

石油勘探开发领域本体构建及应用^①

周树理, 严建文, 包红林, 项亮

(中国石化石油物探技术研究院, 南京 211103)

摘要: 领域本体是知识工程中的重要组成部分, 本体技术是一种重要的知识表示方法, 是知识工程背后的支撑技术. 首先介绍领域本体的构建理论及方法, 从主题域、知识域、本体域提出石油勘探开发领域本体的三维知识模型并基于该模型开发了本体构建软件实现了石油勘探实例的管理及应用.

关键词: 领域本体; 知识工程; 石油勘探开发; 三维模型; 本体构建

Research and Application of Petroleum Exploration and Development Domain-Specific Ontology Construction

ZHOU Shu-Li, YAN Jian-Wen, BAO Hong-Lin, XIANG-Liang

(Sinopec Geophysical Research Institute, Nanjing 211103, China)

Abstract: The domain-specific ontology is an important part of Knowledge engineering. Domain-specific ontology is an important technical knowledge representation, and support technology. This paper introduces the theory and method of constructing Petroleum exploration and development Domain-specific ontology, proposed a three-dimensional model of Petroleum exploration and development from the subject domain, knowledge domain, ontology domain. Based on the model developed ontology construction software to achieve management and application in Petroleum exploration and development domain.

Key words: domain-specific ontology; knowledge engineering; petroleum exploration and development; 3D model; ontology construction

本体作为一种知识共享模式, 为特定领域的人和系统交流提供了极大的便利, 也正因为如此, 本体的研究和应用迅速延伸到知识工程、自然语言处理、信息检索系统、智能信息集成和知识管理、信息交换和软件工程等领域. 而如何对这些不同领域内的知识进行抽取和描述并构建出合适的领域本体已经成为当前的研究热点之一^[1,2].

所谓领域本体是对学科概念的一种描述, 包括学科中的概念、概念的属性、概念间的关系以及属性和关系的约束. 由于知识具有显著的领域特性, 所以领域本体能够更为合理而有效的进行知识的表示. 领域本体一般用来表示某一特定领域范围内的特定知识, 本文主要研究的是基于石油勘探开发领域本体库的构建方法、模型及软件应用.

1 本体构建

1.1 本体的结构

数本体的结构(ontology structure)是一个五元组 $O = \{C, R, Hc, Rel, Ao\}$. 作为不同主体之间进行交流的一种语义基础, 本体由描述某种现实情况的特定术语集, 加上一组关于术语内涵意义的显式假定集合所构成. 在最简单的情况下, 本体只描述概念的分类层次结构; 在复杂的情况下, 本体可以在概念分类层次的基础上, 加入一组合适的关系、公理、规则来表示概念之间的其它关系, 约束概念的内涵解释. 一个完整的本体应由概念(C)、关系(R)、函数(Hc)、公理(Rel)和实例(Ao)五类基本元素构成^[3-6]:

(1)C 这里指的是广义上的概念, 除了可以是一般意义上的概念以外, 也可以是任务、功能行为、策略、

^① 收稿时间:2014-09-12;收到修改稿时间:2014-10-16

推理过程等. 本体中的这些概念通常构成一个继承的分类层次.

(2)R 是领域中概念之间的关系, 形式上定义为 n 维笛卡儿积的子集: $R:C1 \times C2 \times \dots \times Cn$. 在语义上关系对应于对象元组的集合.

(3)Hc 表示函数, n 元函数中的第 n 个元素由前面 n-1 个元素唯一确定, 形式化的定义为: $F:C1 \times C2 \times C3 \dots \times Cn-1 \rightarrow Cn$.

(4)Rel 表示公理是领域中在任何条件下都成立的永真式的描述.

(5)Ao 是指实例即概念中的具体例子, 特定领域的所有实例构成概念在该领域中的值域.

1.2 本体的构建方法

本体是对领域中各个概念之间关系的显性描述. 目前, 领域本体的构建方法还处于探索性研究阶段, 对不同问题的不同认识和解决会得出不同的方法论, 但都包含五条规则:

- (1)明确性和客观性:用自然语言对术语给出明确、客观的语义定义.
- (2)完整性:所给出的定义是完整的, 能表达特定术语的含义.
- (3)一致性:知识推理产生的结论与术语本身的含义不会产生矛盾.
- (4)最大单向可扩展性:向本体中添加通用或专用的术语时, 通常不需要修改已有的内容.
- (5)最少约束:对待建模对象应该尽可能少列出限定的约束条件.

常见的本体构建方法:TOVE、骨架法、IDEF5、METHONTOLOGY、KACTUS、SENSUS 和七步法. 本文提出的本体构建研究以七步法为理论根据, 步骤如图 1 所示.

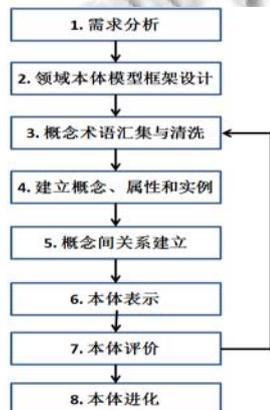


图 1 七步法构建本体步骤

2 石油勘探开发本体框架模型

2.1 知识模型

知识模型中对于知识的描述, 主要可以采用“7W”的模式, 即活动是由何人(Who)何时(When)何处(Where)发起的、发起这个活动的原因(Why)、在这个活动中都遵循了哪些规范、采用了哪些方法(How), 涉及到了哪些(Which)对象、这些对象的特性是什么(What). 通过这个“7W”过程, 可以将某一业务活动中涉及的相关信息完全抽取出来. “7W”的本质是业务对象、业务活动和对象属性这三者彼此之间的关联. 如图 2 所示.

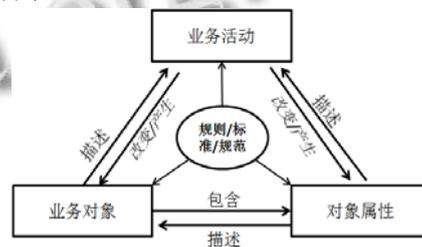


图 2 知识模型

业务活动可以使业务对象发生改变(改变的是对象的属性特征), 或者产生新的业务对象, 对象是业务活动的载体;对象属性是对业务对象的特征描述.

对任何领域而言, 知识模型由业务对象、对象的属性特征、对象涉及的业务活动以及它们之间的关系构成.

2.2 本体的三维模型

基于上述对知识模型的理解, 我们可以将完整的知识模型以主题域、知识域、本体域构成的三维模型来表示. 如图 3.

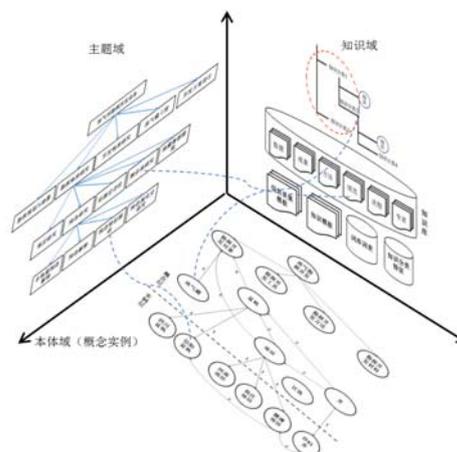


图 3 本体三维模型

在图 3 所示的三维模型中, 无论是主题域、知识域还是本体域, 都是作为知识的一种表现形式而存在, 体现的是石油勘探开发领域内知识的不同方面. 其中, 主题域体现的是业务活动及其之间关系; 知识域体现的是各型各类知识之间的组成和关联关系; 本体域反映的是本领域所有概念(类)、实例及其关联关系. 作为一种知识表达方式, 本体本身即是一种特殊形态的知识.

主题域、知识域、本体域三者构成了一个既在各层面上不断衍生、转化, 又在域间彼此融合、关联的三维空间. 从任何一个域出发, 都与另两个域有错综复杂的关系(图 3 中的域间关联仅作参考).

2.2.1 主题域

主题域体现的是业务活动及其之间关系. 业务活动的划分以方法生命周期为主线, 将专业业务活动与方法管理业务活动有机的串接起来, 尽可能符合油气田勘探、开发管理的约定俗成的习惯, 做到不同业务活动间的业务内容不重复, 并保证能覆盖所有的勘探、开发业务.

根据以上原则和方法, 把油气田勘探开发业务划分为“勘探规划与部署”、“勘探地质研究”、“开发地质研究”、“油气藏工程”、“开发方案设计”等 5 大业务活动类. 如图 4 所示.

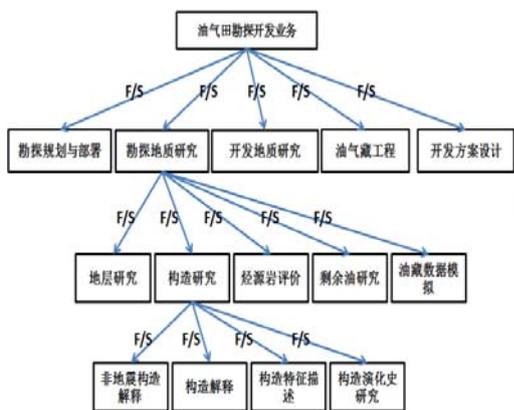


图 4 业务域

2.2.2 知识域

知识域体现为各种知识形态之间的组成和关联关系. 知识域的组成部分包括: 知识库、知识分类、知识模板、词库词表、知识分类特征, 等, 如图 5 所示. 知识库可细分为数据、成果、方法、规范、流程和专家

(专家是知识的重要载体).

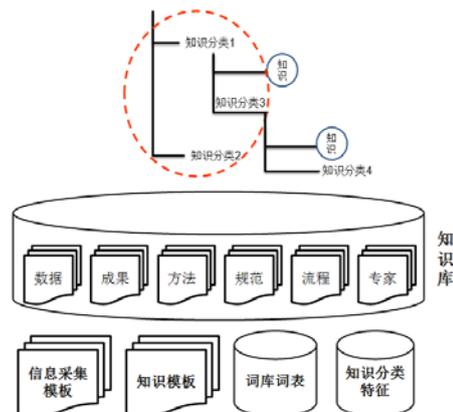


图 5 知识域

无论是包含不同类型知识的知识库、知识分类、词库词表, 还是知识分类特征、知识模板, 都是知识的具体表现形式. 各种形式的知识之间存在着不同类型、错综复杂的关系. 知识之间的关联, 体现了知识的本质特征.

2.2.3 本体域

本体域反映的是本领域所有概念(类)及对象实例. 本体域包含两个层面的内容: 概念层和实例层. 如图 6 所示.

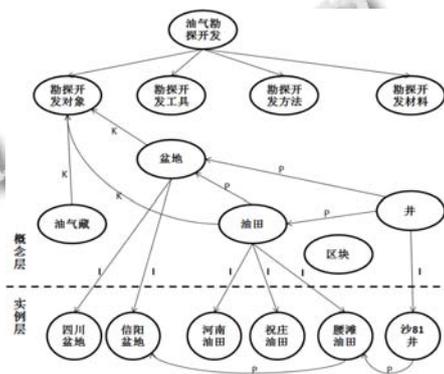


图 6 本体域

2.2.3.1 概念层

概念层体现本领域概念及概念之间的各种关系. 概念层包含了本领域内的通用技术术语, 如石油勘探开发术语、工程技术术语, 等.

概念之间最为典型的关系, 一般有种属关系 (Kind-of)、组成关系 (Part-of) 等. 特别地, 为了使概念层术语及其关系更好地服务于知识搜索和推送, 借鉴

叙词表中的术语关系, 概念关系还包括相关关系、同义关系.

2.2.3.2 实例层

实例层表示业务对象、对象的属性特征以及对象实例之间的关系. 在石油勘探开发领域, 对象实例指的是勘探开发业务活动的直接作用对象, 一般是一些具体的可用位置信息描述的实体, 即所谓“定位类对象”. 如四川盆地、胜利油田、塔里木盆地等.

实例层可以形式化地表示为四元组 $\langle C, R, At, In \rangle$, 其中 C 是本体中对象类概念的集合, 所谓“对象类概念”, 是指可以有直接作用对象实例的概念, 如“盆地”、“油气田”、“井”等, 都属于对象类概念. R 是对象类之间的关系, At 是本体的属性, In 是对象类的实例.

具体的对象实例模型如图 7 所示.

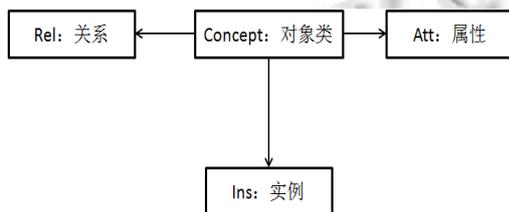


图 7 对象实例模型

石油勘探开发领域典型的对象类概念, 包括: 盆地、油气田、区块、井、油气藏, 等.

所谓属性, 是指对象类所具有的属性特征. 如“井”这一对象类的属性包括井型、井别, 等;

所谓关系, 指对象实例之间的各种可能的关系. 关系既可以存在于不同种类的对象实例之间, 也可以是同种对象的实例之间. 如“井”与“井”之间的相邻关系, 井与油气田之间的所属关系. 井间关系以及井与其他对象类之间的关系, 如图 8 所示.

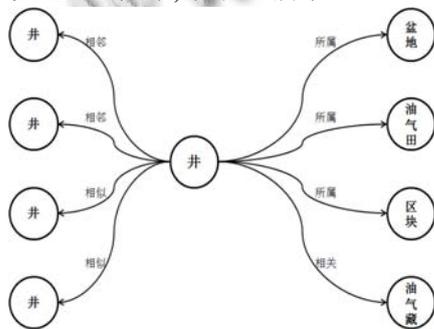


图 8 井的关系图

不同的井实例之间的关系, 通过具体某一属性来建立. 如“相邻”关系, 通过井的“地理位置”来建立, 即, 两口井因为其地理位置相同或相近而具有“相邻”关系; 而井与井“相似”则体现在两口井的井别相同或者井型相同.

3 石油勘探开发本体软件实现及应用

3.1 领域本体软件实现

基于上文所述的领域本体构建方法和模型框架, 开发了本体构建软件. 其主要功能是, 实现对业务主题、知识分类、对象管理. 其中, 对象管理包括对象类和实例的管理, 该功能实现对本体域中实例层(见 2.2.3.2)的管理. 图 9 是对象管理的软件实现界面. 图 10 是对象实例维护界面.

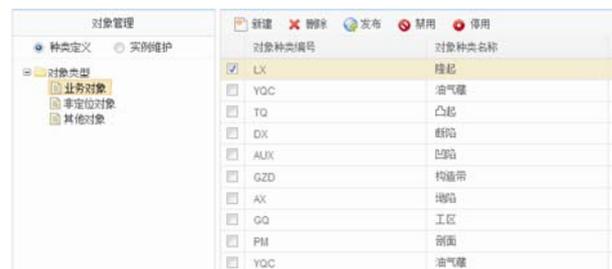


图 9 对象管理界面

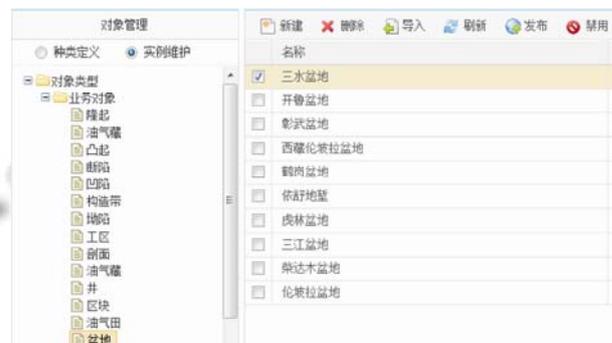


图 10 对象实例维护界面

通过本体构建软件, 本研究初步实现了对对象类概念的定义、对象实例的添加等基本的本体管理功能.

3.2 领域本体的应用

在本体的三维模型(图 3)中, 本体域是由概念和实例组成的空间. 本体域是一种网络结构, 网上每一个结点都是本领域内的一个概念或者实例, 而连接不同结点的线则是概念之间、实例之间或者概念与实例之间的关系. 借助于本体的网络结构, 可以更好地实现

知识检索、关联和推送,对知识共享和知识重用具有重要的推动作用。

石油勘探开发领域本体的应用之一,是推送实例和知识,即应用对象实例之间的关联关系,实现按关系类型的对象实例和知识的推送。

例如,在知识工程系统中,用户打开名为“泌阳凹陷赵凹油田新增石油探明储量报告”的文件,该报告的直接作用对象是“赵凹油田”,则系统借助本体(实例及实例间的关系),自动推送出“所属盆地”、“所属区块”、“相邻油气田”等对象实例。用户可以通过点击这些对象实例链接查找到相关的知识。

4 结语

本文提出石油勘探和开发领域本体的构建方法并通过软件初步实现了对对象类概念及实例的管理及应用。未来将设计实现对该领域本体中概念层的管理,并在此基础上逐步完善。

本体库的构建是一个浩大的人工智能系统工程,领域本体的构建与应用又是其中最具有广阔发展前景的方面,本文提供了在石油勘探开发领域多层次本体构

建方案探讨。本研究所开展的探索和取得的成果,将为石油勘探开发及其他相关领域本体库的构建与应用积累有益的经验。

参考文献

- 1 陈刚,陆汝钤,金芝.基于领域知识重用的虚拟领域本体构造.软件学报,2003,14(3):350-355.
- 2 冯志勇,李文杰,李晓红.本体论工程及其应用.北京:清华大学出版社,2007.
- 3 陈新发,曾颖,李清辉.数字油田建设与实践—新疆油田信息化建设.北京:石油工业出版社,2008.
- 4 杨义忠,王承勇,林淑凤.石油主题词表.北京:石油工业出版社,1994.
- 5 刘宝和.中国石油勘探开发百科全书.北京:石油工业出版社,2008.
- 6 邓小亚.石油领域本体库的构建研究.电子设计工程,2011,19(20).
- 7 袁国铭,李洪奇,樊波.关于知识工程的发展综述.计算技术与自动化,2011,30(1):138-143.