

基于本体的法律知识库的研究与实现^{*})

何庆^{1,2} 汤庸¹ 黄永钊¹

(中山大学计算机科学系 广州 510275)¹ (湖南工程学院 湘潭 411101)²

摘要 基于描述逻辑的本体理论已广泛用于知识表示和推理,而本体在法律领域的研究和应用在国外已开展近 10 年,目前国内还没有这方面的研究与应用成果。本文研究分析了国内外法律知识库的成果,结合中国的法律法规给出了一个法律领域的本体模型 DOLegal,介绍了法律知识库的设计思路及架构。在知识库原型中加入了国内部分法规涉及的行为及处罚,实现了法律知识的表示和推理。文中还对设计中遇到的一些难点进行了研究与分析。

关键词 法律本体,法律知识,知识库,推理规则

The Research and Implementation of Ontology-based Legal Knowledge Base

HE Qing^{1,2} TANG Yong¹ HUANG Yong-Zhao¹

(Department of Computer Science, Sun Yat-Sen University, Guangzhou 510275)¹ (Hunan Institute of Engineering, Xiangtan 411101)²

Abstract The ontology theory based on Description Logics has been widely used in knowledge representation and inference. In foreign countries, researches and application of ontology in legal domain have been carrying out for nearly 10 years. However, there's no domestic research or applicable production at present. This paper investigates and analyzes legal knowledge-base productions at home and abroad, then puts forward an ontology model 'DOLegal' about Chinese law, and introduces the designing and architecture of the legal knowledge base. Integrating actions and punishment involved in some domestic statutes, the knowledge base prototype realizes legal knowledge representation and inference. Some difficulties encountered in design are studied and analyzed in this paper.

Keywords Legal ontology, Legal knowledge, Knowledge base, Inference rule

1 引言

随着社会法制化进程的深入,由国家或地方政府主管部门颁布的各种法规也越来越多,分别适用于不同的行业和领域,但面对众多的法律法规,即使是法律方面的专家也难以自如地掌握和运用。目前国内已有的一些法律信息支持系统^[1],通过关键字把大量的法律法规、法学论文及案例判决书连接在一起,这在一定程度上提高了司法处理的效率,由于这些信息系统都是基于字符串的检索,提供给用户的也是大量文本信息的堆积,计算机不能理解它所存储及处理的文本信息,也就不能在更高层次上向用户提供支持。

本体能在语义和知识层次上表示专业领域的概念及概念间的相互关系,目前已在许多领域得到了广泛的应用,如知识工程、数字图书馆、软件复用、信息检索和 Web 上异构信息的处理、语义 Web 等^[2]。在国外基于本体的法律知识表示及推理这方面的理论研究与应用已开展了一段时间,目前已有的一些系统投入实际应用^[3,4]。本文分析了国内外法律知识库的成果,结合中国的法律法规的特点构建了法律知识库系统。把法律领域的知识加以整理,系统化、形式化存入计算机,有利于法律知识的保存与共享,通过推理机对已有知识的推理可以得到新的知识,大大提高了法律知识的使用效率和准确程度。本系统还提供了法律知识的查询,有利于全民法制观念及法律素质的提高。

2 相关研究

国外理论方面的研究主要以法律领域本体的构建为主,还有一些关于法律知识的提取、推理等方面的研究。

Valente^[5]提出了 FOLaw (Functional Ontology for Law) 法律本体模型,从法律在社会中担任的角色和功能出发描述了法律知识。在 CLIME^[3]项目中用 FOLaw 为模型来构建一个法律信息的服务器,其中包括了 15000 多条法律条款,3500 多个概念,主要用于海事知识及法规的查询和问答。Breuker^[6]提出了另一个法律本体模型 LRI-Core,已在 e-COURT^[4]、E-POWER 等项目中使用。

Engers^[7]对如何自动从法律原文中提取法律知识和规范进行了研究。Chorley^[8]通过基于案例的推理系统自动产生理论以用于对司法案例进行解释。Floris^[9]研究了司法过程中的证据推理。

在国内关于法律知识库的研究和文献很少,只有陈淑燕^[10]描述了一个基于产生式系统的知识库框架,这种形式的知识表示不具备明确的语义,通用性也较差。

3 法律领域本体模型

要表示法律知识,首先就要构建法律领域的本体模型,领域本体应具备以下的角色和功能^[11]。1) 组织和构建领域信息;2) 能推理和解决领域问题;3) 能进行语义检索和查询;

^{*})基金项目:国家自然科学基金项目(60373081);广东省自然科学基金项目(04105503,5003348);教育部“新世纪优秀人才支持计划”资助项目。何庆 博士研究生,主要研究方向为时态知识表示与推理,知识库;汤庸 博士,教授,博士生导师,主要研究方向为数据库,知识工程, CSCW;黄永钊 硕士研究生,主要方向为知识库。

4) 能整合和交换领域信息;5) 能帮助用户理解领域中的知识。

3.1 FOLaw 法律本体

FOLaw 是 Valente 从法律在社会中扮演的角色及功能角度出发而提出来的,模型分为 Normative Knowledge、Meta-legal Knowledge、World Knowledge、Responsibility Knowledge、Reactive Knowledge、Creative knowledge 六大类知识,它不是一个‘纯’的本体,更接近于一个认知框架^[12]。

FOLaw 缺乏组成法律的抽象及核心概念,它是从司法的社会性而不是司法本身出发来描述法律知识的,其中的 World Knowledge 难以全面规范地描述世界的各种常识。因此 FOLaw 不适合描述法律知识及关联的世界知识,但比较适合用于推理和解决法律领域问题。

3.2 LRI-Core 法律本体

Breuker 创建了另一个法律本体模型 LRI-Core,将法律知识划分到世界知识之下,第 2 层概念包括 Mental-Entity、Physical-Entity、Abstract-Entity、Occurrence、Role 五大类。

LRI-Core 虽然将法律知识与世界知识结合在一起表示,但只是一个很笼统的框架,实现法律知识的细节描述和推理时有一定的困难和不足。

3.3 本系统提出的法律本体模型 DOLegal

法律领域的顶层本体应该包含该领域的全面概念和实际应用,结合中国目前的各种法规以及前面两种法律本体的特点,提出以下的法律顶层本体模型 DOLegal,图 1 中描述了 DOLegal 的层次结构,具体介绍如下。

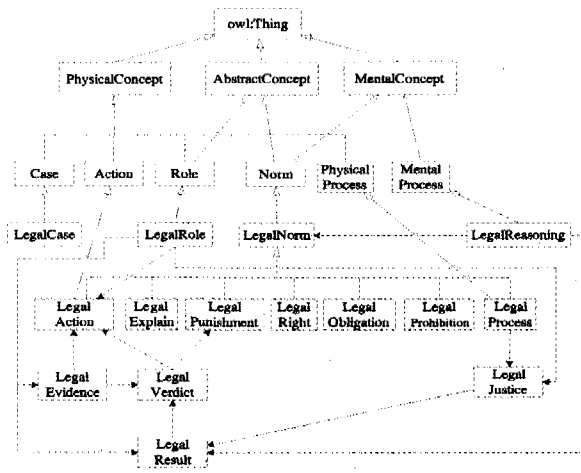


图 1 DOLegal: 法律领域本体模型

LegalNorm: 法律规范,由各种法规中提取出来的法律规范,包括以下 7 个部分:

- 1) LegalAction: 法规中的处理行为及适用条款
- 2) LegalExplain: 法规中的各种解释
- 3) LegalPunishment: 法规中的处罚标准
- 4) LegalRight: 法律中规定的权利
- 5) LegalObligation: 法律中规定的义务及职责
- 6) LegalProcess: 法律事务执行时的各种流程
- 7) LegalProhibition: 法律中禁止的事项

LegalRole: 法律角色,包括违法者、受害者、证人、审判机构、执行机构等

LegalReasoning: 法律知识的推理,分析问题,从 LegalNorm、LegalEvidence 中得出 LegalResult

LegalJustice: 司法审判

LegalResult: 法律结果,包括推理结果及司法审判结果

LegalVerdict: 法律判决

LegalCase: 法律案例,与 LegalRole、LegalAction、LegalVerdict、LegalEvidence、LegalVerdict 关联

LegalEvidence: 法律证据概念

4 法律知识库系统结构

我们给出法律知识库的系统结构,如图 2 所示。

法律知识库元知识:描述各知识库及规则的属性及适用范围。

推理机构:接收应用系统请求,选择适当的知识库及规则完成推理,将结果返回给应用系统,本系统使用的推理机为 Jena,可以处理 OWL 规则及自定义规则。

法律知识库:用一个大而全的知识库表示所有的法律知识是不切实际的,推理效率也会非常低。可以按不同法规以及法律本体 LegalNorm 下的七个组成部分来表示法律知识,各知识库应以前面提到的法律领域本体为基本框架,可以扩充各自的细节内容,本系统中用 Protégé 工具来构建知识库。知识库可以以 OWL 文件形式存储,也可以以推理机认可的格式存到关系数据库中。

规则库:推理机依据规则推出新的知识,可以根据不同的应用需求来创建规则,Jena 可以读取并执行符合其语法的文本形式的规则,规则可以存储在文件中,也可以存储在数据库中。

常规关系数据库:知识库主要用于概念及知识表达,不能在知识库中存放大量实例数据,法规原文、案例及人员等信息必须存储到常规的关系数据库中。

知识库、规则库维护:用户可以通过本体编辑工具构建或维护知识库,还可以将多个知识库合并在一起;规则维护应提供给用户维护规则的接口,同时应提供不同推理机间的规则转换接口。知识库、规则库的各项属性应记入元知识中。

知识、规则提取:从法规原文中自动整理提取法律知识及规则存放到对应的知识库和规则库中,这涉及到 NLP 等复杂的处理过程,目前还难以实现,只能以手工提取为主,国外已有相关的研究及尝试。

案例学习:我国法规中存在很多“情节严重”、“数额较大”、“从重处罚”等非定性描述,通过对已有案例学习,可以为违法行为的定性及惩处尺度提供参考依据。

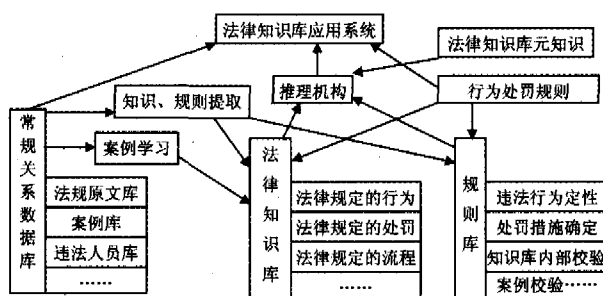


图 2 法律知识库系统架构

5 系统实现

法律知识库加入了《治安管理处罚法》的惩处措施、依据及惩处尺度及部分违法行为描述,为了方便用户比较和使用,同时加入了《刑法》的惩处措施、部分惩处尺度及部分违法行

为描述。

系统分为案例处理、已有案例校验、法律知识查询、知识库维护、知识库校验、规则维护、系统维护七个功能模块。用户可以依据知识库已有概念输入违法行为描述,系统可以确定违法行为涉及的法律条款,根据情节及处理力度给出适当的惩处措施及尺度建议。对于已有案例系统可以判定其行为定性是否准确以及最终惩处是否在恰当以及在合理的惩处尺度之内。用户还可以查询知识库中的各种法律概念,系统会给出相应的法律知识提示。系统在知识库中描述了违法行为定性知识及违法行为处罚知识,具体实现时根据相应的规则进行违法行为定性、给出行为对应的处罚措施及尺度以及校验最终的处罚等。案例的推理过程如图3所示。

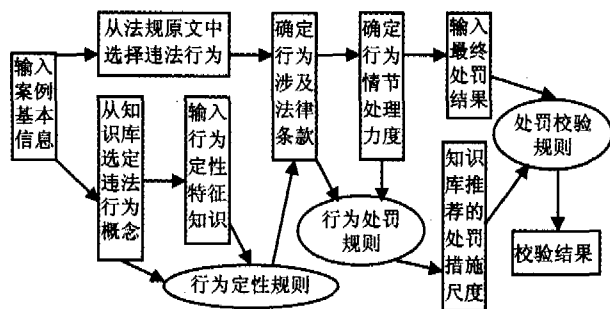


图3 司法案例的定性、处罚及校验过程

这里给出一个具体的规则实例,“非法持有毒品”行为定性规则在Jena中表示如下:

```
[rule1;
  (? action rdf:type ? actionType),
  (? defAction rdf:type ? actionType),
  notEqual(? action, ? defAction),
  (? action base:使用重量单位 ? unit),
  (? defAction base:使用重量单位 ? unit),
  (? action base:持有毒品名称 ? drug),
  (? defAction base:持有毒品名称 ? drug)
  (? defAction base:持有数量 ? quantity),
  (? action base:下限 ? lo),
  (? action base:上限 ? up),
  le(? lo, ? quantity),
  lessThan(? quantity, ? up),
  (? action base:涉及法规条款 ? item)
  →(? defAction base:涉及法规条款 ? item)
]
```

6 下一步的研究计划

目前只是对基于本体的法律知识库进行了初步的研究,开发的应用系统也只是一个原型,还有很多值得进一步研究和深入的方面,具体如下:

进一步完善法律本体模型 DOLegal:国外的研究机构都是与法律研究机构合作,并获得相关行政部门的资助,研究成果也能应用于实践中。本系统只是一个原型系统,参与的人员也没有法律领域的专家,虽然国内尚无此方面的研究,系统的研究及应用前景很好,但需要构建完善的法律本体模型才能成为该领域的标准规范,因此需要与法律领域的专家合作,同时了解具体行政部门的需求,这样才能实现切实可行的法律知识库应用系统。

知识库的存储、检索及推理效率:法律知识库是基于OWL的,其存储形式可以是OWL文本文件,也可以转换成特定的关系数据库格式,其存储及检索均值得研究,目前所用的推理机Jena是把整个知识库读入内存后在进行各种推理的,知识库的组织对推理效率的影响有待更深入的研究,可以关注国内外的最新研究现状及成果以提高知识库的存储、检索及推理效率。此外本体的合并及概念的冲突检测及处理也是值得关注的研究方向。

规则的形式化表示:Jena可以读取并执行符合其语法的文本形式的规则,规则可以存储在文件中,也可以存储在数据库中。目前规则标记语言的标准,但有多个规则标记语言的提案,如SWRL、WRL等。可以研究规则的形式化描述,以合理的形式存储到OWL文件或数据库中,还可以研究不同推理机间的规则转换接口。

其他方面的研究:考虑Agent在法律知识库中的应用,进一步考虑LegalReasoning的表示及实现,考虑法律知识库与常识库的结合。

结论 法律知识库可以大大提高法律知识的使用效率和准确程度,有利于全民法制观念的普及和法律素质的提高,有较高的应用前景和推广价值。本文对国内外法律知识库的研究和应用进行了总结和比较,提出了一个法律领域的本体模型DOLegal,基于此模型开发了一个法律知识库原型系统,系统的细节还需进一步的研究,也需要在实际应用中得到更好的完善。

参考文献

- 1 <http://www1.chinalawinfo.com/index.asp>. 北大法宝. 北大法律信息网
- 2 李善平,尹奇峰,胡玉杰,郭鸣,付相君. 本体论研究综述. 计算机研究与发展, 2004(7):1041~1052
- 3 CLIME (legal information retrieval) IST 25414, 1998-2001, www.bmtch.co.uk/clime/index.html
- 4 e-COURT (information management for criminal court sessions) IST-28199, 2000-2003. www.intrasoft-intl.com/e-court
- 5 Valente A, Breuker J. A functional ontology of law. In: G. Bargellini and S. Binazzi, eds. Towards a global expert system in law. CEDAM Publishers, Padua, Italy 1994
- 6 Breuker J. Constructing a legal core ontology: LRI-Core. <http://www.ws.onto.ufal.br/Papers/brazil-2004-ontology-ws.pdf>
- 7 Engers T M, Ron G, Sayah K. A Case Study on Automated Norm Extraction. In: T. Gordon, ed. Legal Knowledge and Information Systems, Jurix 2004; The Seventeenth Annual Conference. Amsterdam: IOS Press, 2004. 49~58
- 8 Chorley A, Bench T. AGATHA: Automation of the Construction of Theories in Case Law Domains. In: T. Gordon, ed. Legal Knowledge and Information Systems, Jurix 2004; The Seventeenth Annual Conference. Amsterdam: IOS Press, 2004. 89~98
- 9 Floris B, Prakken H. Reinterpreting arguments in dialogue: an application to evidential reasoning. In: T. Gordon, ed. Legal Knowledge and Information Systems, Jurix 2004; The Seventeenth Annual Conference. Amsterdam: IOS Press, 2004. 119~129
- 10 陈淑燕,瞿高峰. 通用法规知识库系统的设计. 计算机工程, 2001, 27(11):90~92
- 11 Valente A. Types and roles of legal ontologies. In: V.R. Benjamins, et al. eds. Law and the semantic web, LNAI 3369, 2005. 65~76
- 12 Breuker J, Hoekstra R. Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law. In: Proceedings of EKAW Workshop on Core ontologies, CEUR, 2004