

ПРИКЛАДНЫЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ ЯЗЫКОВОЙ РЕАЛЬНОСТИ (НА ПРИМЕРЕ УГОЛОВНОГО ПРАВА)

И. Е. Воронина

Воронежский государственный университет

Поступила в редакцию 15.03.2013 г.

Аннотация. Обсуждаются лингвистические аспекты моделирования правовых явлений и процессов. В качестве предметной области рассматривается уголовное право.

Ключевые слова: представление знаний, право, моделирование языковой реальности.

Annotation. In the article discusses the linguistic aspects of modeling of legal phenomena and processes in a domain of a criminal law.

Keywords: knowledge representation, ontology, law the modeling the language reality.

ВВЕДЕНИЕ

Для моделирования и описания правовых явлений и процессов актуальна задача создания систем, основанных на знаниях. Такие системы, включающие в себя элементы экспертной обработки и интеллектуального поиска, могут стать, как минимум, полезным инструментом в правотворческой и правоприменительной деятельности.

Онтологическое моделирование, как метод формализации уголовно-правовых норм, является важным средством достижения совпадения или корреляции словаря конкретного субъекта квалификации преступлений с унифицированным словарём терминов уголовного законодательства. Онтология, охватывающая ключевые понятия и категории всех норм уголовного законодательства, может обеспечить понятийную совместимость и единство информационно-поискового языка различных онтологий в области квалификации преступлений, индексирование в процессе квалификации данных и поиск необходимой информации для оценки действий лиц, совершивших общественно опасные деяния.

Принятие решений в юриспруденции носит субъективный характер, что обусловлено недостаточной степенью формализации материала. В подобных условиях, как следствие, не приходится говорить не только о какой-либо автоматизации принятия решений, но даже о написании обучающих программ, позволяющих, например, проверять правильность решения

учебных юридических задач, не говоря уже о задачах-казусах.

Следует заметить, что одним из главных препятствий на пути формализации является наличие нечетко определенных понятий предметной области. В связи с этим необходимо выяснить, что вообще следует понимать под четко определенным понятием, каковы его признаки, можно ли сформулировать алгоритм определения принадлежности понятия к четко определенным.

Кроме того, выделение ключевых понятий является важным шагом в процессе создания онтологии предметной области. Онтология необходима как для анализа юридических текстов, так и для создания прототипа учебной юридической экспертной системы, пусть даже основывающейся на неполной базе знаний.

Общий список выделенных ключевых понятий составил порядка 700, причем среди них оказались как отдельные слова, так и словосочетания. Список отношений (всего 26), связал ключевые понятия уголовного права.

ПРОБЛЕМА НЕЧЕТКИХ ОПРЕДЕЛЕНИЙ

Одной из причин возникновения неоднозначности при принятии решений в уголовном праве является тот факт, что не все ключевые понятия имеют четкое легальное определение. Для проведения дальнейших рассуждений следует решить, что мы будем считать четко определенным понятием.

Для того чтобы понятие считалось точно определенным, необходимо, чтобы оно удовлетворяло нижеперечисленным признакам:

1. Определение понятия есть в законе или подзаконном акте.

2. Определение понятия содержится только в одном акте.

3. Если определение понятия содержится в нескольких актах, то это либо одно и то же определение, либо они дополняют друг друга.

4. Определения в разных законах не должны противоречить друг другу.

5. Данный закон или подзаконный акт не утратил юридическую силу.

6. Определение понятия не содержит в себе (или содержит незначительное количество, не влияющее на смысл) неопределенных или нечетко определенных терминов.

7. Определение понятия включает в себя исчерпывающий перечень элементов.

8. Определение понятия включает в себя неисчерпывающий перечень элементов, но его можно сделать исчерпывающим, позаимствовав элементы из другого закона или акта.

9. Можно дать однозначную количественную характеристику термина.

10. Термин можно однозначно вычлени из определения, он не будет смешиваться с другим сходным термином.

11. Термин не должен составлять большую часть определения либо быть слишком громоздким.

12. Цепочки взаимоопределяющих дефиниций не должны ссылаться друг на друга.

Обозначим $k_1 - k_{12}$ – значения признаков четко определенного понятия. Тогда для того, чтобы определить, является ли понятие четко определенным, применим следующие правила.

1. Если определение термина есть в законе или подзаконном акте, то $k_1 = 1$, иначе $k_1 = 0$.

2. Если определение содержится только в одном акте, то $k_2 = 1$, иначе $k_2 = 0$.

3. Если $k_2 = 0$ и $k_1 = 0$, то $k_3 = 0$ и $k_4 = 0$.

4. Если $k_2 = 0$ и $k_1 < 0$, то если определение содержится в нескольких актах, и это одно и то же определение, и они дополняют друг друга, то $k_3 = 1$, иначе $k_3 = 0$.

5. Если $k_2 = 1$, то k_3 – не определено.

6. $k_4 = k_3$.

7. Если $k_1 = 0$, то k_5 – не определено, иначе если данный закон или подзаконный акт не утратил силу, то $k_5 = 1$, иначе $k_5 = 0$.

8. Если определение не содержит в себе (или содержит незначительное количество, не влия-

ющее на смысл) неопределенных или нечетко определенных терминов, то $k_6 = 1$, иначе $k_6 = 0$.

9. Если определение включает в себя исчерпывающий перечень элементов, то $k_7 = 1$, иначе $k_7 = 0$.

10. Если $k_7 = 0$, то если определение включает в себя неисчерпывающий перечень элементов, но его можно сделать исчерпывающим, позаимствовав элементы из другого закона или акта, то $k_8 = 1$, то $k_8 = 0$.

11. Если $k_7 = 1$, то k_8 не определено.

12. Если термин не количественный, то k_9 не определено.

13. Если термин количественный и ему можно дать однозначную количественную характеристику, то $k_9 = 1$, иначе $k_9 = 0$.

14. Если термин можно однозначно вычлени из определения и он не будет смешиваться с другим сходным термином, то $k_{10} = 1$, иначе $k_{10} = 0$.

15. Если термин не составляет большую часть определения либо не слишком громоздкий, то $k_{11} = 1$, иначе $k_{11} = 0$.

16. Если цепочки взаимоопределяющих дефиниций не ссылаются друг на друга, то $k_{12} = 1$, иначе $k_{12} = 0$.

17. Если значения для всех $k_i = 1$ (кроме тех, которые неопределены), то определение считается четко определенным, иначе – нечетко определенным.

Вышеуказанные правила позволяют определить значения признаков четко определенного понятия для всех ключевых понятий отрасли. Будем рассматривать только те понятия, среди значений признаков которого встречается хотя бы один 0 (такие понятия и будем считать нечетко определенными), значения признаков, которые не определены, обозначим «*».

Вышеуказанные правила позволяют определить значения признаков четко определенного понятия для всех ключевых понятий отрасли. Будем рассматривать только те понятия, среди значений признаков которого встречается хотя бы один 0 (такие понятия и будем считать нечетко определенными), значения признаков, которые не определены, обозначим «*». Ниже представлен фрагмент таблицы значений признаков для нечетко определенных понятий.

Значения признаков четкости для нечетких понятий

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	..
Близкие	1	1	*	*	*	1	0	0	..
Беспомощное состояние	0	0	*	*	*	1	1	*	..
Хулиганские побуждения	0	0	*	*	*	1	1	0	..

Для вычисления веса нечетко определенных понятий, используя простую формулу:

$$\text{Вес} = \frac{\sum_{i=1}^n k_i}{n}, n=12.$$

Так как нечетко определенные понятия имеют различные веса из диапазона $[0,1]$, то можно их ранжировать. В результате все нечетко определенные понятия можно будет разбить на группы и каждой группе поставить в соответствие значение лингвистической переменной.

Учитывая диапазон весов и количество понятий, имеющих определенный вес (например, вес 0,25 имеет больше половины понятий), можно выделить следующие группы (табл. 2):

Процентное соотношение групп нечетко определенных понятий представлено на рис.

Можно сделать вывод, что более половины всех нечетких понятий (64 %) составляют понятия, имеющие вес 0,25, остальные понятия распределены по группам приблизительно равномерно.

ФОРМАЛИЗОВАННОЕ ПРЕДСТАВЛЕНИЕ НЕЧЕТКОСТИ

Первичные экспертные данные, полученные в ходе исследований, представляют собой совокупность отображений вида $E \rightarrow X_i, i \in 1, n$, где E – множество объектов, X_i – множество значений переменной x_i , i – индекс, нумерующий переменные, участвующие в исследовании.

В нашем случае E – множество нечетко определенных понятий ($e^{(1)} - e^{(92)}$), X – множество признаков четкости, n – количество признаков, $n=12$, $K_1 - K_{12}$ – множества значений перемен-

Таблица 2

Лингвистическая шкала нечеткости

Группа весов	Значение лингвистической переменной
0,08	незначительный
0,17	малозначительный
0,25	небольшой значимости
0,33–0,42	средней значимости
0,5	значительный
0,58	весьма значимый
0,67–0,75	особо значимый

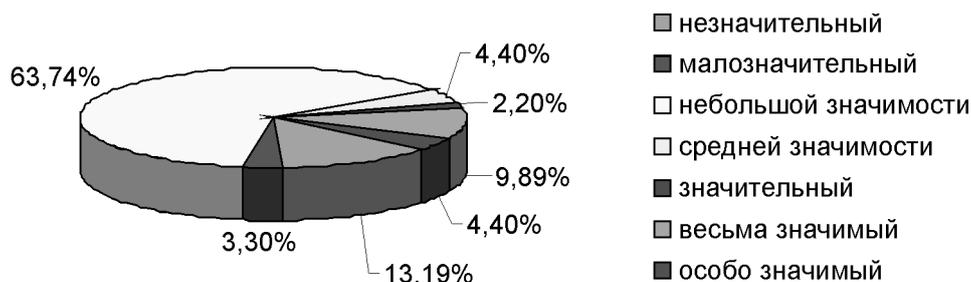


Рис. Процентное соотношение групп нечетко определенных понятий

ных $k_1 - k_{12}$, набор отображений – множество нечетких понятий.

Имеется 12 переменных $k_1 - k_{12}$. Каждая из них имеет множество значений из трех элементов $M = \{m^{(1)}, m^{(2)}, m^{(3)}\}$, где $m^{(1)} - 0$, $m^{(2)} - 1$, $m^{(3)}$ – не определено (обозначим данное значение «*»). E – множество нечетко определенных понятий ($e^{(1)} - e^{(92)}$). Фрагмент матрицы данных представлен в табл. 3.

Пусть значения переменных $k_1 - k_{12}$ будут следующими:

- k_1 : $k_1 - 1, \neg k_1 - 0, k_1^* -$ не определено
- k_2 : $k_2 - 1, \neg k_2 - 0, k_2^* -$ не определено
- k_3 : $k_3 - 1, \neg k_3 - 0, k_3^* -$ не определено
- k_4 : $k_4 - 1, \neg k_4 - 0, k_4^* -$ не определено
- k_5 : $k_5 - 1, \neg k_5 - 0, k_5^* -$ не определено
- k_6 : $k_6 - 1, \neg k_6 - 0, k_6^* -$ не определено
- k_7 : $k_7 - 1, \neg k_7 - 0, k_7^* -$ не определено
- k_8 : $k_8 - 1, \neg k_8 - 0, k_8^* -$ не определено
- k_9 : $k_9 - 1, \neg k_9 - 0, k_9^* -$ не определено
- k_{10} : $k_{10} - 1, \neg k_{10} - 0, k_{10}^* -$ не определено
- k_{11} : $k_{11} - 1, \neg k_{11} - 0, k_{11}^* -$ не определено
- k_{12} : $k_{12} - 1, \neg k_{12} - 0, k_{12}^* -$ не определено

Введем еще одну переменную: y – уровень четкости. Ее значения:

- y_1 – незначительный;
- y_2 – малозначительный;
- y_3 – небольшой значимости;
- y_4 – средней значимости;
- y_5 – значительный;
- y_6 – весьма значимый;
- y_7 – особо значимый.

Чтобы перейти от достаточно детального описания, которое дается совместно переменными, к более агрегированному, обобщенному описанию, необходимо задать функцию

$$Y = f(k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, k_{11}, k_{12}).$$

На основании матрицы данных, представленной в табл. 3, представим логическую фун-

кцию:

$$y_1 = (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& k_7^* \& k_8^* \& \neg k_9 \& \neg k_{10} \& \neg k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& k_4^* \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& \neg k_{10} \& k_{11} \& \neg k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& \neg k_{10} \& \neg k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& \neg k_{10} \& \neg k_{11} \& k_{12});$$

$$y_2 = (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& \neg k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& \neg k_{11} \& k_{12});$$

$$y_3 = (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& k_4^* \& k_5^* \& k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& \neg k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& k_7^* \& k_8^* \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& \neg k_4 \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& k_4^* \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& \neg k_3 \& k_4^* \& k_5^* \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12});$$

$$y_4 = (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& k_6 \& k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& \neg k_{12}) \vee (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& \neg k_6 \& k_7 \& k_8^* \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& k_6 \& k_7 \& k_8^* \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12});$$

$$y_5 = (\neg k_1 \& \neg k_2 \& k_3^* \& k_4^* \& k_5^* \& k_6 \& k_7 \& k_8^* \& k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& k_3^* \& k_4^* \& k_5^* \& k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12});$$

$$y_6 = (k_1 \& \neg k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& \neg k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& \neg k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12});$$

$$y_7 = (k_1 \& k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12});$$

Таблица 3

Фрагмент матрицы данных

	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	k_{11}	k_{12}
$e^{(1)}$ Беспомощное состояние	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(3)}$	$m^{(3)}$	$m^{(2)}$	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(2)}$	$m^{(2)}$	$m^{(2)}$
$e^{(2)}$ Близкие	$m^{(2)}$	$m^{(2)}$	$m^{(3)}$	$m^{(3)}$	$m^{(3)}$	$m^{(2)}$	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(2)}$	$m^{(2)}$	$m^{(2)}$
$e^{(3)}$ Важный личный документ	$m^{(1)}$	$m^{(1)}$	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(1)}$	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(2)}$	$m^{(2)}$	$m^{(2)}$
$e^{(4)}$ Взрывоопасные объекты	$m^{(2)}$	$m^{(1)}$	$m^{(2)}$	$m^{(2)}$	$m^{(2)}$	$m^{(1)}$	$m^{(1)}$	$m^{(1)}$	$m^{(3)}$	$m^{(2)}$	$m^{(2)}$	$m^{(2)}$
...												

$\& \neg k_2 \& k_3 \& k_4 \& k_5 \& \neg k_6 \& \neg k_7 \& \neg k_8 \& \neg k_9 \& k_{10} \& k_{11} \& k_{12}) \vee (k_1 \& k_2 \& k_3 \& k_4 \& k_5 \& k_6 \& \neg k_7 \& \neg k_8 \& k_9^* \& k_{10} \& k_{11} \& k_{12})$.

Функция f позволяет избежать излишней детализации и перейти к содержательно более агрегированному описанию в терминах уровня нечеткости.

Рассчитывая на то, что между объектами предметной области всегда предполагается наличие семантического отношения [1]:

СЕМАНТИЧЕСКОЕ ОТНОШЕНИЕ ($\langle \text{объект1} \rangle$ И $\langle \text{объект2} \rangle$) ИМЕЕТ МЕСТО, или

СЕМАНТИЧЕСКОЕ ОТНОШЕНИЕ ($\langle \text{объект1} \rangle$ И $\langle \text{объект2} \rangle$) ИМЕЕТ МЕСТО С ОПРЕДЕЛЕННОЙ ДОЛЕЙ УВЕРЕННОСТИ

($\text{SemR} (\text{Comp}_1 \longrightarrow \text{Comp}_2)$), или, когда подразумевается степень уверенности x_i , $\text{SemR} (\text{Comp}_1 \xrightarrow{x_i} \text{Comp}_2)$), в данном случае правило $\text{SemR} (\text{Comp}_1 \xrightarrow{x_i} \text{Comp}_2)$ может трансформироваться в **КЛЮЧЕВОЕ ПОНЯТИЕ** ($e^{(n)} \xrightarrow{Y_i} \text{Нечетко определенное}$), то есть в простое определение принадлежности ключевого понятия к нечетким с указанием степени нечеткости). Это представление может лечь в основу исследования влияния нечетких понятий на принятие решения, что в настоящее время было проиллюстрировано на примере учебных задач [2]. Кроме того, подобная трактовка позволяет применять для исследования те же методы, что и в вопросах сочетаемости языковых единиц [1].

Все вышеупомянутые ключевые понятия и отношения были получены вручную и представляют собой образец для сравнения с экспериментальным материалом, который можно получать, опробывая различные алгоритмы. Так, например, анализ данных, полученных в результате программной обработки Общей части Уголовного кодекса РФ, вычисление силы связи между словами по алгоритму Гинзбурга действительно отражает реальные взаимосвязи ключевых понятий данного текста. Алгоритм может использоваться для выявления терминологических словосочетаний, которые принци-

пиально не могут быть получены в рамках статистических подходов без учета контекста [3].

ЗАКЛЮЧЕНИЕ

В ходе выделения ключевых понятий возникли определенные трудности, которые требуют дополнительной проработки и повлекут за собой изменения в алгоритмах:

- неоднозначность в определении, все ли элементы словосочетания составляют одно ключевое понятие, или целесообразнее разбить его на несколько понятий;

- неоднозначность в определении количества элементов в словосочетании, претендующем на роль ключевого понятия;

- большое количество элементов в некоторых ключевых понятиях-словосочетаниях;

- элементы некоторых ключевых понятий стоят не подряд (например, «охрана общественной безопасности, общественного порядка» дает два ключевых понятия – «охрана общественной безопасности», «охрана общественного порядка».

Очевидно и то, что от уровня формализации лингвистической реальности, результатов исследований в этой области напрямую зависит и то, чего можно достичь на пути создания специализированного языка моделирования и описания отрасли уголовное право.

СПИСОК ЛИТЕРАТУРЫ

1. Воронина И.Е. Компьютерное моделирование лингвистических объектов: монография / И. Е. Воронина – Воронеж: Издательско-полиграфический центр Воронежского гос. ун-та, 2007. – 177 с.

2. Воронина И.Е. Проблемы формализации принятия решений в юриспруденции. / И. Е. Воронина, Л. В. Дудкина // Сб. трудов межд. конференции «Актуальные проблемы прикладной математики, информатики и механики. Воронеж, 22–24 июня, 2009» – Воронеж: Изд.-полиграф. центр Воронежского гос. ун-та, 2009. – Ч. 1. – с. 98–100.

3. Воронина И.Е. Функциональный подход к выделению ключевых слов: методика и реализация / И. Е. Воронина, А. А. Кретов, И. В. Попова, Л. В. Дудкина // Вестник Воронеж. гос. ун-та. Серия Системный анализ и информационные технологии. – 2009. – № 1. – С. 68–72.

Voronina Irina Ye. – Associated Professor of Software & Information System Administering Chair, Department of Applied Mathematics, Computer Science & Mechanics, Voronezh State University. E-mail: irina.voronina@gmail.com

Воронина Ирина Евгеньевна – к. т. н., доцент кафедры программного обеспечения и администрирования информационных систем факультета ПММ, ВГУ. E-mail: irina.voronina@gmail.com