

УДК 004.822

КАРАСЮК В.В.

## ОНТОЛОГИЧЕСКАЯ ПАРАДИГМА ОБРАБОТКИ КОНТЕНТА ДЛЯ ЦЕЛЕЙ ОБУЧЕНИЯ

### Введение

Автоматическое (автоматизированное) формирование базы знаний современной системы электронного обучения является когнитивно значимой задачей в контексте создания онтологической структуры предметной области. Извлечение полезной информации из текстов естественного языка требует выделения информационных ресурсов, pertinentных информационным потребностям пользователя и лексико-морфологического анализа. Обычно для этого используют специализированные лингвистические процессоры, которые анализируют текст и строят семантическую сеть, являющуюся структурой базы знаний.

Задача автоматизированного построения онтологических структур, описывающих знания определенной предметной области, является привлекательной и актуальной для многих сфер деятельности и исследований, особенно с динамически изменяющимся содержанием.

Постановка задачи: разработать систему электронного обучения для специальности «правоведение», которая базируется на онтологических принципах описания базы знаний и обеспечивает нормы представления правовой информации.

Цель работы: представить характеристику подхода и особенности вывода онтологических моделей правовой информации для систем электронного образования при коллективной работе над исходной информацией.

### Особенности правовой информации

Правовая информация определена в законе Украины «Про информацию» как отдельный вид информации. С точки зрения содержания правовая информация характеризуется большими объемами; структурными особенностями; сложностью процессов автоматизированной обработки и, соответственно, отсутствием эффективных программных инструментов. Каждый нормативный акт имеет необходимые реквизиты, которые отражают его юридическую силу, предмет регулирования, сферу действия, придают ему официальность. Традиционными достоинствами языка права являются четкость, определенность, стереотипность, единообразие, его сухость, доступность для понимания. С другой стороны, формально правовая информация в основном представлена в виде неструктурированной текстовой информации, достаточно синонимична, имеет временные ограничения своей легальности, к ней предъявляются требования достоверности, своевременности и т.п. [1]. Также правовая (законодательная) информация нуждается в представлении «как в законе», то есть реконструкция содержания правовых норм в виде «искусственных» предложений на основе отдельных понятий не подходит для целей юридической практики и, соответственно, обучения юристов.

### Структура представления знаний

В разрабатываемой системе использован онтологический подход к организации базы знаний в области правовой информации, которая организует семантическую сеть понятий и относящихся к ним описаний. Поддерживая синонимичность определений и законодательное формулирование понятий, в структуру онтологии ввели множество синонимов понятий и множество текстовых описаний, представляющих законодательные определения.

На теоретико-множественном уровне такая онтология представляется в виде:

$$O = \langle P, R, F \rangle, \quad (1)$$

где  $P$  – конечное множество концептов (понятий, терминов) предметной области, которую представляет онтология  $O$ ;  $R$  – конечное множество отношений между концептами (понятиями, терминами) заданной предметной области;  $F$  – конечное множество функций интерпретации (аксиоматизации), заданных на концептах и / или отношениях онтологии  $O$ . Отметим, что

единственным ограничением, которое накладывается на множество  $P$ , является его конечность и непустота.

$$P = \{P_i\}, \quad (2)$$

где  $P_i$  – отдельное понятие (концепт), имеющее собственное семантическое представление, которое связано с множеством конкретных фактов и множеством допустимых синтаксических конструкций. То есть, онтология обеспечивает словарь для представления и обмена знаниями в области права и множество связей, установленных между терминами в этом словаре [2].

Формально понятие  $P_i$  (2) представляется в виде набора словосочетаний  $W_j^i$ , которые представляют собой синонимы понятия  $P_i$  и набора законодательных (точных) определений этого понятия  $Z_j^i$ :

$$P_i = (W_1^i, \dots, W_n^i); \quad (3)$$

$$(W_1^i, \dots, W_n^i)R_j(Z_1^i, \dots, Z_n^i),$$

где  $R_j$  определяет связь между отдельными синонимическими терминами и их точным (законодательным) определением либо пояснением. В общем случае количество элементов набора  $W_j^i$  может не совпадать с количеством элементов в наборе  $Z_j^i$ .

Элементом онтологии  $O$  также является связь  $R_r$  между понятиями или группой понятий:

$$(P_n, \dots, P_m)R_r(P_k, \dots, P_l). \quad (4)$$

На совокупности понятий онтологии (3), как над множествами, можно выполнять операции: рефлексивность; симметричность; транзитивность; линейность. А над онтологиями и их частями можно выполнять операции: объединение; пересечение; вычитание; выборка.

В процессе развития информационного наполнения системы коллектив пользователей стремится к построению идеальной онтологии. Этим термином обозначаем динамическую базу знаний, которая постоянно совершенствуется и развивается коллективом пользователей.

### Принципы работы с онтологией

Система разрабатывается таким образом, что предусматривается работа пользователя (эксперта) с каждым предложением источников информации последовательно. То есть приблизительная последовательность работы будет следующей:

- пользователь читает очередное предложение и осознает его содержание;
- выделяет понятия, которые упоминаются в предложении;
- определяет связи между группами понятий;
- устанавливает дополнительные связи, которые уточняют содержание основных связей или описывают некоторым образом сложные понятия, в том числе связывает их с законодательными определениями.

На этом этапе пользователь выделил знания, которые несет в себе предложение и которые можно добавить в онтологию. Далее идет этап внесения знаний в базу данных.

На рис. 1 показан пример разбора первых предложений первого параграфа учебника по уголовному праву. Здесь «Тексты» это упорядоченный по источникам и деревьям разделов источников набор предложений исходных текстов (собственно источник построения онтологии). «Понятия» – набор термов, которые несут основную смысловую нагрузку элементов текста. Понятия являются основным элементом онтологии. «Связки» – это набор термов, используемых для задания характера связи между понятиями в тексте. «Связи между понятиями» подразумевают два аспекта: создание групп понятий и связей между группами понятий. Связи между понятиями – это второй по значимости элемент онтологии. «Указание мест употребления понятий в текстах» – это раздел, хранящий непосредственную связь конкретных слов в тексте и понятий из базы (используется для возможности перехода от рассмотрения пользователем текста к рассмотрению онтологии во всей взаимосвязанности понятий). «Указания употребления связей в текстах» – средство для ускорения навигации по

текстам, содержащее привязку предложения к конкретной связи. По сути, эти данные обеспечивают возможность рассмотрения источников построения онтологии (т.е. обратный переход от онтологии к тексту).

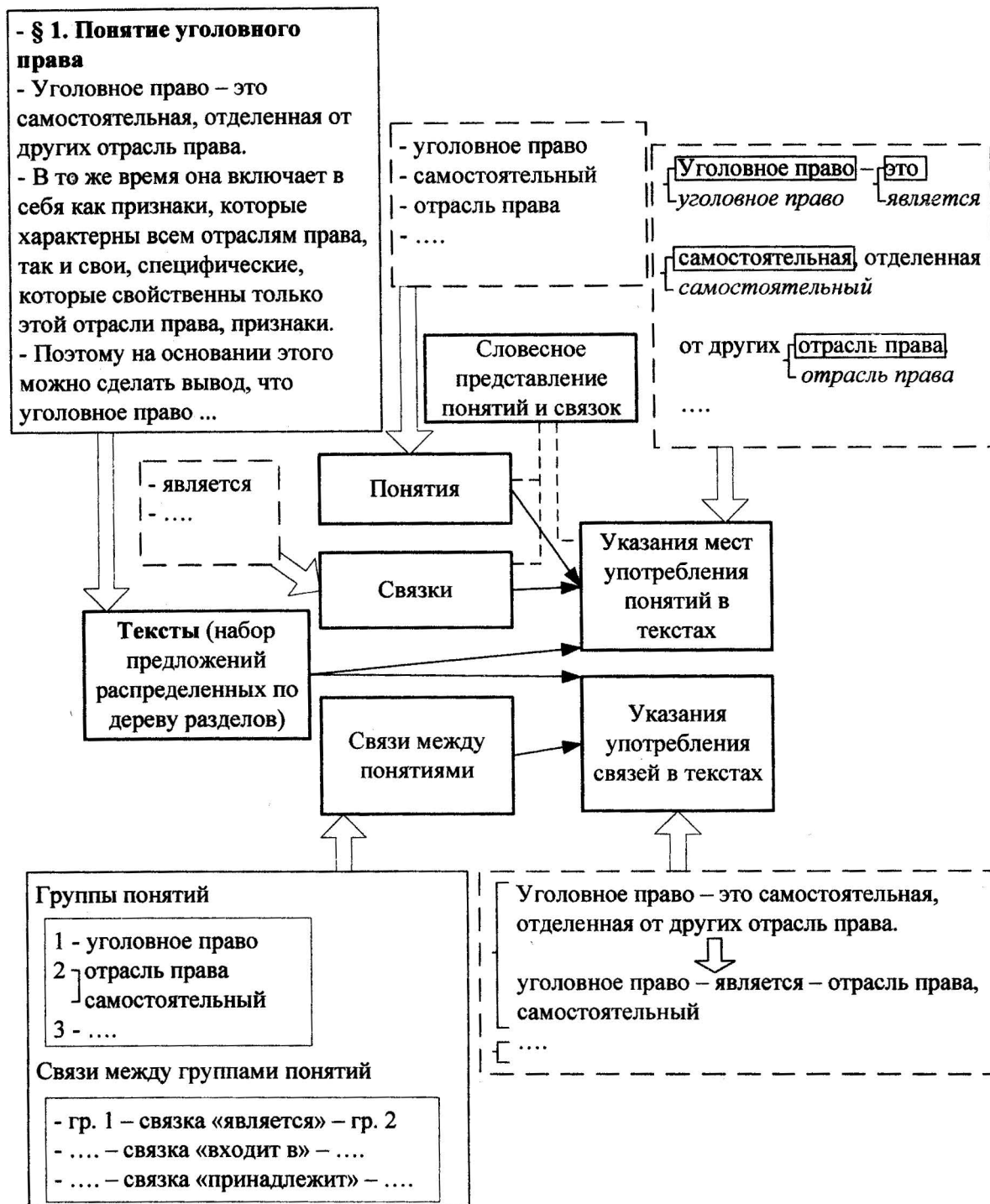


Рис. 1 Пример заполнения базы данных системы электронного обучения информацией из области уголовного права

Два последних компонента базы данных обеспечивают сохранение связей в первом звене цепочки развития системы: исходные тексты – онтология – анализирующие компоненты. Для частичной автоматизации поиска понятий и связок в текстах введена компонента «Словесное представление понятий и связок», в которой содержатся возможные словесные написания понятий и связок, определенные пользователем.

### **Особенности коллективной работы над онтологией**

В технологии наполнения базы знаний системы использованы принципы самоорганизации, в соответствии с которыми каждый пользователь (студент, преподаватель, сторонний пользователь) имеет возможность дополнять и править базу знаний правовой информации. Каждый пользователь может добавить в базу знаний фрагмент отредактированной онтологии согласно своему представлению о структуре изучаемой информации.

Цель данного подхода заключается в объединении усилий многих пользователей для совершенствования онтологии предметной области на основе начальной онтологии и включения в нее понятий и их связей из доступных источников электронной информации, в том числе веб-сайтов. Осуществляется коллективное развитие существующей онтологии. При этом отдельные, потенциально различные онтологии от разных пользователей анализируются и «сливаются» с начальной для создания уточненной, «идеальной» онтологии, которая замещает начальную онтологию предметной области в системе. Для того чтобы сравнить онтологии различных пользователей (представленных в виде графовой модели) и не внести в базу знаний «недостоверную информацию», разрабатывается метрика оценки подобия онтологий. Следует отметить, что проблема возникновения упорядоченности или самоорганизации информации в сложных системах, которые обмениваются с окружающей средой информацией, представляет общенаучный интерес. Это достаточно актуально с точки зрения изучения эволюционных процессов.

### **Выводы**

В результате проведенных исследований сформирована структура базы данных, которая реализует базу знаний; разработаны программные приложения, которые ориентированы на технологию "клиент-сервер" и обеспечивают построение семантической сети на сервере с возможностью многопользовательской работы. Система принята в опытную эксплуатацию в Центре информационных технологий Национального университета «Юридическая академия Украины им. Я.Мудрого».

Перспективные исследования предполагается выполнить в направлении представления нечетких связей между понятиями в базе знаний. Также предусматривается исследование влияния принципов самоорганизации на качество создаваемой множеством пользователей онтологии в предметной области правоведения.

### **ЛИТЕРАТУРА**

1. Tatsyi, V. Семантическая сеть знаний в правоведении = Semantic network of knowledge in science of law / V. Tatsyi, A. Getman, S. Ivanov, V. Karasiuk, O. Lugoviy, O. Sokolov // Automation, Control, and Information Technology (ACIT 2010): Proceedings of the IASTED International Conference, held June 15 – 18 2010 in Novosibirsk, Russia / The International Association of Science and Technology for Development. – Anaheim, USA, Calgary, Canada, Zurich, Switzerland: ACTA Press 2010. p. 218 – 222.
2. Карасюк, В. Модель базы знаний онтологической системы. / В.В. Карасюк, С.Н. Иванов // Труды научно-технической конференции «Компьютерное моделирование в наукоемких технологиях» (КМНТ-2010). Часть 1. – ХНУ им. В.Н. Каразина: - Харьков, 2010. с. 159 – 162.

**КАРАСЮК Владимир Васильевич** - к.т.н., доцент кафедры информатики и вычислительной техники Национального университета «Юридическая академия Украины имени Ярослава Мудрого».

Научные интересы:

*- искусственный интеллект, системы электронного обучения.*