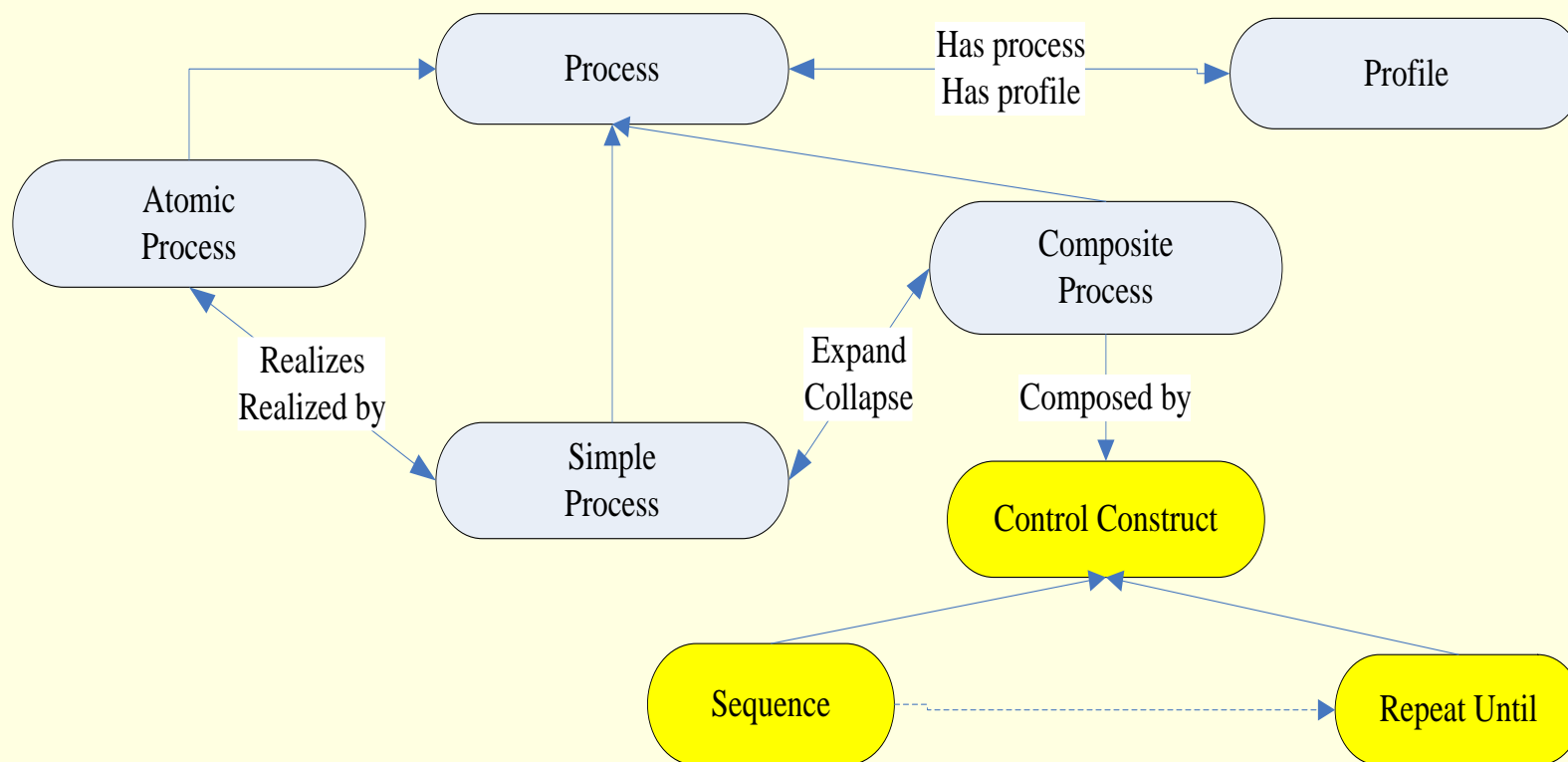
本体定义:

◆ OWL-S顶层本体视图



本体定义:

◆ OWL-S的Service Model



提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

本体构建：形式化

- 五元组表示： $O = \{C, R, F, A, I\}$
 - C – Concepts
 - 概念集合，通常以Taxonomy形式组织
 - 例如：武大校友
 - R – Relations
 - 描述概念或实例之间语义关系的集合
 - 例如：subClassOf, birthplace
 - F – Functions
 - 一组特殊的关系，关系中第n个元素的值由其他n-1个元素的值共同确定
 - 例如：price-of-a-used-car由car-model, manufacturing-date 和 kilometers确定

本体构建：形式化

- 五元组表示： $O = \{C, R, F, A, I\}$
 - A – Axioms
 - 公理
 - 例：如果A是B的子女，B是C的子女，则A是C的子孙
 - I – Instances
 - 描述具体的个体
 - 例：Peter是概念学生的实例

本体构建：描述方法

■ 资源描述框架RDF

■ Resource Description Framework

■ RDF数据模式

■ 资源 Resource

- 使用URI唯一标识一个资源
- 一个资源通常表示一个事物（Thing）

■ 属性 Property

- 一种特殊类型的资源，用以描述资源与资源见的关系

■ 语句 Statement

- 由3种资源组成的三元组(Triple)
- 主语rdf:subject, 谓语rdf:predicate以及宾语rdf:object

本体构建：形式化示例

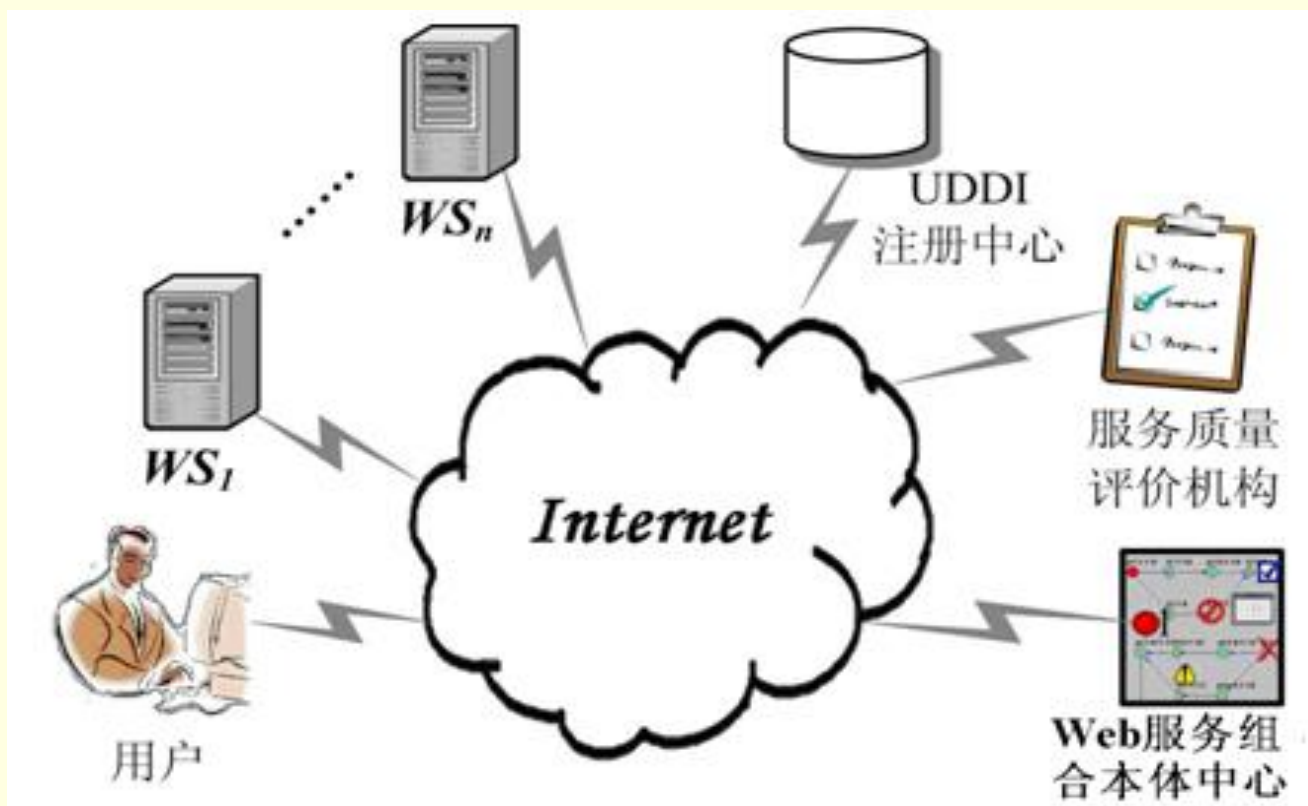
- 本体的简化形式： $O = \{C, I, T, P\}$
 - C – Concepts
 - 描述领域中的抽象概念，通常以Taxonomy形式组织
 - 例：学生和老师是两个概念
 - I – Instances
 - 描述具体的实例
 - 例：学生Peter是概念学生的实例
 - T – ISA
 - 概念与概念之间、实例与概念之间的关系
 - 例：subClassOf关系和instanceOf关系
 - P – Properties
 - 本体中用于描述实例信息的其他语义关系
 - 例：instance-attribute-value (AVP)

提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

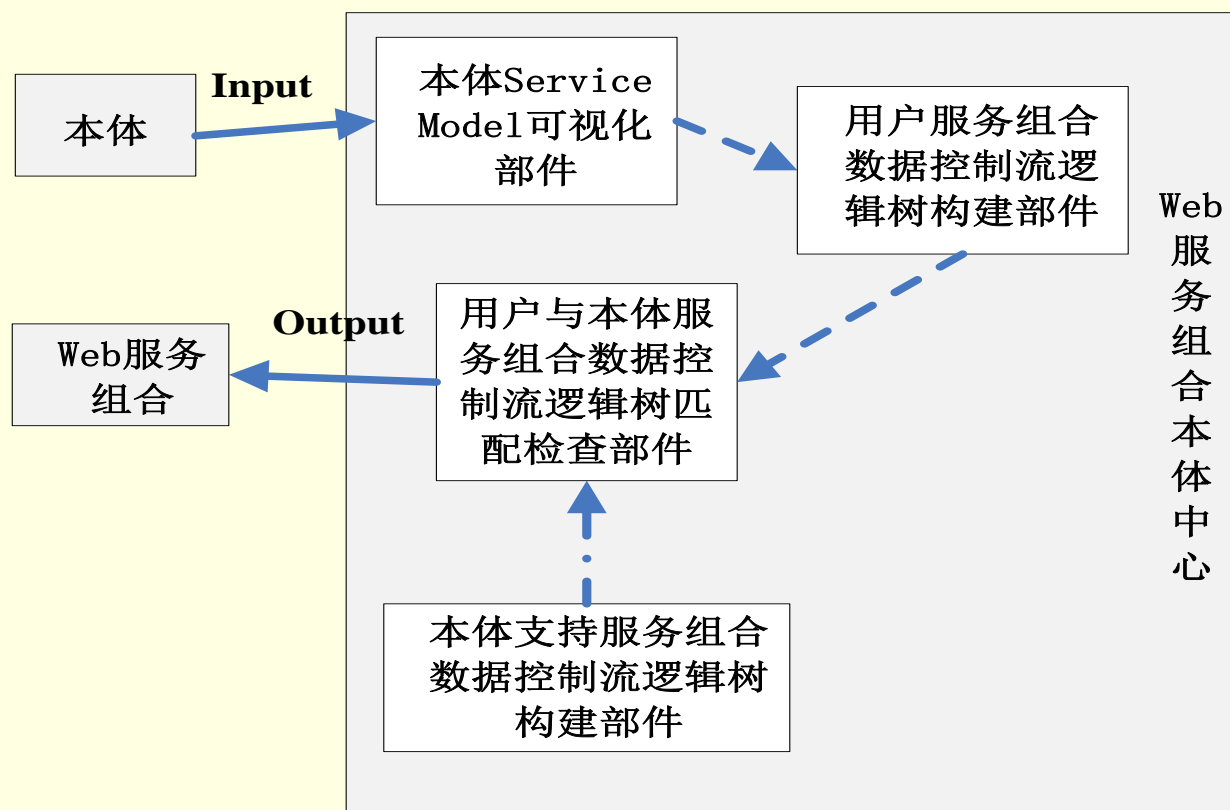
本体应用：

- ◆ 一种基于本体的Web服务组合概念建模应用框架



本体应用：

◆一种基于本体的Web服务组合概念建模应用框架



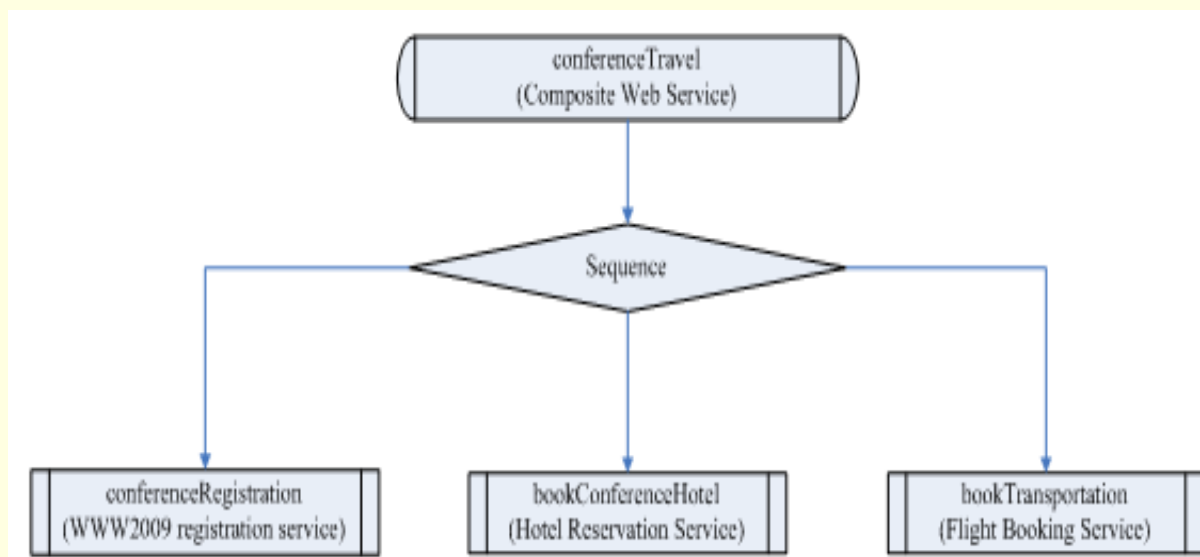
本体应用：

◆ Web服务组合本体中心的内部工作机制：

- ◆ 步骤一：对本体的**Service Model**进行可视化；
- ◆ 步骤二：通过用户对本体可视化组件的拖拽操作，构建符合用户需求的服务组合数据控制流逻辑树；
- ◆ 步骤三：构建本体内所支持的所有服务组合数据控制流逻辑树集合；
- ◆ 步骤四：将用户服务组合数据控制流逻辑树与本体内所支持的所有服务组合数据控制流逻辑树集合中的所有元素进行一致性校验，若两者不相符，则提示用户进行修改，直到符合本体要求为止，若两者相符，则输出符合**BPEL4WS**规范的服务组合执行脚本。

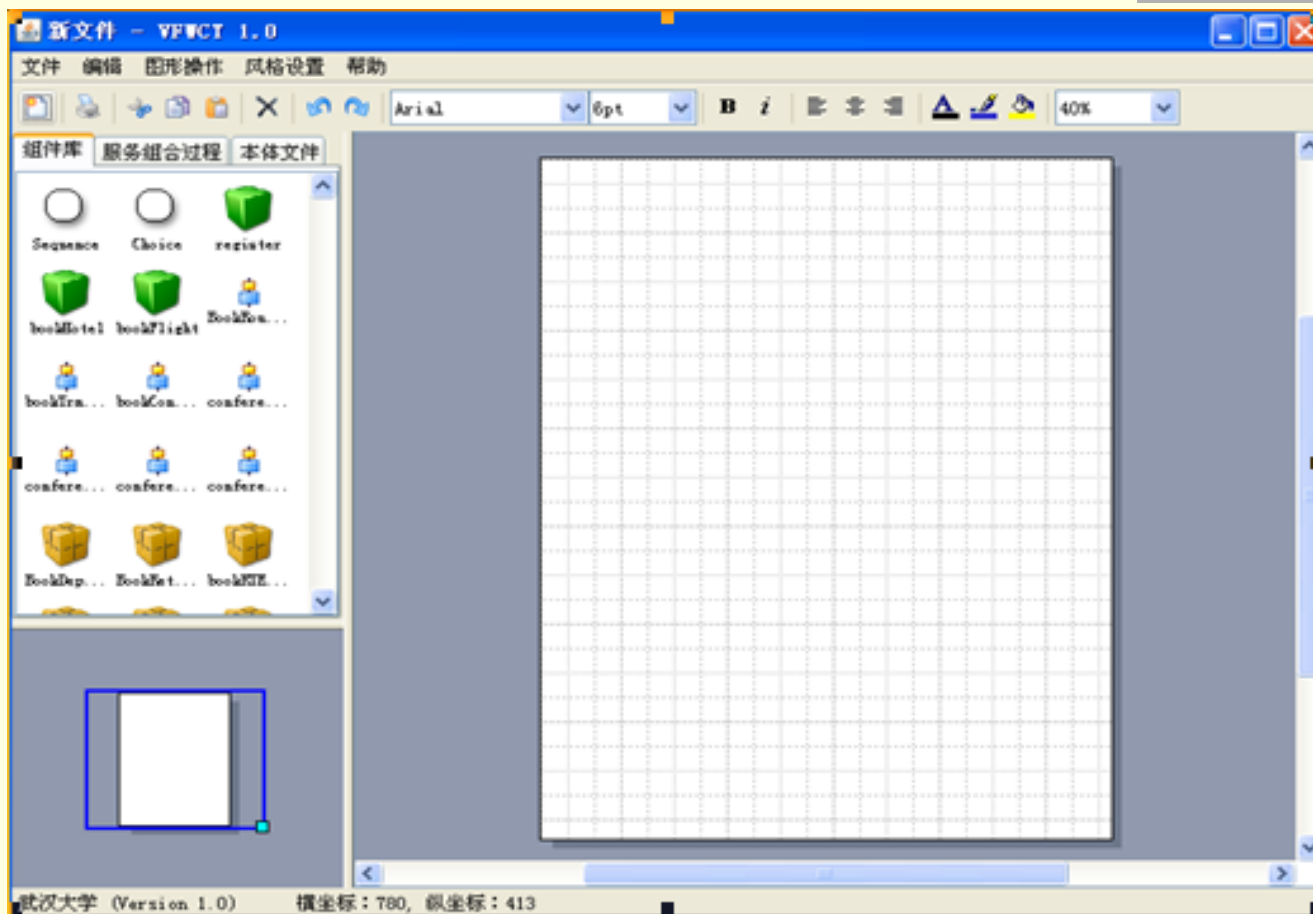
本体应用：

使用来自MINDSWAP的关于参加学术会议、预定往返机票及预定旅馆的Web服务组合本体conferenceTravel.owl作为介绍本文所提出的基于OWSCCM模型的Web服务组合概念建模方法及其实现工具VFWCT的实例研究。



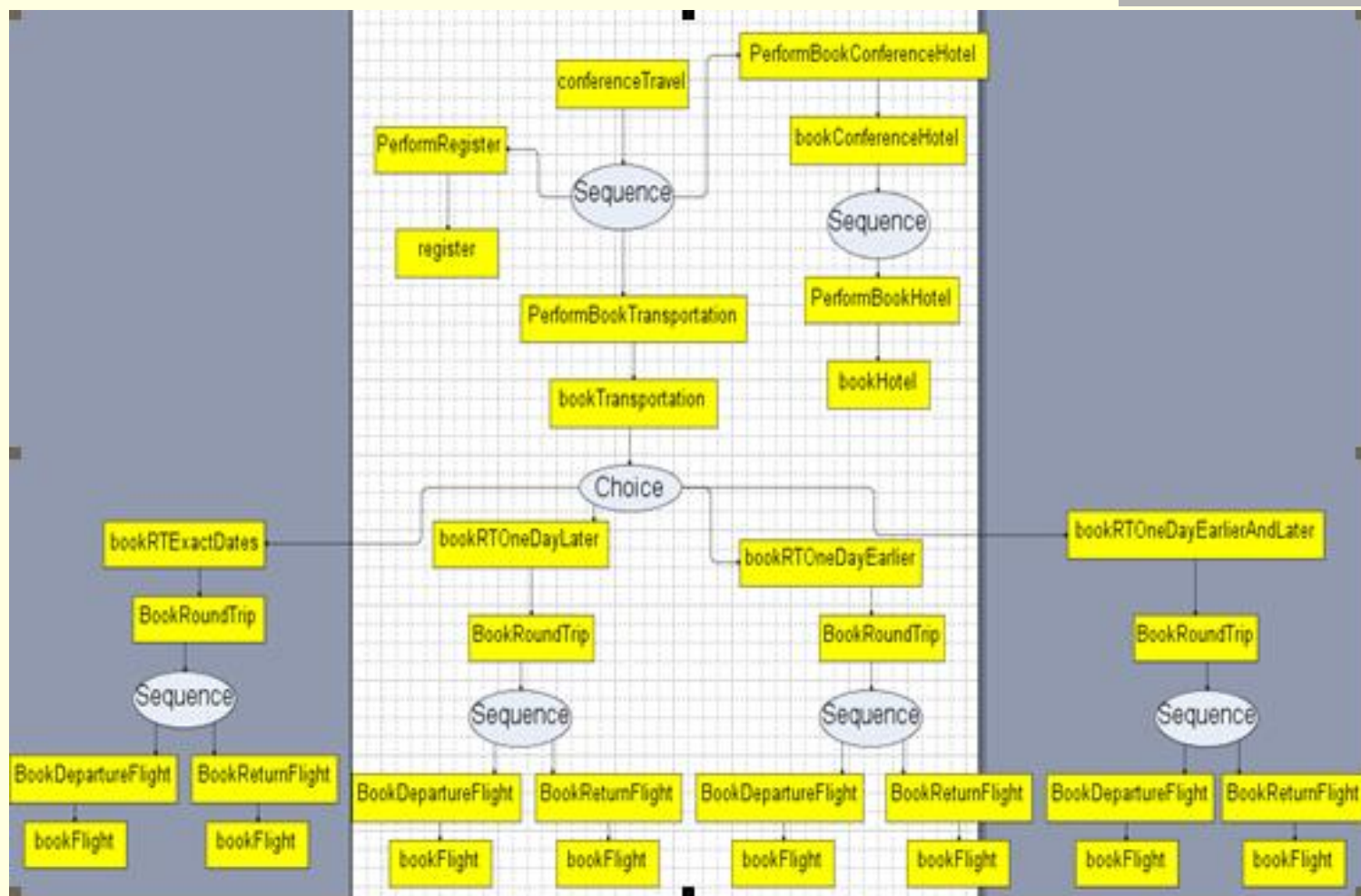
学术会议行程安排Web服务组合

本体应用:



Service Model元素可视化界面

本体应用:



用户服务组合控制数据流逻辑树

本体应用:



请选择QoS参数约束

要求

响应时间：高

成本：较高

安全性：中

可靠性：低

信誉度：高

确定 取消

This dialog box is titled "请选择QoS参数约束" (Please select QoS parameter constraints). It contains a section labeled "要求" (Requirements) with five parameters, each with a dropdown menu: "响应时间" (Response time) set to "高" (High), "成本" (Cost) set to "较高" (Medium-High), "安全性" (Security) set to "中" (Medium), "可靠性" (Reliability) set to "低" (Low), and "信誉度" (Reputation) set to "高" (High). At the bottom are "确定" (OK) and "取消" (Cancel) buttons.

QoS参数约束选择界面



请输入启动参数

conferenceItravel_conference: WWW2009

conference_startDate: 2009-08-01

conference_endDate: 2009-08-10

conference_city: Wuhan

conferenceItravel_city: Beijing

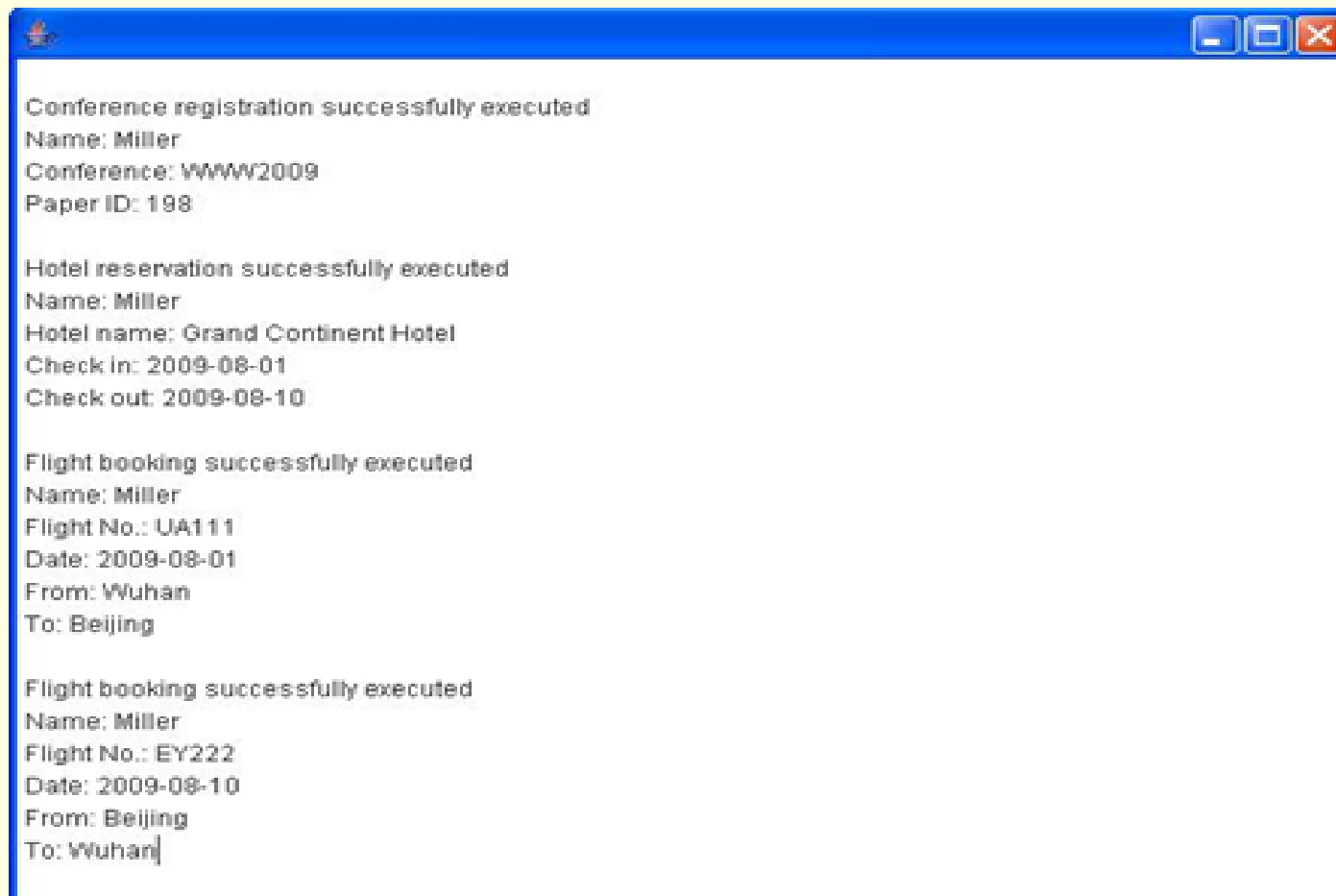
conferenceItravel_person: Miller

提交 取消

This dialog box is titled "请输入启动参数" (Please enter startup parameters). It contains six text input fields with labels: "conferenceItravel_conference" (WWW2009), "conference_startDate" (2009-08-01), "conference_endDate" (2009-08-10), "conference_city" (Wuhan), "conferenceItravel_city" (Beijing), and "conferenceItravel_person" (Miller). The date fields have dropdown arrows. At the bottom are "提交" (Submit) and "取消" (Cancel) buttons.

服务组合启动参数输入界面

本体应用：



会议旅行服务组合运行结果

提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

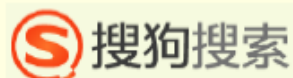
知识图谱基础:



250概念
4M实例
6000属性
500M三元组
在线更新



350K概念
10M实例
100属性
120M三元组



15K概念
40M实例
4000属性
1B三元组
Google KB核心



50M义项
50+种语言
262M三元组



850K概念
8M实例
70K属性



Google KG

15K概念
600M实例
20B三元组

WordNet
7种欧洲语言
跨语言链接



NELL

OpenIE
(Reverb, OLLIE)

知识图谱基础:

■ 知识图谱 --- 百度百科

- 知识图谱，也称为科学知识图谱，它通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合，并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。为学科研究提供切实的、有价值的参考。

■ Google知识图谱 --- 维基百科

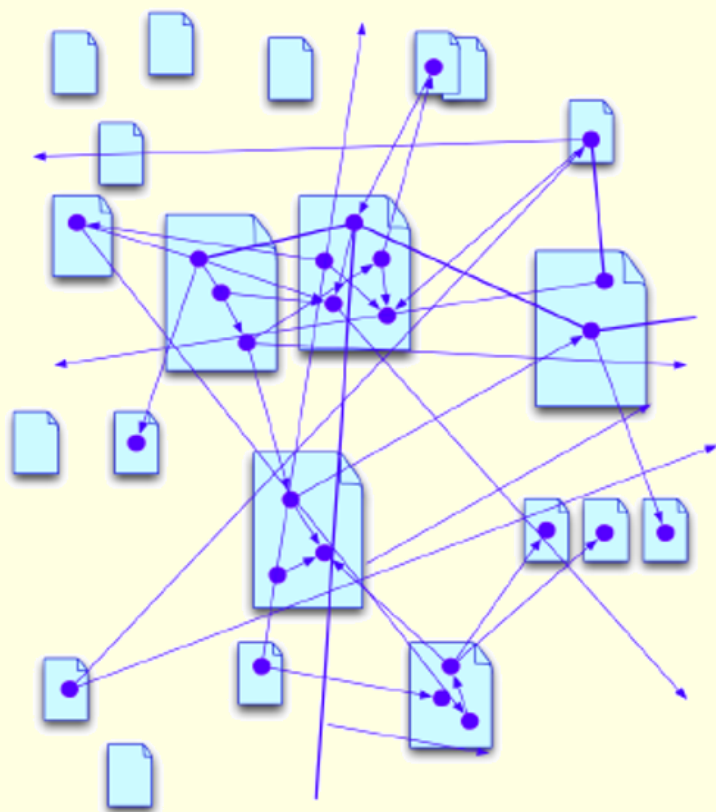
- 实体及其之间的关系图。
- 规模：5亿个对象，35亿个事实和关系



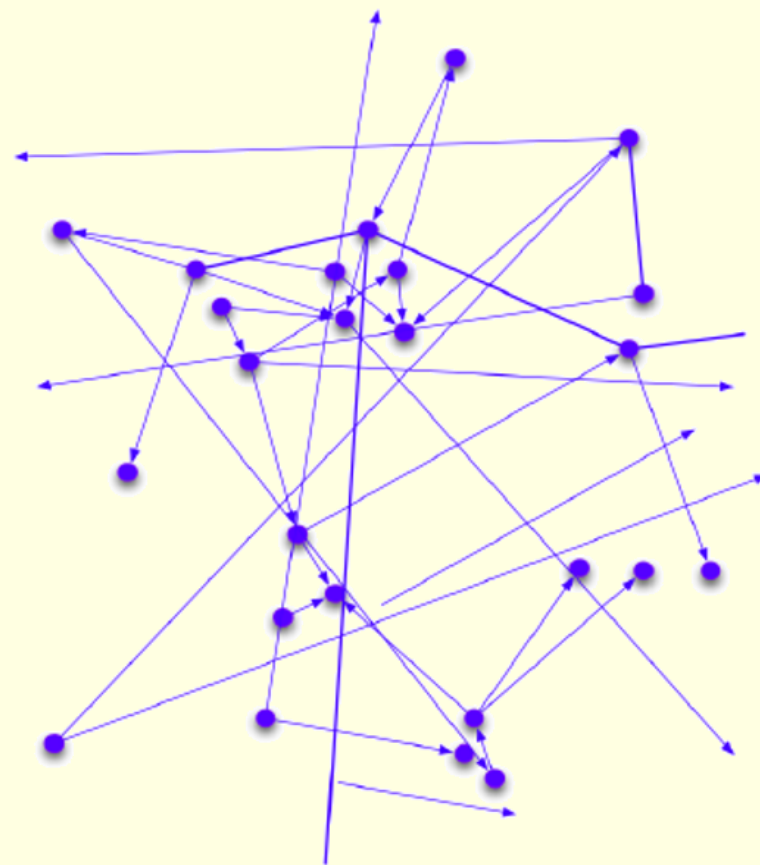
知识图谱的本质：知识库？语义网络？
知识图谱的形式：RDF？Graph？

知识图谱基础：文档万维到数据万维

文档万维网



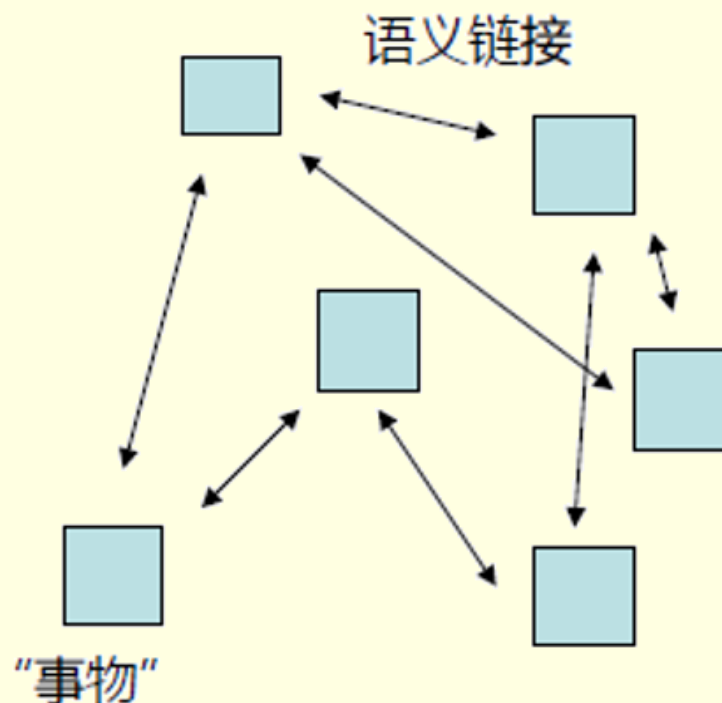
数据万维网



知识图谱基础：数据万维网

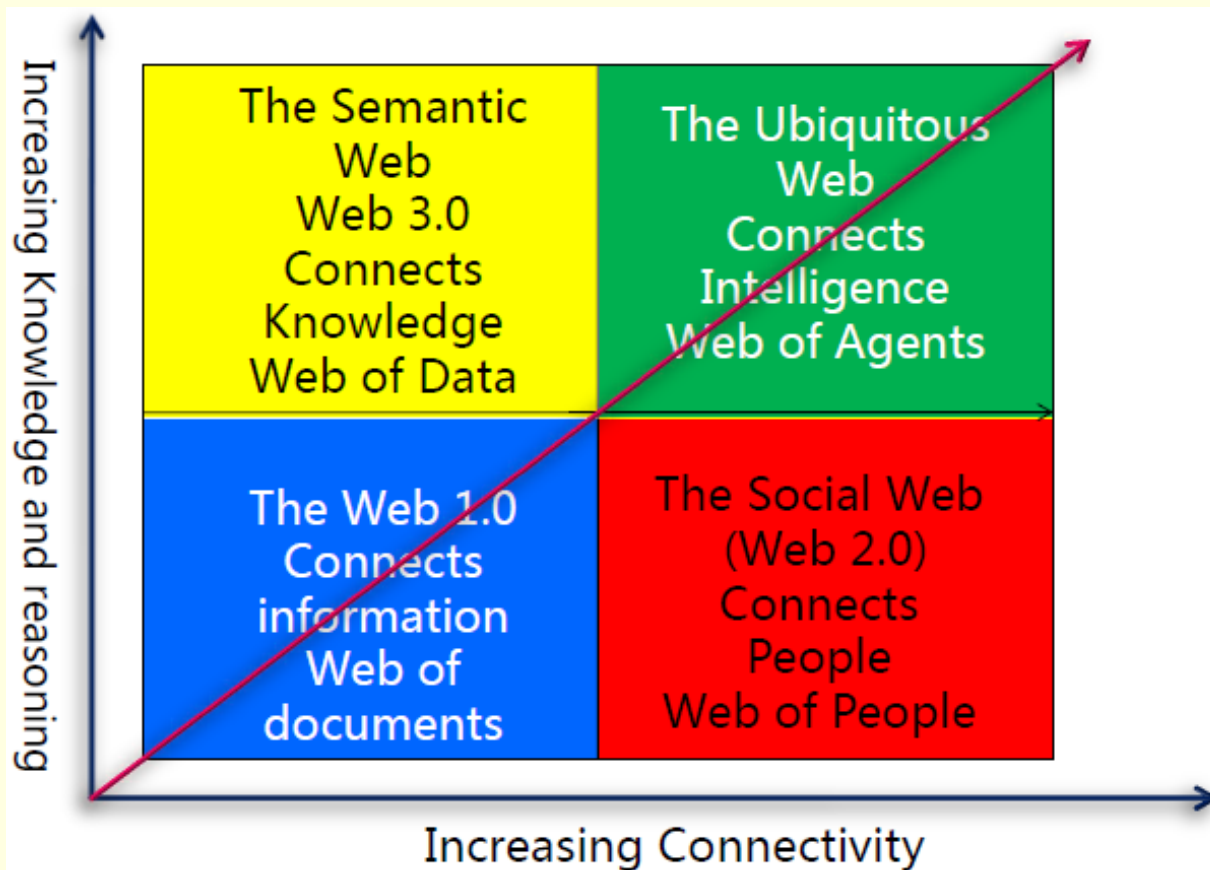
■ 特征：

- Web上的事物拥有唯一的URI
- 事物之间由链接关联（如人物、地点、事件、建筑物）
- 事物之间链接显式存在并拥有类型
- Web上数据的结构显式存在



知识图谱基础：万维网的发展

- Agent Webs that know, learn and reason as human do



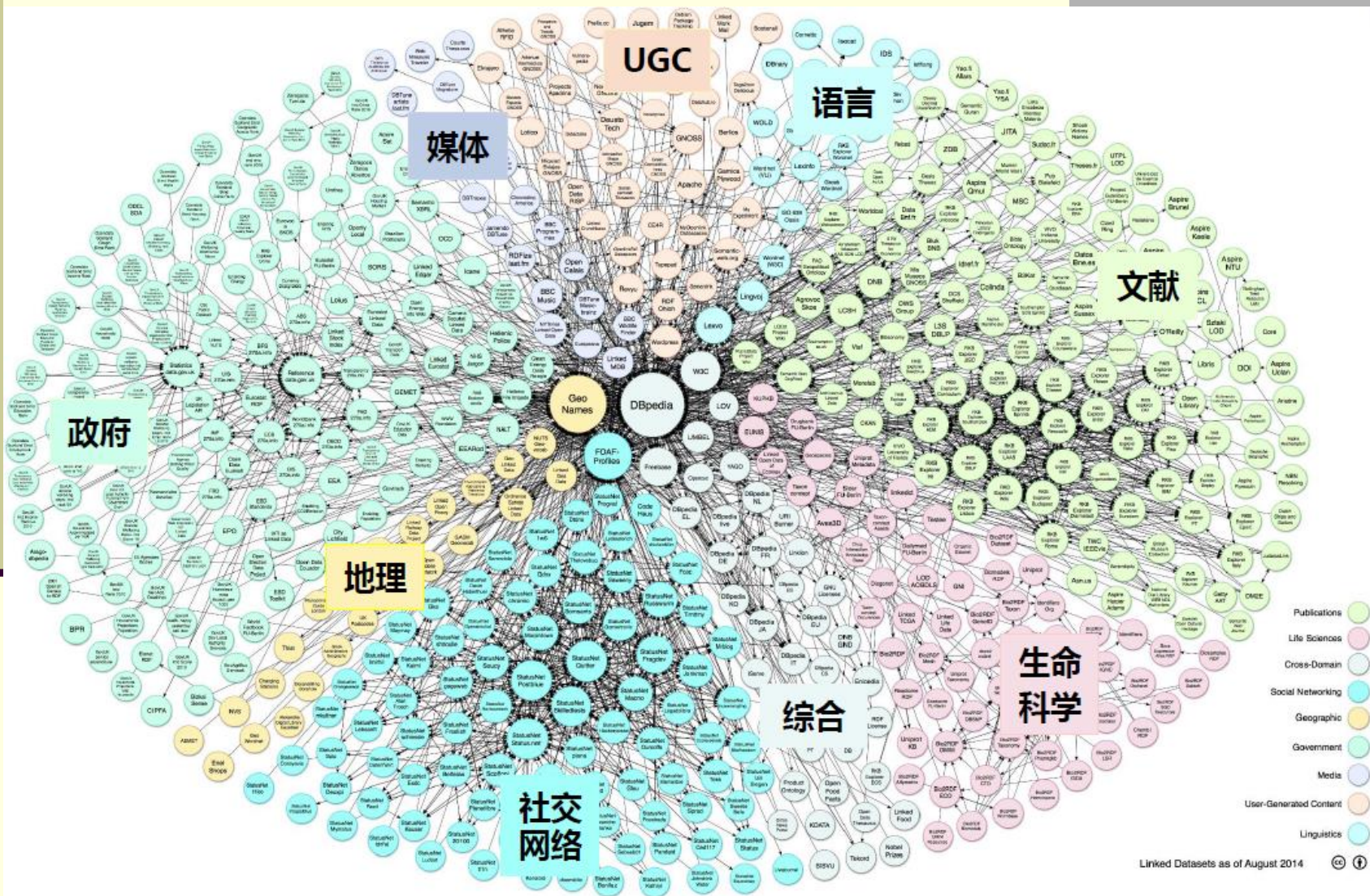
知识图谱基础：领域无关知识图谱

类别	名称	其他
人工构建	ResearchCyc	http://www.cyc.com/platform/researchcyc
	WordNet	wordnet.princeton.edu
基于维基百科	DBPedia	dbpedia.org
	YAGO	yago-knowledge.org
	Freebase	freebase.com
	WikiTaxonomy	http://www.hits.org/english/research/nlp/download/wikitaxonomy.php
	BabelNet	babelnet.org
开放知识抽取	KnowItAll	openie.cs.washington.edu
	NELL	rtw.ml.cmu.edu
	Probase	http://research.microsoft.com/en-us/projects/probase/
中文知识图谱	百度知心	www.baidu.com
	搜狗知立方	www.sogou.com

知识图谱基础：领域无关知识图谱

<i>name</i>	<i># of concepts</i>	<i># of isA pairs</i>
Freebase	1,450	24,483,434
WordNet	25,229	283,070
WikiTaxonomy	111,654	105,418
YAGO	352,297	8,277,227
DBPedia	259	1,900,000
ResearchCyc	≈ 120,000	< 5,000,000
KnowItAll	N/A	< 54,753
TextRunner	N/A	< 11,000,000
OMCS	173,398	1,030,619
NELL	123	< 242,453
Probase	2,653,872	20,757,545

知识图谱基础：特定领域知识图谱



提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

结构化数据转化为语义资源：

■ 结构化数据

- 大部分结构化数据都被存储在关系型数据库中。

■ 将结构化数据转化为知识的RDF描述—D2R

- D2R是一种XML-based语言，用来达到上述的映射目标
- D2R 映射步骤

- 从关系型数据库中选取一个或者一组相似的类
- 把选取中的记录按列分组
- 为每个实例进行URI或者blank node分配
- 为每个Instance 创建属性

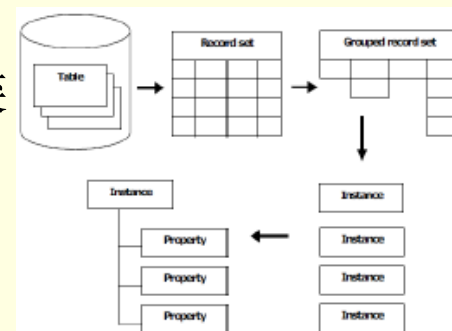


Figure 1. The D2R mapping process.

D2R映射过程

D2R MAP - A Database to RDF Mapping Language. WWW (Posters) 2003

半结构化知识抽取:

■ Taxonomy知识抽取

- 从Web Table中提取实例与概念之间的上下位语义关系

■ 演职员表 编辑

演员表

角色	演员	备注
耿浩	黄渤 ^[6]	二手音响店老板，过气歌手
郝义	徐峥	耿浩的发小，知名制片人；乐天派，花花公子，泡妞达人
康小雨	袁泉	文艺小清新的女汉子
周丽娟	周冬雨	90后小镇杀马特姑娘
莎莎	马苏	东北老妹儿，泼辣火爆的特殊服务人员
东东	陶慧	阿凡达女郎，草台班子的舞蹈演员
旅店老板	岳小军	低调的艺术大叔



半结构化知识抽取:

■ Taxonomy知识抽取

1. 给出 “种子 (seeds) ” 作为搜索的起始。

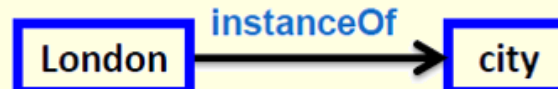
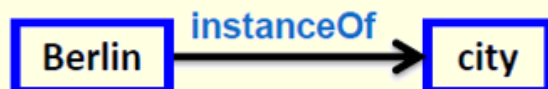
`cities = {Paris, Shanghai, Brisbane}`

2. 搜索包含一个或多个 “种子” 的表格

Paris	France
Shanghai	China
Berlin	Germany
London	UK

Paris	Iliad
Helena	Iliad
Odysseus	Odysee
Rama	Mahabaratha

3. 从表格中抽取概念-实例关系



非结构化资源的知识学习:

■ Open Information Extraction

- 学习一般性模型来表示关系
- 学习领域相关正则表达式

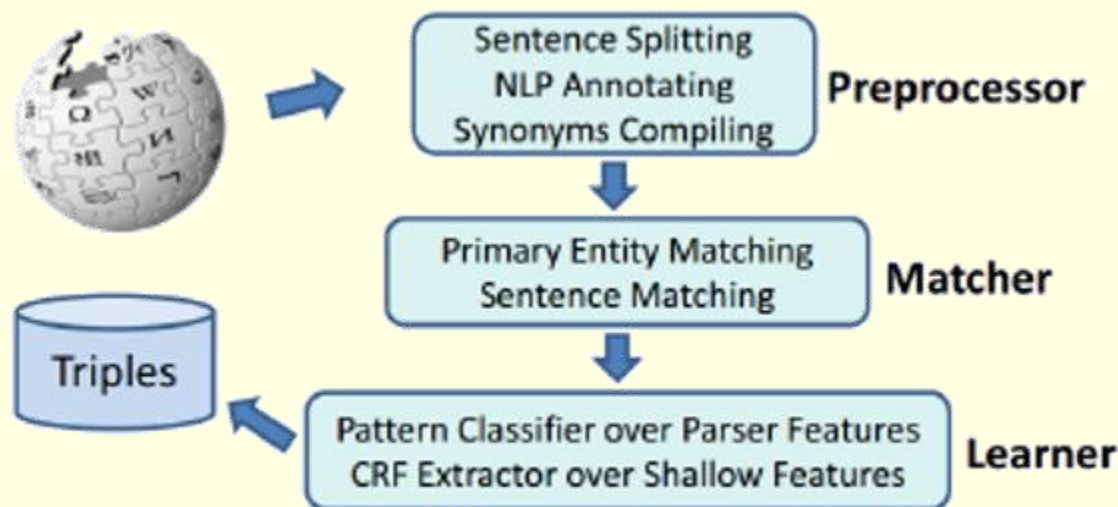


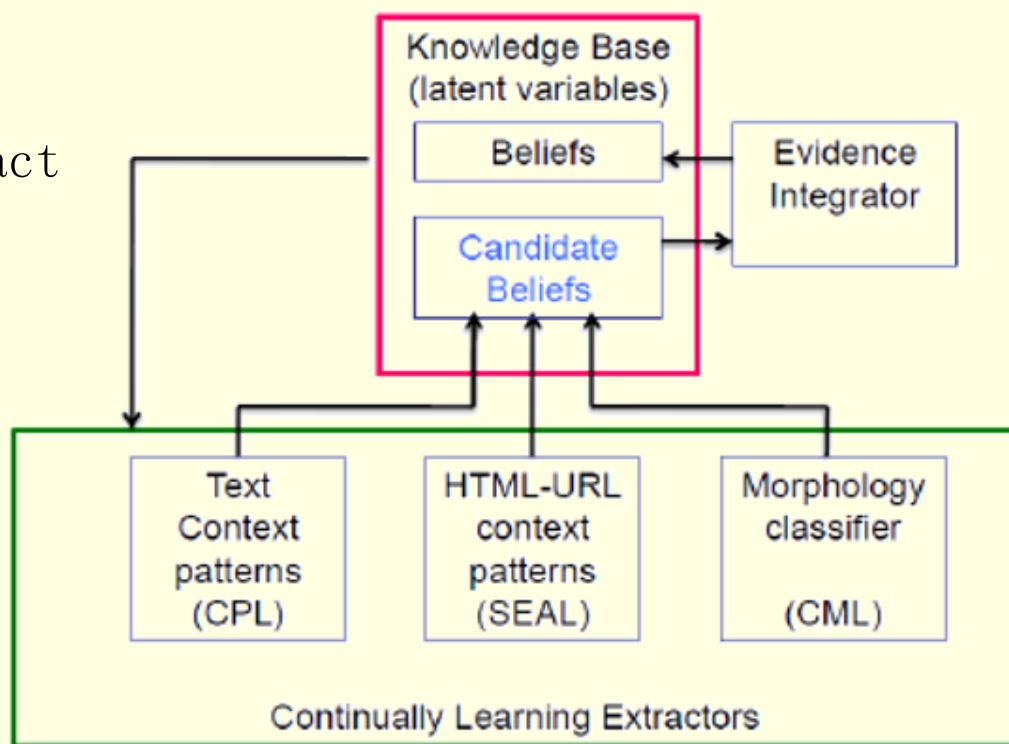
Figure 1: Architecture of WOE.

Open information extraction using Wikipedia . ACL 10

非结构化资源的知识学习：

■ NELL (Never Ending Language Learning)

- 大规模信息抽取系统
- 500-600个概念和关系
- 3.2M的“低可信度” fact
- 500K高可信度的fact



NELL系统图

非结构化资源的知识学习：

■ Probase (A Probabilistic Knowledgebase)

■ 目标

- 通过注入“一般性知识”到计算机中，来更加了解人的交流，形成知识库

■ 建立Probase

- 先用迭代的方法建立核心的taxonomy
- 找到哪些属性和哪些类可以用来回答哪些问题，比如（中国，人口，14亿）可以用来回答“中国居住了多少人？”，虽然“人口”没有出现在问题中
- 用一个非监督的bootstrapping 算法反复扫描网页文档集合来得到很多instance的关系
- 用一种概率的数据集成机制来融合目前已有的结构化数据，例如Freebase, IMDB, Amazon

Probase: a probabilistic taxonomy for text understanding. SIGMOD 2012

知识图谱构建方法：小结

- 互联网海量数据中富含各类知识
- 不同类型的资源需要不同的知识图谱获取方法
- 知识获取仍然面临很多挑战
 - 增量式知识获取
 - 长尾知识获取
 - 知识的演化
 - 高精确度的知识获取
 - 开放taxonomy 知识的建立
 - 异构知识的处理
 - 利用群体智慧的知识获取
 - 推理规则的学习
 -

提 纲

- 自动推理证明
 - 自动推理证明发展史
 - 谓词逻辑
 - 消解原理
- 基于本体的知识表达及应用
 - 本体定义
 - 本体构建方法
 - 本体应用
- 基于知识图谱的知识表达及应用
 - 知识图谱基础
 - 知识图谱构建方法
 - 知识图谱应用

知识图谱应用：语义数据集成

搜狗百科

李娜

搜索词条

编辑词条

收藏

分享

投诉

登录

李娜

编辑词条

图册



李娜(8)

李娜（1982年2月26日-），出生于湖北武汉，前中国女子网球运动员，亚洲首位大满贯单打冠军得主，亚洲第一位入选国际网球名人堂的球员。

2011年，李娜在法网公开赛封后；之后，她三次跻身澳网决赛，并于2014年夺冠后退役。^[1]在其15年职业生涯中，21次打入WTA女单赛事决赛，共获得了9个WTA和19个ITF单打冠军。

快速导航

大事记

光影集锦

荣誉记录

词条图册

关系表

城市百科

知 知乎精选

中文名 李娜

外文名 Li Na

别名 娜姐

毕业院校 华中科技大学（新闻学专业）

运动项目 网球

所属运动队 中国网球队

出生地 湖北省武汉市江岸区

出生日期 1982年2月26日

国籍 中国

性别 女

民族 汉

星座 双鱼座



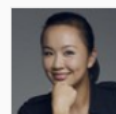
姜山



曾毅



杨魏玲花



彭帅



张帅



凤凰传奇

申请成为词条顾问 ?

您可以享受以下特权：



私人定制
个人百科



个人品牌
跨屏展示



万元顾问
基金奖励

词条信息

创建者：孙阳阳西红柿

编辑次数：256 次

词条浏览：867778 次

最近更新：19.10.28

相关搜索

知识图谱应用：互联网语义搜索

- 在网络搜索时，经常会出现多义的词条。如“李娜”可表示网球运动员李娜和歌手李娜。通常搜索结果会以结果列表的形式给出。



知识图谱的语义链接，使得搜索引擎可以用基于实体的搜索来代替基于字符串的搜索，从而实现搜索时的歧义消除。

知识图谱应用：互联网语义搜索

Google

cards of house

全部

图片

影片

购物

地图

更多

设定

工具

約 2,420,000,000 項搜尋結果 (0.79 秒)

House of Cards (American TV series) - Wikipedia

[https://en.wikipedia.org/wiki/House_of_Cards_\(America...\)](https://en.wikipedia.org/wiki/House_of_Cards_(America...)) [翻譯這個網頁](#)

House of Cards is an American political thriller web television series created by Beau Willimon. It is an adaptation of the 1990 BBC miniseries of the same title and based on the novel of the same title by Michael Dobbs. The first 13-episode season was released on February 1, 2013, on the streaming service Netflix.

No. of seasons: 6 Original network: Netflix

Related shows: House of Cards Production location(s): Baltimore, Maryland, Jo...

[Plot](#) · [Cast and characters](#) · [Production](#) · [Reception](#)

紙牌屋- 维基百科，自由的百科全书

<https://zh.wikipedia.org/wiki/紙牌屋> [轉為繁體網頁](#)

《紙牌屋》（英語：House of Cards）是美國一部政治權謀題材的電視連續劇，於網路上播出，是Netflix原創劇集。本劇由鮑爾·威利蒙（英語：Beau Willimon）創作并改編自...

季数: 6 集数: 73

每集长度: 43–59 分钟 发行公司: Netflix

[演員及角色](#) · [劇情概要](#) · [制作](#) · [播出](#)

House of Cards (TV Series 2013–2018) - IMDb

<https://www.imdb.com/title/> [翻譯這個網頁](#)

★★★★★ 評分：8.8/10 - 439,703 票

Created by Beau Willimon. With Kevin Spacey, Michel Gill, Robin Wright, Kate Mara. A Congressman works with his equally conniving wife to exact revenge on ...

影片

House of Cards: Season 6 | Official

House of Cards | The Final Season | Netflix

House of Cards | Teaser [HD] | Netflix

紙牌屋 (House of Cards)

連續劇

92% 的使用者喜歡這個電視節目

Google 使用者

《紙牌屋》是美國一部政治權謀題材的電視連續劇，於網路上播出，是Netflix原創劇集。本劇由鮑爾·威利蒙創作并改編自安德魯·戴維斯的BBC同名電視影集。兩部電視影集都是基於麥可·多布斯同名小說創作的。第一季全部13集於2013年2月1日在串流媒體網站Netflix首播，第二季的13集於2014年2月14日同時放出。 [維基百科](#)

開播時間：2013 年 2 月 1 日

劇終時間：2018 年 11 月 2 日

季数：6

編劇：波·威利蒙，麥可·多布斯，安德魯·戴維斯，法蘭克·普格列斯，[更多](#)

獲獎紀錄：[全球獎最佳戲劇類電視女主角](#)，[更多](#)

播映服務供應商

Netflix

訂閱

演員名單

[查看另外 15+ 項](#)

85

知识图谱应用：自动问答系统



Q & A

