

# ПРИМЕНЕНИЕ ЭКСПЕРТНО-КЛАССИФИКАЦИОННЫХ МЕТОДОВ ПРИ ПРОЕКТИРОВАНИИ ПРОФЕССИОНАЛЬНЫХ И ОБРАЗОВАТЕЛЬНЫХ СТАНДАРТОВ

В.В. НИКИТИН

Предложена экспертно-классификационная процедура для совершенствования онтологии объектов профессиональной деятельности при разработке государственных профессиональных и образовательных стандартов, в которой существенное место занимают алгоритмы автоматической классификации.

**Ключевые слова:** онтология предметной области, объекты профессиональной деятельности, алгоритмы классификации.

## ВВЕДЕНИЕ

Важный начальный этап процесса проектирования профессиональных и образовательных стандартов — формирование и анализ множества объектов профессиональной деятельности (ОПД) специалиста, т. е. видов его трудовой деятельности и объектов, на обеспечение функционирования и развития которых эта деятельность направлена. Эффективное выполнение работ, связанных с такой деятельностью, требует от специалиста определенного набора умений и компетенций. Образовательные программы должны быть построены таким образом, чтобы обеспечивать формирование у обучающихся необходимого набора умений и компетенций.

Формирование множества ОПД для определенной предметной области<sup>1</sup> и его анализ представляют собой слабо формализованную и трудоемкую задачу. Сегодня общепризнанным способом описания предметной области служит онтологическое моделирование. В общем случае онтологическая модель, или онтология, задается тремя конечными подмножествами: концептов, связей и функций интерпретации. На практике наблюдается тенденция к использованию упрощенных онтологических моделей, в которых множество ОПД представлено множеством (словарем) терминов, или понятий, имеющим определенную структуру [1]. Структура

может задаваться, например, графом, вершины которого соответствуют понятиям, а дуги — связям между ними. Если этот граф иерархический, вершины более высокого уровня могут интерпретироваться как более сложные (обобщающие) понятия, или классы более элементарных, «исходных» для этого уровня ОПД. Именно классы ОПД служат основой для формирования набора компетенций специалиста и затем уже профессиональных и образовательных стандартов.

Задачу анализа предметной области в целях формирования профессиональных и образовательных стандартов можно понимать как задачу выявления структуры множества ОПД, т. е. задачу построения онтологии предметной области на базе ее «первичного» описания в виде неструктурированного множества. Однако для этого необходимо каким-то образом задать связи между ОПД. Трудность состоит в том, что для задания таких связей необходимо уже иметь (хотя бы в виде модели) некоторую онтологию, т. е. некоторое представление о структуре множества ОПД. Получается замкнутый круг. Поэтому на практике возникает задача не столько создания онтологии «с нуля», сколько улучшения уже существующей онтологии (полученной, например, экспертным или нормативным путем). Цель такого улучшения — сделать структуру множества ОПД более адекватной поставленной задаче формирования образовательных стандартов. Конкретные онтологии представляют интерес только в том случае, если ведется постоянная работа по их совершенствованию.

Однако создание и совершенствование даже относительно простых онтологических моделей

<sup>1</sup> В данном случае предметная область — это область профессиональной деятельности специалиста.



представляют значительные трудности для высокотехнологичных, динамично развивающихся отраслей, для которых характерны частое появление новых объектов и связанных с ними понятий, существование большого числа синонимичных понятий, быстрое устаревание некоторых типов объектов, непрерывные изменения в описании и интерпретации отдельных понятий и др.

Сложность задачи связана не только с наличием быстрых изменений в самой предметной области, но и с большим объемом исходной информации. Число ОПД может исчисляться сотнями, а число классов, в которые их желательно объединить, — десятками. При решении задачи вручную эксперт вынужден резко ограничивать число рассматриваемых вариантов онтологии, основываясь на своих субъективных оценках. Поэтому для повышения объективности и обоснованности экспертизы целесообразно применять формализованные критерии и процедуры.

Задача построения онтологии на базе неструктурированного множества ОПД рассматривалась в работе [2]. В этой постановке эксперты должны были определить наличие и меру связи между всеми парами ОПД, причем каждая связь задавалась (по определенным правилам) некоторым числом. В настоящей работе предполагается, что некоторая структура на множестве ОПД уже задана, и требуется «реструктурировать» это множество, т. е. указать новую структуру, более адекватную задаче формирования образовательных стандартов. В такой постановке задача экспертов существенно упрощается: от них требуются уже не численные оценки, а качественная информация значительно меньшего объема и в более привычных для них качественных терминах.

## 1. ПОСТАНОВКА ЗАДАЧИ

Пусть задана некоторая онтология, т. е. множество понятий, имеющее определенную структуру. Пусть, далее, эта структура иерархическая — если каждому понятию поставить в соответствие вершину некоторого графа, то две вершины соединяются дугой в том и только в том случае, если они находятся на разных уровнях. При этом:

— если вершина более высокого уровня соединена дугами с несколькими вершинами одного из предыдущих уровней, то она интерпретируется как интегрированное (обобщающее) понятие или *класс* вершин (понятий) этих уровней, соединенных с нею дугами;

— граф не обязательно является деревом — каждая вершина более низкого уровня может соединяться дугами с несколькими вершинами более высокого уровня, т. е. входить одновременно в несколько классов;

— граф может содержать изолированные вершины, не включенные ни в один из классов.

Чтобы иметь возможность сравнивать разные онтологии, введем два критерия.

- Критерий качества классификации вершин графа онтологической структуры. Классификация выполняется алгоритмически, но после этого эксперты оценивают полученный результат, исходя уже не из формальных, а из содержательных соображений: «естественности» объединения понятий более низкого уровня в обобщающие понятия, наличия изолированных (не отнесенных ни к одному из классов) ОПД, имеющейся дополнительной информации, не учтенной при классификации и т. п.
- Критерий равномерности распределения весов классов по уровням онтологической структуры. Этот критерий допускает формализацию.

Пусть каждой вершине присвоен некоторый вес. По определению весом класса является сумма весов входящих в него вершин. Поясним содержательный смысл введенного параметра. Если все вершины равноценны, то естественно присвоить им один и тот же вес, например, 1. Тогда вес класса — это число входящих в него вершин и характеризует его размер. Возможные подходы к присвоению весов обсуждаются далее в § 2.

Будем считать, что онтология  $A$  лучше онтологии  $B$  по критерию равномерности, если дисперсия  $\sigma_{cp}^2(A)$  весов классов в среднем для всех уровней структуры  $A$  меньше, чем структуры  $B$ , где  $\sigma_{cp}^2(A)$

определяется выражением  $\sigma_{cp}^2 = \frac{1}{l_B - 1} \sum_{l=2}^{l_B} \sigma_l^2$ . Здесь

$l_B$  — номер самого верхнего уровня онтологической структуры (уровни нумеруются снизу вверх),  $\sigma_l^2$  — дисперсия весов классов  $l$ -го уровня.

Окончательное решение о том, какая из двух онтологий лучше, принимается экспертами на основании их сравнения по двум описанным выше критериям.

Задача заключается в том, чтобы, исходя из заданной онтологии, построить новую онтологию, лучшую (в определенном нами смысле), чем заданная.

Далее описана экспертно-классификационная процедура решения поставленной задачи. На одном из этапов этой процедуры используется алгоритм автоматической классификации. Для простоты изложения и возможности содержательной интерпретации получаемых результатов описание процедуры дается на примере конкретной онтологии.

**2. ОПИСАНИЕ ПРОЦЕДУРЫ**

Опишем процедуру решения задачи на примере онтологии для предметной области «Информатика и информационно-коммуникационные технологии» [1].

Эта онтология имеет трехуровневую иерархическую структуру. На верхнем уровне находятся 6 классов объектов профессиональной деятельности.

1. Компоненты аппаратных средств информационных технологий.

2. Корпоративные информационные системы и услуги, связанные с их проектированием и эксплуатацией.

3. Компоненты программных средств информационных технологий.

4. Компонент ориентированные информационные системы.

5. Коммуникационные средства, системы и услуги, связанные с их проектированием и эксплуатацией.

6. Средства микроэлектронной техники и услуги, связанные с их проектированием и производством.

На втором уровне каждый из этих классов (обобщающих понятий) подразделяется на несколько подклассов (более простых понятий). Рассмотрим подробнее класс 2 «Корпоративные информационные системы и услуги, связанные с их проектированием и эксплуатацией». Он подразделяется на три подкласса.

2.1. Инструментальные программные средства интеграции корпоративных информационных систем.

2.2. Услуги по поддержке жизненного цикла корпоративных информационных систем.

2.3. Функциональные подсистемы корпоративных информационных систем.

На самом нижнем, третьем уровне класс 2 включает в себя 44 ОПД (табл. 1). Существенно, что в данной онтологии (и эта ситуация типичная) объектам нижнего уровня соответствуют изолированные вершины.

Другими словами, указанные 44 ОПД не распределены по подклассам второго уровня: они даются одним списком (второй столбец табл. 2) и перенумерованы в алфавитном порядке (первый столбец).

Процедура улучшения онтологии состоит из трех этапов.

*Первый этап* — использование уже имеющейся (подлежащей улучшению) онтологии для формирования пространства признаков, в котором далее будет строиться новая классификация вершин нижнего уровня.

**Нижний уровень класса  
«Корпоративные информационные системы и услуги,  
связанные с их проектированием и эксплуатацией»**

Но-мер	Наименование ОПД	Код
200	Анализ баз данных	111
201	Аренда приложений в сетях	010
202	Инструменты для моделирования корпоративных информационных систем	100
203	Информационно-поисковые услуги	001
204	Исследования рынка инструментальными средствами	001
205	ИТ-консалтинг	010
206	Лингвистические средства	111
207	Операторские центры (call center/help desk)	110
208	Офисные приложения	001
209	Поддержка бизнес-процессов	101
210	Поддержка групповой работы (groupware)	110
211	Проектирование баз данных	111
212	Проектирование информационных систем	111
213	Разработка архитектуры информационной системы	111
214	Разработка документации	111
215	Разработка программ на заказ (custom)	001
216	Сетевое администрирование	010
217	Системы административного управления	011
218	Системы бухгалтерского и складского учета	001
219	Системы планирования ресурсов (MRP, ERP и др.)	001
220	Системы поддержки принятия решений	001
221	Системы управления документооборотом	001
222	Системы управления запасами	001
223	Системы управления знаниями	001
224	Системы управления персоналом	001
225	Системы управления производством	001
226	Системы управления сбытом (CRM и т. п.)	001
227	Системы финансового анализа	001
228	Службы каталогов	110
229	Средства OLAP/хранилища данных	110
230	Торговые услуги в сети Интернет	001
231	Услуги «Co-location»	110
232	Услуги «Hot-line»	110
233	Услуги в сфере системной интеграции	010
234	Услуги по временному найму персонала	001
235	Услуги по обеспечению информационной безопасности	010
236	Услуги по классификации контента	001
237	Услуги по обеспечению непрерывности деятельности	010
238	Услуги по преобразованию данных	110
239	Услуги по реинжинирингу	001
240	Услуги сетевых систем хранения данных	110
241	Услуги центров обработки данных	110
242	Установка программных средств	010
243	Экспертиза качества информационных систем	101



## Классификация ОПД на 4 класса

Класс	Номер	Наименование ОПД	Значение функции принадлежности			
1	200	Анализ баз данных	1	0	0	0
	206	Лингвистические средства	1	0	0	0
	211	Проектирование баз данных	1	0	0	0
	212	Проектирование информационных систем	1	0	0	0
	213	Разработка архитектуры информационной системы	1	0	0	0
	214	Разработка документации	1	0	0	0
	202	Инструменты для моделирования корпоративных информационных систем	0,8	0,2	0	0
2	201	Аренда приложений в сетях	0	1	0	0
	207	Операторские центры (call center/help desk)	0	1	0	0
	210	Поддержка групповой работы (groupware)	0	1	0	0
	217	Системы административного управления	0	1	0	0
	228	Службы каталогов	0	1	0	0
	229	Средства OLAP/хранилища данных	0	1	0	0
	231	Услуги «Co-location»	0	1	0	0
	232	Услуги «Hot-line»	0	1	0	0
	238	Услуги по преобразованию данных	0	1	0	0
	240	Услуги сетевых систем хранения данных	0	1	0	0
241	Услуги центров обработки данных	0	1	0	0	
3	205	ИТ-консалтинг	0	0	1	0
	216	Сетевое администрирование	0	0	1	0
	233	Услуги в сфере системной интеграции	0	0	1	0
	235	Услуги по обеспечению информационной безопасности	0	0	1	0
	237	Услуги по обеспечению непрерывности деятельности	0	0	1	0
	242	Установка программных средств	0	0	1	0
	243	Экспертиза качества информационных систем	0	0,4	0,6	0
4	203	Информационно-поисковые услуги	0	0	0	1
	204	Исследования рынка инструментальными средствами	0	0	0	1
	208	Офисные приложения	0	0	0	1
	209	Поддержка бизнес-процессов	0,2	0	0	0,8
	215	Разработка программ на заказ (custom)	0	0	0	1
	218	Системы бухгалтерского и складского учета	0	0	0	1
	219	Системы планирования ресурсов (MRP, ERP и др.)	0	0	0	1
	220	Системы поддержки принятия решений	0	0	0	1
	221	Системы управления документооборотом	0	0	0	1
	222	Системы управления запасами	0	0	0	1
	223	Системы управления знаниями	0	0	0	1
	224	Системы управления персоналом	0	0	0	1
	225	Системы управления производством	0	0	0	1
	226	Системы управления сбытом (CRM и т. п.)	0	0	0	1
	227	Системы финансового анализа	0	0	0	1
	230	Торговые услуги в сети Интернет	0	0	0	1
234	Услуги по временному найму персонала	0	0,4	0	0,6	
236	Услуги по классификации контента	0	0	0	1	
239	Услуги по реинжинирингу	0	0,4	0	0,6	

Этот этап выполняется с помощью экспертов. Однако от экспертов, в отличие от методики, описанной в работе [2], уже не требуется определять меру близости между каждой парой ОПД указанием конкретного числа. Перед ними ставится более простая задача: пользуясь уже имеющейся онтологией, для каждого ОПД указать, к какому из трех подклассов второго уровня этой онтологии он относится.

Если эксперт затрудняется выбрать какой-то один подкласс, он может отнести этот ОПД одновременно к двум или даже к трем подклассам. В результате каждый ОПД получает трехпозиционный код из единиц и нулей. Например, если эксперт отнес ОПД «Сетевое администрирование» к подклассу «2.2. Услуги по поддержке жизненного цикла корпоративных информационных систем», этому ОПД ставится в соответствие код 010. Если же, по мнению эксперта, ОПД нельзя отнести ни к одному из подклассов (т. е. эксперт фактически отказывается от классификации этого объекта), ему присваивается код 000 (в рассматриваемом случае таковых не оказалось).

Поскольку разные эксперты могут вкладывать в одни и те же понятия разное содержание, они могут присвоить одним и тем же ОПД разные коды. В нашем случае оказалось, что мнения экспертов расходятся незначительно, поэтому при наличии разногласий код ОПД определялся простым большинством голосов.

Коды, проставленные группой из 10 экспертов — специалистов в данной области для всех 44 ОПД данного класса, приведены в третьем столбце табл. 1.

В случае большей разбросанности мнений возможен и другой способ кодирования. Пусть, например, ОПД «Службы каталогов» восемь экспертов из 10 отнесли к первому классу, 7 — ко второму и ни один — к третьему. Тогда этому ОПД приписывается код (0,8; 0,7; 0). Для каждого  $j$ -го ОПД полученный код можно интерпретировать как вектор значений  $k$  параметров  $x_j = (x_j^{(1)}, \dots, x_j^{(k)})$  (для рассматриваемого примера  $k = 3$ ), т. е.  $j$ -й ОПД может быть представлен точкой  $x_j$  в  $k$ -мерном пространстве  $X$ .

*Второй этап* — автоматическая классификация (кластер-анализ) 44-х ОПД нижнего уровня последовательно на 4, 3 и 2 класса в  $k$ -мерном пространстве  $X$ , полученном на первом этапе (с последующей оценкой экспертами результатов автоматической классификации).

В качестве расстояния в пространстве  $X$  между любыми двумя ОПД можно принять расстояние по Хеммингу (число несовпадающих разрядов в их двоичных кодах) либо по Евклиду, если коды оп-

ределяются путем усреднения мнений экспертов, т. е. на каждой позиции кода может стоять любое число между 0 и 1.

Поскольку некоторые ОПД могут действительно (а не только по мнению экспертов) относиться к двум или трем подклассам одновременно, в работе применяется алгоритм размытой автоматической классификации [3].

В размытом случае каждый классифицируемый объект (точка  $x$  в пространстве признаков  $X$ ) относится ко всем классам одновременно, но с разной степенью принадлежности («достоверности»). Эти величины задаются вектор-функцией принадлежности  $H(x) = (h_1(x), \dots, h_r(x))$ , где  $h_i(x)$  — функция принадлежности точки  $x$  к  $i$ -му классу. Вид размытости задается ограничениями на функцию  $H(x)$  [3]. В алгоритме размытой классификации предусмотрено наличие «фонового» класса, в который попадают объекты, «не поддающиеся» классификации (в нашем примере это все ОПД с кодом 000).

По мнению экспертов, наиболее естественную содержательную интерпретацию имеет классификация ОПД на 4 класса (табл. 2).

1-й класс можно интерпретировать как совокупность ОПД, выполняющих функции общей программной поддержки.

2-й класс — функции организационно-программной поддержки.

3-й класс — функции интеграции сети и поддержания жизненного цикла.

4-й класс — прикладные подсистемы корпоративных информационных систем.

Поскольку в рассматриваемом примере исходная онтология только трехуровневая, классифицировались только ОПД нижнего уровня. В случае существенно многоуровневой исходной онтологии описанную процедуру можно затем использовать для классификации вершин второго, третьего и далее уровней.

*Третий этап* улучшения онтологии — перераспределение классов по уровням онтологической структуры так, чтобы на каждом уровне распределение весов стало более равномерным.

Обратим внимание, что на втором уровне самым большим оказался 4-й класс: он охватывает почти половину всех ОПД. Это обстоятельство на первый взгляд может показаться случайным: если бы классификация прикладных задач, решаемых в корпоративных информационных системах, была не столь детальной (непонятно, например, чем отличаются «офисные приложения» от «поддержки бизнес-процессов», и зачем выделять эти функции, если далее они детализируются), число элементов в 3-м классе могло бы существенно сокра-



таться. Однако размер, или вес класса лучше определять не по числу вошедших в него ОПД, которое может быть в значительной мере связано с недостаточной продуманностью принципов составления списка исходных понятий при разработке онтологии. Желательно, чтобы приписываемые веса так или иначе были связаны с назначением рассматриваемой онтологии. Если онтология используется для разработки профессиональных и образовательных стандартов, веса могут отражать, например, время, необходимое для обучения данной области профессиональной деятельности, относительную потребность в специалистах, выполняющих соответствующие функции, — так называемый коэффициент актуальности, и т. п. (при решении конкретных прикладных задач значения весов классов должны находиться с помощью специальной экспертной процедуры оценивания).

Рассмотрим, для простоты, вторую возможность. Рассмотрим коммерческую компанию среднего размера — с численностью персонала 2—3 тыс. чел. В такой компании для оказания услуг по интеграции сети и поддержания жизненного цикла корпоративных информационных систем (сетевое администрирование и т. п. — см. класс 3) едва ли потребуется больше 10 чел. Функции общей программной поддержки (см. класс 1) едва ли потребуют более 30 чел., а число программистов, разрабатывающих и совершенствующих приложения по заказам функциональных подразделений компании, может исчисляться сотнями, и то их, как правило, не хватает.

Поэтому при оценке весов классов было бы правильно приписывать ОПД этих классов веса, характеризующие «массовость» соответствующей специальности. Например, если ОПД 3-го класса вес 1, то ОПД 1-го и 4-го классов необходимо приписать веса соответственно 3 и 20. При таких весах ОПД большой вес 4-го класса уже не кажется случайностью. Даже если все прикладные функции объединить в одной — «поддержка бизнес-процессов», 4-й класс окажется больше, чем 1-й и 3-й, вместе взятые.

Таким образом, вес четвертого класса «Прикладные системы корпоративных информационных систем» значительно больше весов остальных классов этого уровня и лишь ненамного уступает весу класса высшего уровня «Корпоративные информационные системы». А это означает, что для улучшения онтологии в смысле равномерности распределения весов классов по уровням этот класс следует перенести на следующий (верхний) уровень онтологической структуры.

В результате применения описанных трех этапов экспертно-классификационной процедуры

получен новый вариант онтологии для предметной области «Информатика и информационно-коммуникационные технологии», отличный от исходного варианта. В новом варианте уже не 6 классов ОПД, а 7 — на основе класса «Корпоративные информационные системы и услуги, связанные с их проектированием и эксплуатацией» сформированы два класса — класс с тем же названием, но меньшего веса, а также новый класс: «Прикладные информационные системы».

*Четвертый этап* — экспертное сравнение построенной онтологии с исходной.

Из 10 экспертов, участвовавших в описанной экспертной процедуре, 8 оценили полученную классификацию ОПД на 4 класса как более содержательную, чем априорная (т. е. не основанная на классификации ОПД нижнего уровня) классификация на 3 класса исходной онтологии. Кроме того, все эксперты положительно оценили тот факт, что в новой онтологии нет изолированных вершин.

Далее они признали целесообразным сделанное по формальному критерию перенесение класса 4 на более высокий уровень онтологической структуры.

В результате эксперты решили, что по совокупности двух критериев новая онтология существенно лучше исходной.

## ЗАКЛЮЧЕНИЕ

Предложенная экспертно-классификационная процедура совершенствования онтологии предметной деятельности позволяет формировать классы объектов профессиональной деятельности, более адекватные задаче проектирования профессиональных и образовательных стандартов, что, в свою очередь, повышает эффективность самого проектирования.

## ЛИТЕРАТУРА

1. *Никитин В.В.* Информационно-методические обеспечение формирования перечня направлений и специальностей в области информационно-коммуникационных технологий. — М.: МАКС Пресс, 2006. — 272 с.
2. *Классификация объектов профессиональной деятельности специалиста при проектировании профессиональных и образовательных стандартов / В.В. Никитин и др.* // Проблемы управления. — 2007. — № 4. — С. 51—55.
3. *Бауман Е.В., Дорофеев А.А.* Классификационный анализ данных. // Тр. Междунар. конф. по проблемам управления / ИПУ РАН. — М., 1999. — Т. 1. — С. 62—77.

*Статья представлена к публикации членом редколлегии А.А. Дорофеевом.*

**Никитин Виктор Васильевич** — канд. техн. наук, проректор, Государственный университет — Высшая школа экономики, г. Москва, ☎ (495) 771-32-38, e-mail: vnikitin@hse.ru.