

УДК 004.822

КАРАСЮК В. В.

ОНТОЛОГИЯ ПРАВОВОГО ЗНАНИЯ: ИСПОЛЬЗОВАНИЕ ТАКСОНОМИЧЕСКОЙ СТРУКТУРЫ

Введение

Правовая информация – это совокупность документированных или публично распространенных сведений о праве, его системе, источниках, реализации, юридических фактах, правоотношениях, правопорядке, правонарушениях, борьбе с ними и их профилактике и т.д. Это определение приведено в соответствии с законом «Об информации» [1]. Из него следует, что в эту совокупность включается любая информация, которая способствует решению правовой задачи.

Однако эта совокупность не является однородной. Накопленные за время существования правовой системы информационные ресурсы представлены в различных распределенных хранилищах, по-разному структурированы и описаны, представлены на разных языках, используют разностороннюю терминологию. А что касается доступных электронных ресурсов, то они имеют к тому же разнообразные форматы.

Поэтому актуальной проблемой становится разработка механизмов и средств для объединения различных информационных ресурсов в единую структурированную систему, навигация по которой будет доступной рядовому пользователю. Это особенно важно с позиций совершенствования систем обучения, когда значительная часть ресурсов переносится в электронное пространство, а пользователи часто не очень квалифицированы.

Одним из современных способов интеграции знаний является использование онтологического описания. Онтология – это структурная спецификация некоторой предметной области, ее формализованное представление, которое включает словарь (или имена) указателей на термины предметной области и логические выражения, описывающие, как они соотносятся между собой. Подавляющее большинство онтологий представляют собой иерархические структуры с понятиями (концептами) в узлах. Основной проблемой онтологического инжиниринга является сложность создания онтологии.

Поэтому в основу наполнения базы знаний положены принципы коллективной работы над онтологией. В их основе лежит свойство самоорганизации информации. Понятие самоорганизующихся систем имеет важное значение для искусственно созданных структур, потому что их прототипы в физической жизни явно демонстрируют самоорганизацию.

Таким образом, задача состоит в создании единой структуры правового знания, разработке методики наполнения этой структуры и организации доступа к созданной структуре, поиска необходимой информации в ней, в первую очередь для целей обучения.

Правовая информация

Любой правовой документ имеет свои атрибуты (метаданные), которые представляют его сферу действия; предмет регулирования; юридическую силу; название правового акта; вид документа; наименование органа, принявшего акт; время его вступления в законную силу и прекращения действия; дата регистрации в Министерстве юстиции; статус документа и так далее. Нормативно-правовой акт имеет четко установленную структуру, четко предписываемые правила поведения субъектов права, в нем выделены ответственность, полномочия, права и обязанности субъектов.

Особенностями языка права являются четкость, определенность, стереотипность, единообразие, его сухость, доступность для понимания. В нормативном высказывании всегда наличествуют нормативные термины: "запрещено", "обязан", "имеет право" и т. п. Правовая информация является основой функционирования правовых систем.

С точки зрения представления правовой информации в информационных системах правовая информация имеет следующие особенности:

- большой объем, например, действующая база объединяет около 500 000 документов;

- в основном представлена в виде неструктурированной текстовой информации;
- правовая информация достаточно синонимична;
- эта информация должна быть достоверной;
- должна быть своевременной – вовремя поступать субъекту деятельности;
- имеет временные ограничения своей легальности;
- расположена в отдельных базах данных, хранилищах, локальных сетях, для каждого ее вида зачастую применяется свой формат представления;
- имеет высокую содержательную изменчивость.

Некоторые исследователи отмечают [2] самоорганизацию правовой информации. Суть самоорганизации поясним так: например, в онтологии могут быть ветви верхнего уровня: уголовное право, гражданское право, административное право и другие. Если какой-либо пользователь ошибочно включил концепт из области уголовного права в ветвь гражданского права, то в процессе эксплуатации системы найдется пользователь, который отредактирует положение указанного концепта и включит его в ветвь уголовного права. Эти синергетические принципы в информационных системах возможно использовать за счет краудсорсинга. Математической модели данного процесса для подобных социальных систем не существует.

Онтологическая структура правового знания

Онтологические структуры (онтологии) широко используемые для описания ресурсов в семантических web-приложениях, корпоративных базах данных, системах обработки документов и т.д. [3]. В работе [4] приводится перечень особенностей, которые характерны для онтологических ресурсов. Перечислим наиболее значимые: формальность; наличие отношений класс-экземпляр; свойства концептов могут иметь информацию о свойствах (слотах); ограничения на область значений свойств; дизъюнктивные классы заведомо не имеют общих экземпляров; наличие аксиом, позволяющих создавать произвольные логические ограничения.

Эти особенности делают онтологии наиболее привлекательными в правовой проблемной области. В настоящее время существует ряд практических программных систем, которые доказали свою эффективность в области права. Обзоры характеристик некоторых систем и методологий их построения можно найти в [5]. Интересными системами являются FOLaw (Functional Ontology of Law) [6], LRI Core [7], Frame-based Ontology, CLO (Core Legal Ontology), Jurwordnet [8]. Следует также упомянуть системы юридического вывода на основе прецедента (legal case-based reasoning - LCBR) [9]. Платформа Ontomedia работает с Core Mediation Ontology и онтологический субдомен используется для посредничества с отдельными субдоменами (например, проблемная область потребителя, проблемная область модели семьи, проблемная область здоровья и т.п.).

Таким образом, целью данной работы является описание методики электронного представления структуры правовых знаний на основе онтологий, возможностей развития средств наполнения этой структуры актуальными знаниями за счет использования метода краудсорсинга и оценка качества выполняемого наполнения.

Тезаурус: объединение возможностей

При планировании работы над созданием онтологии вначале планируется построение глоссария терминов, который включает термины (концепты и их экземпляры, атрибуты, действия и т.п.), которые составляют предметную область, и их естественно-языковые описания. При увеличении объема глоссария до «существенных» размеров, строятся деревья классификации концептов [10].

Эта структура в дальнейшем может быть преобразована в таксономию путем включения, в виде аксиом, знаний, которые не содержатся в словаре (рис.1). Для этого следует провести лексико-семантический анализ глоссария. Это может быть сделано с применением методов обработки натурального языка (НЛП – естественно-языковой процессинг).

Согласно определению международной организации по стандартизации (ИСО), тезаурус является словарем, управляемым языком индексации, формально организованным для того, чтобы установить явные априорные отношения между понятиями. Это определение устанавливает лексические единицы и семантические отношения между этими единицами в качестве элементов, составляющих тезаурус. Тезаурусные отношения (род-вид, часть-целое,

комплекс-элемент, причина-следствие) [11] налагаются на структуру таксономии, то есть идентифицируются основные таксономии предметной области.

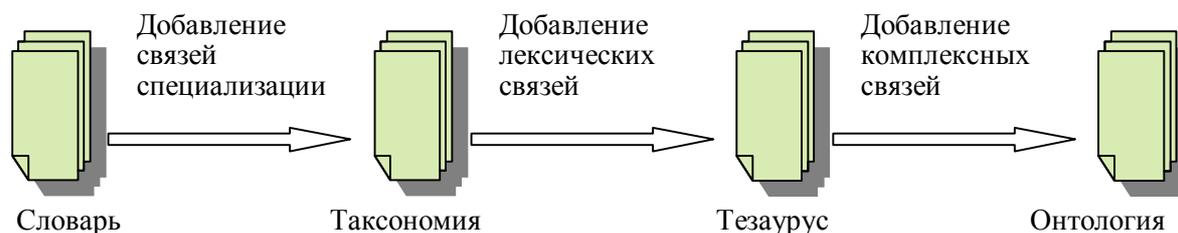


Рис. 1 Последовательность построения онтологии

В этом процессе таксономия понимается в качестве основного инструмента понятий категоризации, которая имеет формальную и машинно обрабатываемую структуру, содержащую термины и их определения в формате XML. Таксономия, в отличие от глоссария, уже отражает иерархические отношения между терминами. Дальнейший этап интеграции – тезаурус, который включает в себя другие взаимоотношения между терминами (например, синонимика отношений), и, наконец, последний этап – онтология, которая будет включать в себя иерархию, основные отношения тезауруса, новые и более сложные семантические отношения, и ограничения, выраженные в виде логических аксиом. Для правовой онтологии важным является включение официальных законодательных определений концептов. Для целей обучения онтология должна иметь терминологический характер, поэтому целесообразно использовать процесс поэтапного создания, который предполагает естественную эволюцию глоссария от своего первоначального формата словаря в онтологию (рис. 1).

Краудсорсинг

Отличительным принципом краудсорсинга является возможность разделения работы на небольшие части (модули), которые могут выполнять разные исполнители. В принципе, к работе на условиях краудсорсинга допускаются любые исполнители, изъявившие желание. Однако в большом количестве практических систем идет речь о дополнительных условиях (процедурах) допуска исполнителей к работе с онтологией. В нашем случае к работе привлекаются студенты юридического университета, что уже само по себе является неким квалификационным условием допуска.

Для технической поддержки работы разработан программный пакет в виде четырех подсистем, с использованием современных технологий визуального программирования, на языке Java. В пакете программ, названном JURONT, используется web интерфейс пользователя и автоматизированный режим работы с базой знаний, в том числе автоматизированное наполнение онтологии из текстовых документов. Практическая значимость пакета определяется аппликация пользователя. Задачами аппликации пользователя являются: 1) навигация в онтологии; 2) выбор фрагментов текстов-источников, соответствующих элементам онтологии; 3) просмотр текстов-источников в полном объеме; просмотр текстов в виде текста, разделенного по разделам; в виде маркированного текста и другие.

Пакет установлен в локальной сети юридического университета. Пользователи (студенты) получили задание (на добровольных началах) в процессе изучения дисциплины «Уголовное право Украины» наполнять онтологию концептами, которые полностью охватят изучаемый материал. По оценке экспертов, полнота извлечения понятий из текста в проанализированных ветвях онтологии составила более 90%. То есть результат достаточно хороший. При этом неохваченными в онтологии остались понятия, которые имели сложную формулировку или многозначное определение.

Выводы

Таким образом, в работе проведен анализ процесса построения онтологии правовых знаний. Исследованы особенности правовой информации и зарубежный опыт формализации

правовых знаний. Сформулирован вывод, что наиболее современным механизмом описания правовых знаний является онтология. Определена структура базы данных, которая реализует базу знаний; разработаны программные приложения, которые ориентированы на технологию "клиент-сервер" и обеспечивают построение онтологии на сервере. Для наполнения онтологии актуальными знаниями применяется методика краудсорсинга – коллективной работы над наполнением онтологии. Оценка качества знаний в системе показала неплохой результат в целом, хотя были обнаружены некоторые несоответствия в онтологии.

Перспективные направления исследований. Предполагаемая доработка методики и программного обеспечения должна автоматически решать проблему поиска места в онтологии (существующий узел в структуре), к которому ближе всего относится новое понятие. Эта проблема может быть решена за счет разработки метрики для оценки подобия онтологий. В результате программный комплекс автоматически предложит пользователю наилучшее место возможного расположения нового концепта, который добавляется в структуру знаний. Также перспективной проблемой является доработка программного обеспечения для решения задач: информационно-справочной; информационно-поисковой; упорядочения структуры базы знаний; построения и расширения индивидуальной базы знаний для каждого пользователя.

ЛИТЕРАТУРА

1. Закон України «Про інформацію» N 2657-XII // Відомості Верховної Ради України (ВВР), 1992. – N 48. – ст.650
2. Trujillo B. Self-Organizing Legal Systems: Precedent and Variation in Bankruptcy // Utah Law Review, 2004. - p. 483-562. [Электронный ресурс]. - Режим доступа: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=924673.
3. Тузовский А. Системы управления знаниями (методы и технологии) / А.Ф. Тузовский, С.В. Чириков, В.З. Ямпольский. – Томск: Издательство научно-технической литературы, 2005. – 260 с.
4. Соловьев В. Онтологии и тезаурусы / В.Д. Соловьев, Б.В. Добров, В.В. Иванов, Н.В. Лукашевич – Казань: КГУ, 2006. – 157 с. [Электронный ресурс]. - Режим доступа: <http://window.edu.ru/resource/722/41722>.
5. Nardi J. Foundational Ontologies for Semantic Integration / J. C. Nardi, R. D. Falbo, J. P. Almeida // A Systematic Literature Review: I3E, - IFIP AICT 399. - p. 238–249.
6. Breuker J. Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law // J. Breuker, R. Hoekstra // Proceedings of the EKAW04 Workshop on Core Ontologies in Ontology Engineering, Northamptonshire, UK. - p. 15 – 27. Режим доступа: <http://dare.uva.nl/document/8751>.
7. Gangemi A. Some onto-logical tools to support legal regulatory compliance, with a case study / A. Gangemi, A. Prisco, M. Sagri, G. Steve, D. Tiscornia // Workshop WORM Core, LNCS: Springer Verlag. – 2003. - p. 607–620. Режим доступа: <http://www.loa.istc.cnr.it/Papers/WORM-CORE.pdf>
8. Sagri M. Jur-WorNet / M. Sagri, D. Tiscornia, F. Bertagna // Second International Wordnet Conference - GWC - 2004. Brno: Masaryk University. - p. 305-310.
9. Henderson J. Dynamic arguments in a case law domain / J. Henderson, T. Bench-Capon // ICAIL '01: proceedings of the 8th international conference on artificial intelligence and law. New York: ACM Press. - 2001. – p 60–69. DOI:10.1145/383535.383542
10. Хала А. О построении онтологии для правовой области с применением технологии METHONTOLOGY / Е.А. Хала // Національна академія наук України. Збірник наукових праць ІПМЕ ім. Г.С.Пухова. Вип. 64. - Київ: 2012. - с. 64 - 71.
11. Широков В. Комп'ютерна лексикографія: [текст]/ Широков В.А. – Київ: Науково-видавниче підприємство «Видавництво “Наукова думка” НАН України», 2011. – 351 с.

КАРАСЮК Владимир Васильевич – к.т.н, доцент кафедры Информатики и вычислительной техники Национального юридического университета им. Ярослава Мудрого.

Научные интересы:

– искусственный интеллект, онтологические системы знаний.