

УДК 519.7:006

В. В. НИКИТИН**ПРОЕКТИРОВАНИЕ ОНТОЛОГИИ ОБЪЕКТОВ
ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ СПЕЦИАЛИСТА
ПРИ РАЗРАБОТКЕ ПРОФЕССИОНАЛЬНЫХ
И ОБРАЗОВАТЕЛЬНЫХ СТАНДАРТОВ**

Рассматривается применение методов классификационного анализа на одном из этапов разработки государственных профессиональных и образовательных стандартов. *Профессиональные и образовательные стандарты; онтология объектов профессиональной деятельности; алгоритмы классификации*

Одна из центральных задач управления процессами подготовки профессиональных кадров заключается в разработке государственных профессиональных и образовательных стандартов. Решение этой задачи представляет собой многоэтапный процесс формирования, анализа, оценивания и обработки больших массивов информации [1].

На первом этапе решения этой задачи формируются классы объектов профессиональной деятельности (ОПД). Исходной информацией здесь является множество ОПД и связей между этими объектами.

На втором этапе оценивается актуальность элементов профессиональной деятельности, т.е. актуальность обучения по тематике, определяемой определенным этапом жизненного цикла (ЖЦ) для конкретного класса объектов профессиональной деятельности. Оценка актуальности является интегральной характеристикой, которая определяется тремя составляющими:

- состоянием рынка труда;
- перспективами развития научно-технического прогресса;
- социальной престижностью профессий и рода занятий.

На третьем этапе формируется перечень направлений профессиональной подготовки, точнее, его промежуточный вариант — множество интегрированных элементов профессиональной деятельности (ИЭПД).

На четвертом этапе формируется множество первичных продисциплин (ППД) — групп дидактических единиц (ДЕ).

Важный начальный этап процесса проектирования профессиональных и образовательных стандартов состоит в определении объектов профессиональной деятельности (ОПД) специалиста, относительно которых «выстраиваются» его профессиональная деятельность и профессиональные компетенции. Образовательные программы, в свою очередь, должны быть построены таким образом, чтобы обеспечивать формирование у обучающихся заданного набора компетенций.

Формирование множества ОПД для определенной предметной области и их классификация представляют собой слабо формализуемую и трудоемкую задачу. Это в первую очередь сказывается на формировании содержания образования в высокотехнологичных, динамически развивающихся отраслях, для которых характерны частое появление новых объектов и связанных с ними понятий, существование большого числа синонимичных понятий, быстрое устаревание некоторых типов объектов, изменения в описании и интерпретации отдельных понятий. Особенно ярко это проявляется на комплексных объектах, которые являются сложной агрегацией множества простых объектов и определяются обобщающими понятиями. Важно отметить, что именно эти понятия, как правило, используются в качестве определений ОПД в стандартах высшего профессионального образования.

В статье рассматривается применение формализованных методов анализа сложно-организованных многомерных данных с использованием как компьютерных, так и экспертных процедур для решения задач, воз-

никающих на этапе формирования классов объектов профессиональной деятельности.

1. ТЕКУЩЕЕ СОСТОЯНИЕ ПРОБЛЕМЫ

Сегодня общепризнанным способом идентификации таких сложных объектов служат онтологии, на основе которых можно определить не только набор более простых понятий, составляющих сложное понятие, но и связи между понятиями (например, степень сходства и различия между ними), что позволяет сформировать набор обобщающих понятий (классов ОПД). Эти понятия являются образующими понятиями для последующего формирования набора компетенций специалиста.

Создание онтологий в различных предметных областях является сегодня одной из самых актуальных задач. Ее решением на разных уровнях занимаются исследовательские центры и международные организации. Одним из наиболее крупных проектов является проект UN SPSC (United Nations Standard Products and Services Code — Стандартный классификатор товаров и услуг), разработка которого ведется экспертной группой UNDO (United Nations Development Organization) и крупной компанией Dun & Bradstreet. Проект развивается на основе опыта использования транснациональных систем и стандартов (SWIFT, EDIFACT и др.) [1]. В настоящее время этот классификатор насчитывает более 13 тыс. категорий товаров и услуг.

Задача формирования обобщающих понятий, используемых при проектировании профессиональных и образовательных стандартов для той или иной предметной области, может быть понята как задача формирования классов простых (исходных) ОПД. Далее под ОПД всюду понимаются именно исходные ОПД. Сложность этой задачи связана, прежде всего, с объемом исходной информации. Число ОПД исчисляется сотнями, а число классов, в которые их нужно объединить, — десятками. При решении задачи вручную эксперт вынужден резко ограничивать число рассматриваемых вариантов классификации, основываясь на своих субъективных оценках. Для повышения объективности и обоснованности классификации желательно использовать формализованные критерии и многовариантные процедуры.

2. МЕТОД РЕШЕНИЯ

Для формализации постановки задачи введем следующие обозначения:

$V = \{vn n = 1, \dots, NV\}$	множество ОПД; элемент множества представляет собой код ОПД
$\chi = \{\chi n n = 1, \dots, NV\}$	вектор наименований ОПД, элементы вектора — наименования ОПД
$\mu = \{\mu n\}, n = 1, \dots, NV$	вектор описаний ОПД, элементы вектора — текстовые описания ОПД

Пусть $R = \{rij\}, i, j = 1 \dots NV, i \neq j$ — матрица «расстояний» между ОПД, элемент rij которой служит численной характеристикой различия между i -м и j -м ОПД ($0 \leq rij \leq 1$).

Цель обработки матрицы расстояний состоит в разбиении всего множества элементов v_1, \dots, v_{NV} на такие подмножества (классы) $q'_1, \dots, q'_{N'}$, чтобы расстояния rij между элементами, попавшими в один класс, были возможно меньше, а между элементами, попавшими в разные подмножества, — возможно больше. Аналогичная постановка задачи предложена в [2], с той разницей, что в [2] используется матрица связей, являющихся мерой «сходства» между ОПД. Такую задачу будем называть задачей классификации исходного множества элементов, а полученные классы $q'_1, \dots, q'_{N'}$ интерпретировать как классы объектов профессиональной деятельности.

В приведенной формальной постановке матрица расстояний R считается заданной (в [2] она задается экспертами — специалистами в данной предметной области). Однако на практике информация о связях между ОПД обычно отсутствует, и ее нужно предварительно получить из содержательных соображений. Ниже предлагается метод построения матрицы расстояний исходя из онтологии предметной деятельности.

Как уже говорилось, онтология предметной деятельности является сегодня общепризнанным способом описания той или иной области профессиональной деятельности. Несмотря на наличие разных определений понятия «онтология», в большинстве практических применений под онтологией понимается некое описание структуры понятий, используемых в данной предметной области [3]. Пример такой онтологии для предметной области «Информатика и информационно-коммуникационные технологии» приведен в [4].

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) контент-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). Например, класс «Корпоративные информационные системы и услуги, связанные с их проектированием и эксплуатацией» 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.

Наконец, самый нижний, третий уровень (уровень исходных понятий, которые и называются объектами профессиональной деятельности, ОПД) представлен 44 ОПД (табл. 1). Существенно, что в данной онтологии (и эта ситуация типичная) связи между ОПД не определены. Это видно из того, что 44 ОПД не распределены даже по подклассам второго уровня: они даются одним списком (графа 2) и перенумерованы в алфавитном порядке (графа 1).

Как отмечается в [3], онтологии имеют смысл только в том случае, если ведется постоянная работа по их совершенствованию. Полезным инструментом для такого совершенствования может служить методика классификационного анализа [5–7]. Покажем возможности применения этой методики на примере той части онтологии предметной области «Информатика и информационно-коммуникационные технологии», которая представлена в табл. 1. Идея заключается в том, чтобы, используя уже существующую онтологию, ввести численную меру близости (расстояние) между исходными ОПД, после чего структурировать множество ОПД с помощью

метода автоматической классификации. Построенная таким образом «снизу» структура, как правило, не будет полностью совпадать с построенной «сверху» из априорных соображений онтологией и может подсказать возможности улучшения исходной онтологии.

Для построения матрицы расстояний между исходными ОПД поставим в соответствие каждому ОПД трехпозиционный код из единиц и нулей, определяющий, к каким подклассам (2.1, 2.2 или 2.3) данный ОПД относится.

Таблица 1

Перечень исходных понятий для класса ОПД «Корпоративные информационные системы и услуги, связанные с их проектированием и эксплуатацией»

200	Анализ баз данных	001
201	Аренда приложений в сетях	001
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	110
208	Офисные приложения	001
209	Поддержка бизнес-процессов	001
210	Поддержка групповой работы	110
211	Проектирование баз данных	001
212	Проектирование информационных систем	111
213	Разработка архитектуры информационной системы	111
214	Разработка документации	111
215	Разработка программ на заказ	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

Окончание табл. 1

228	Службы каталогов	110
229	Средства OLAP/хранилища данных	110
230	Торговые услуги в сети Интернет	001
231	Услуги «Co-location»	111
232	Услуги «Hot-line»	111
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	010
242	Установка программных средств	010
243	Экспертиза качества информационных систем	010

Например, ОПД «Сетевое администрирование» относится к подклассу «2.2. Услуги по поддержке жизненного цикла корпоративных информационных систем» и ему ставится в соответствие код 010. Коды, проставленные экспертами – специалистами в данной области для всех 44 ОПД данного класса, приведены в графе 3 табл. 1.

Отметим одну принципиальную трудность такого кодирования. Она связана с тем, что исходные понятия практически любой онтологии не имеют однозначного определения. Поэтому

1) эксперты могут относить один и тот же ОПД к двум или даже трем подклассам одновременно (например, ОПД 210, 212 и др.);

2) разные эксперты могут присваивать одним и тем же ОПД разные коды.

Однако опыт показал, что в среднем мнения экспертов расходятся незначительно, поэтому при наличии разногласий код ОПД определялся простым большинством голосов.

Проведем теперь автоматическую классификацию ОПД последовательно на 4, 3 и 2 класса, приняв в качестве расстояния между каждой парой ОПД расстояние по Хеммингу (число несовпадающих разрядов в их кодах). Поскольку некоторые ОПД могут относиться к двум или трем подклассам одновременно, воспользуемся алгоритмом размытой классификации [5].

Наиболее естественную содержательную интерпретацию имеет классификация на 3 класса (табл. 2).

Таблица 2

№ кл	Код	Наименование ОПД	Функция принадлежности
1	206	Лингвистические средства	1 0 0
1	212	Проектирование информационных систем	1 0 0
1	213	Разработка архитектуры информационной системы	1 0 0
1	214	Разработка документации	1 0 0
1	231	Услуги «Co-location»	1 0 0
1	232	Услуги «Hot-line»	1 0 0
1	207	Операторские центры	0,8 0,2 0
1	210	Поддержка групповой работы	0,8 0,2 0
1	228	Службы каталогов	0,8 0,2 0
1	229	Средства OLAP/хранилища данных	0,8 0,2 0
1	238	Услуги по преобразованию данных	0,8 0,2 0
1	240	Услуги сетевых систем хранения данных	0,8 0,2 0
1	202	Инструменты для моделирования корпоративных информационных систем	0,6 0,2 0,2
1	217	Системы административного управления	0,4 0,3 0,3
2	205	ИТ-консалтинг	0 1 0
2	216	Сетевое администрирование	0 1 0
2	233	Услуги в сфере системной интеграции	0 1 0
2	235	Услуги по обеспечению информационной безопасности	0 1 0
2	237	Услуги по обеспечению непрерывности деятельности	0 1 0
2	241	Услуги центров обработки данных	0 1 0
2	242	Установка программных средств	0 1 0
2	243	Экспертиза качества информационных систем	0 1 0
3	200	Анализ баз данных	0 0 1
3	201	Аренда приложений в сетях	0 0 1
3	203	Информационно-поисковые услуги	0 0 1
3	204	Исследования рынка инструментальными средствами	0 0 1

Окончание табл. 2

3	208	Офисные приложения	0 0 1
3	209	Поддержка бизнес-процессов	0 0 1
3	211	Проектирование баз данных	0 0 1
3	215	Разработка программ на заказ	0 0 1
3	218	Системы бухгалтерского и складского учета	0 0 1
3	219	Системы планирования ресурсов (MRP, ERP и др.)	0 0 1
3	220	Системы поддержки принятия решений	0 0 1
3	221	Системы управления документооборотом	0 0 1
3	222	Системы управления запасами	0 0 1
3	223	Системы управления знаниями	0 0 1
3	224	Системы управления персоналом	0 0 1
3	225	Системы управления производством	0 0 1
3	226	Системы управления сбытом (CRM и т.п.)	0 0 1
3	227	Системы финансового анализа	0 0 1
3	230	Торговые услуги в сети Интернет	0 0 1
3	234	Услуги по временному найму персонала	0 0 1
3	236	Услуги по классификации контента	0 0 1
3	239	Услуги по реинжинирингу	0 0 1

1-й класс можно интерпретировать как совокупность ОПД, выполняющих функции общей поддержки корпоративных информационных систем.

2-й класс — услуги по поддержанию жизненного цикла корпоративных информационных систем.

3-й класс — прикладные подсистемы корпоративных информационных систем.

Обратим внимание, что 3-й класс оказался самым большим: он включает 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), количество элементов в 3-м классе могло бы существенно сократиться. Однако размер класса лучше определять не по количеству вошедших в него ОПД, кото-

рое в значительной мере связано с недостаточной продуманностью принципов составления списка исходных понятий при разработке онтологии. Более объективным измерителем является относительная потребность в специалистах, выполняющих соответствующие функции.

Рассмотрим, например, среднюю компанию с численностью персонала 2–3 тыс. чел. В такой компании для оказания услуг по поддержанию жизненного цикла корпоративных информационных систем (сетевое администрирование и т. п., класс 2) едва ли потребуется больше 10 чел. Функции общей поддержки (класс 1) едва ли потребуют более 50 чел., а количество программистов, разрабатывающих и совершенствующих приложения по заказам функциональных подразделений компании, может исчисляться сотнями, и их, как правило, не хватает. Поэтому при оценке размеров классов было бы правильно приписывать ОПД этих классов соответствующие веса. Например, если ОПД 2-го класса приписать вес 1, то ОПД 2-го и 3-го классов должны иметь веса соответственно 5 и 20.

При таких весах большой размер 3-го класса уже не кажется случайностью. Даже если все прикладные функции объединить в одной — «поддержка бизнес-процессов», 3-й класс окажется больше, чем 1-й и 2-й, вместе взятые.

Таким образом, результаты автоматической классификации показывают, что подкласс 2.3 второго уровня, который мы вначале обозначили как «Прикладные системы корпоративных информационных систем», по размерам ненамного уступает самому классу первого уровня «Корпоративные информационные системы». А это означает, что он должен находиться на первом уровне онтологии и рассматриваться как самостоятельный класс ОПД.

В результате мы получаем улучшенный вариант онтологии для предметной области «Информатика и информационно-коммуникационные технологии», приведенной в [4]. В этом 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 размера, и новый класс «Прикладные информационные системы».

3. ПРИЛОЖЕНИЕ РЕЗУЛЬТАТОВ

Данная методология и алгоритмы [1–3] нашли применение при проектировании новых образовательных стандартов высшего профессионального образования по направлениям «Бизнес-информатика» и «Программная инженерия», а также при разработке профессиональных стандартов для отрасли информационных технологий [8].

ВЫВОДЫ

Предложенный метод построения онтологии предметной области на основе классификационного анализа позволяет повысить формализуемость и объективность процесса проектирования профессиональных и образовательных стандартов.

СПИСОК ЛИТЕРАТУРЫ

1. **Никитин, В. В.** Информационно-методическое обеспечение формирования перечня направлений и специальностей в области информационно-коммуникационных технологий / В. В. Никитин. М. : МАКС Пресс, 2006. 272 с.
2. **Никитин, В. В.** Классификация объектов профессиональной деятельности специалиста при проектировании профессиональных и образовательных стандартов / В. В. Никитин, С. В. Мальцева, А. А. Дорофеюк [и др.] // Проблемы управления. 2007. № 4. С. 51–55.
3. **Никитин, В. В.** Формирование номенклатуры направлений подготовки специалистов на основе многопараметрической модели профессиональной деятельности / В. В. Никитин, С. В. Мальцева, А. А. Дорофеюк [и др.] // Автоматизация и современные технологии. 2008. № 5. С. 38–44.
4. **Проектирование** нового перечня направлений (специальностей) и государственных образовательных стандартов среднего профессионального и высшего профессионального образования для динамично развивающихся высокотехнологичных отраслей на основе технологий управления знаниями. Отчет ГУ ВШЭ № 1327. М., 2006.
5. **Мандель, А. С.** Экспертно-статистические системы в задачах управления и обработки информации : ч. I / А. С. Мандель // Приборы и системы управления. 1996. № 12.
6. **Мандель, А. С.** Экспертно-статистические системы в задачах управления и обработки информации : ч. II / А. С. Мандель // Приборы и системы управления. 1997. № 2.
7. **Дорофеюк, А. А.** Опыт использования экспертных комиссий в задачах организационного управления / А. А. Дорофеюк, А. Л. Чернявский // Методы и алгоритмы анализа эмпирических данных. М. : Ин-т проблем управления, 1988.
8. **Профессиональные** стандарты в области информационных технологий. М. : АПКИТ, 2008. 616 с.

ОБ АВТОРЕ

Никитин Виктор Васильевич, проректор гос. ун-та «Высшая школа экономики». Дипл. инж.-системотехн. (Моск. ин-т электрон. машиностр., МИЭМ, 1977). Канд. техн. наук по техн. киберн. и информатике (МИЭМ, 1984). Иссл. в обл. проф. и образовател. стандартов, электрон. государства, архитектуры предприятий.

