

УДК 004.9

РАЗРАБОТКА СИСТЕМЫ КЛАССИФИКАЦИИ ПРОФЕССИОНАЛЬНЫХ КОМПЕТЕНЦИЙ В ИТ-СФЕРЕ

Питухин Е.А., Тулаева А.И., Шабалина И.М.

ФГБОУ ВО «Петрозаводский государственный университет», Петрозаводск,
e-mail: eugene@petsu.ru, anike@petsu.ru, Ishabalina@opti-soft.ru

В настоящей статье рассматривается вопрос классификации перечня компетенций профессий, востребованных работодателями ИТ-сферы. Классификация осуществляется путем структурирования компетенций по ключевым признакам: типу, уровню владения и т.д. В качестве исследуемой области выступает область информационных технологий как одна из динамически развивающихся сфер деятельности современного общества. Для анализа были выбраны вакансии наиболее популярных профессий, востребованных лидерами среди мировых работодателей в ИТ-области. Для обработки полученного перечня вакансий был разработан алгоритм классификации компетенций, основанный на анализе текстов. Произведена оценка основных метрик качества работы алгоритма. В результате работы предложенного алгоритма классификации была получена онтологическая информационная модель востребованных компетенций ИТ-сферы. Статистический анализ полученной модели позволил выявить особенности компетентностной структуры каждой из выбранной профессий, получить количественные характеристики параметров классификации, а также выявить закономерности в потребностях работодателей ИТ-области.

Ключевые слова: онтология, работодатель, ИТ-сфера, компетенции, классификация, алгоритм

THE DEVELOPMENT OF SYSTEM FOR CLASSIFICATION OF PROFESSIONAL COMPETENCIES IN IT-SPHERE

Pitukhin E.A., Tulaeva A.I., Shabalina I.M.

Petrozavodsk State University, Petrozavodsk, e-mail: eugene@petsu.ru,
anike@petsu.ru, Ishabalina@opti-soft.ru

This article is devoted to classification of professional competencies, which were demanded by foreign employers. We could structure them and developed the ontology of competencies. As the researched area we selected the sphere of the information technologies as the most demanded field of activity. We analyzed the most popular vacancies from the most known foreign employers in IT-sphere. We developed the algorithm for classification of competencies. It is based on the analysis of texts. We estimated the work quality of our algorithm. We made a thorough analysis of the results of classification. At last we could receive the information model of competencies in IT-sphere.

Keywords: ontology, employer, information technology, competencies, classification, algorithm

Одной из главных целей настоящего исследования является создание единой онтологии компетенций, которая позволила бы сравнивать и сопоставлять между собой требования к компетенциям, как со стороны работодателей, так и со стороны университетов.

Мы придерживаемся парадигмы, что рынок образовательных услуг считается вторичным по отношению к рынку труда, который сам является вторичным по отношению к рынку производства товаров и услуг [1, 2]. Поэтому главным принципом обеспечения возможности сравнения компетенций является, на наш взгляд, приведение в соответствие системы «образовательных» компетенций учебных планов университетов к системе «профессиональных» компетенций работодателя, а не наоборот.

В статье [3] авторами подробно изложена идея информационной системы, описывающей процессы получения компетенций, трудоустройства и управления учебными планами в системе «вуз – студент – работо-

датель». Для создания такой системы авторы предлагают:

1) связать трудоустройство с процессом получения компетенций через разработку общей онтологии компетенций, объединяющей образовательные и профессиональные компетенции;

2) разработать процесс обратной связи, влияющей на содержание учебных планов вузов на основе мониторинга процессов трудоустройства и получения компетенций.

Большинство профессиональных требований (или, другими словами, компетенций), которые предъявляет работодатель к соискателю работы, находят свое отражение в объявлениях о работе. Таким образом, одним из основных источников информации о компетенциях являются тексты объявлений на требуемые вакансии, которые размещают работодатели в открытом доступе.

В рамках данного подхода мы принимаем допущение, что «скрытая» часть вакансий, которая не публикуются в открытом до-

студе и реализуется по внутрифирменным каналам, знакомству и т.п., не оказывает существенного влияния на структуру «ядра» компетенций. Поскольку в выбранной нами для рассмотрения IT-отрасли при приеме на работу существует особая специфика, заключающаяся в широком использовании интернет-технологий при поиске работы и высокой эффективности размещения объявлений на специализированных сайтах, популярных в среде IT-специалистов.

Методология исследования основана на одновременном сочетании компетентностного подхода к описанию предметной области и информационного подхода к формализации компетенций различных категорий в виде онтологической модели.

Основными задачами исследования являются разработка системы измерения компетенций для создания многомерной модели компетенций, а также создание алгоритма классификации, с помощью которых неструктурированное множество «профессиональных» компетенций можно было представить в виде онтологии.

Исходные данные и выбор признаков классификации

В качестве первого шага, чтобы приступить к формированию многомерной модели компетенций, были собраны актуальные данные о наиболее востребованных профессиях в IT-сфере. Были проанализированы вакансии наиболее известных мировых работодателей IT-отрасли на таких сайтах поиска работы, как job.com, jobview.monster.com, indeed.com, apple.com, visa.com, hp.com и др.

Исследовались вакансии для TOP-10 рейтинга наиболее популярных профессий IT-отрасли, регулярно публикуемого на ресурсе jobsearch.about.com. «Here are the top 10 jobs for computer science majors, based on income, employment outlook, and job satisfaction» [4].

Сформированный перечень состоял из следующих профессий, названия которых стали значениями первого параметра измерения OCCUPATION = {Database administrator, Computer hardware engineer, Web developer, Computer network architect, Computer programmer, Computer systems analyst, Information security analysts, Java developer, Project manager, Software engineer}. Для каждой из 10 профессий были рассмотрены вакансии от не менее чем 10 различных работодателей. В результате исследования 100 вакансий был получен общий перечень из 396 уникальных компетенций с указанием количества повторов, если такие наблюдались несколько раз.

Число повторений, представленное параметром COUNT = {1, 2, 3...10}, было выбрано в качестве количественной меры многомерного концептуального представления характеризующей частоту востребованности компетенции.

Сравнительный анализ позволил выявить различные наиболее часто встречающиеся смысловые категории компетенций, которые были выделены в параметр TYPE, представляющий независимое измерение. Параметр TYPE может принимать следующие значения: TYPE = {Knowledge, Skills, Experience}.

Последующий лексический анализ обнаружил, что каждый тип характеризуется некоторым уровнем владения компетенцией. Он описывается показателем LEVEL = {Min, Mid, Max}.

В нашей первой работе [5], посвященной проблеме классификации компетенций, кроме вышеуказанных значений параметра TYPE имелись и другие, TYPE = {Certificat, Degree, Understanding}. В то же время параметр LEVEL принимал 15 различных значений, LEVEL = {None, Basic, General, Possession, Conceptual, Ability, Proficiency, Comfortable, Familiarity, Excellent, Broad, Advanced, Strong, Strength, Expert, Direct}. Основываясь на современных российских образовательных и профессиональных стандартах, нам пришлось пересмотреть параметры классификации и оптимизировать их. Так, показатель TYPE = {Certificat, Degree} вынесен за рамки вышеописанной классификации. Он выделен сейчас как отдельный класс EDU, который характеризуется требованиями к соискателю иметь в наличии диплом о соответствующем образовании. Показатель TYPE = {Understanding} в силу своей схожести объединен с TYPE = {Skills}. Что касается уровней владения компетенцией, все они были также укрупнены исходя из их смысловых значений. Так, LEVEL = {None, Basic, General} теперь соответствует LEVEL = {Min}. LEVEL = {Excellent, Broad, Advanced, Strong, Strength, Expert, Direct} объединен в LEVEL = {Max}. Остальные значения показателя уровня владения отнесены к LEVEL = {Mid}.

Параметры классификации TYPE и LEVEL являются не категориальными, а порядковыми признаками, имеют как лексическое (текстовое), так и соответствующее им целочисленное значение {1, 2, 3}, позволяющее их ранжировать в порядке силы влияния, что важно для последующего анализа онтологии.

Итоговым результатом классификации является параметр «OBJECT», содержащий

предметную область компетенции, за вычетом TYPE, LEVEL, грамматик (союзов и предлогов), а также знаков препинания. Например, OBJECT = {Computer Science, Microsoft SQL Server 2008, Communication, Math, XML, Databases, Java, Linux, etc...}.

Таким образом, предлагаемая система классификации позволяет создать онтологическую модель компетенций в четырех измерениях: «OCCUPATION» – «TYPE» – «LEVEL» – «OBJECT», где количественной мерой является параметр COUNT. Данная система соответствует многомерному концептуальному представлению по Кодду (multi-dimensional conceptual view) [6].

Алгоритм классификации ИТ-компетенций

Для формирования соответствующей многомерной модели компетенций был разработан алгоритм классификации, с помощью которых неструктурированное множество «профессиональных» компетенций можно было представить в виде онтологии.

Ниже представлен общий алгоритм классификации компетенций.

1. Вызвать процедуру для выбора пользовательского режима разделения/неразделения компетенций, если результат выбора = Ложь, перейти на строку 8.

2. Вызвать алгоритм 1 для разделения компетенций при наличии в ней нескольких типов одновременно.

3. Вызвать алгоритм 2 для разделения компетенции при наличии «,» и «in» или «and».

4. Вызвать алгоритм 3 для разделения компетенции при наличии «with» и «and» или «,» или «or».

5. Вызвать алгоритм 4 для разделения компетенции при наличии «of» и «and» или «,».

6. Вызвать алгоритм 5 для разделения компетенции при наличии «in» и «and» или «or».

7. Конец условия.

8. Вызвать алгоритм 6 для разделения компетенции на отдельные теги (слова).

9. Вызвать алгоритм 7 для определения класса Edu и типа Type компетенций.

10. Вызвать алгоритм 8 для определения уровня владения компетенцией Level.

11. Вызвать алгоритм 9 для определения грамматик Gramm.

12. Вызвать алгоритм 10 для определения предметной области Object.

Важной особенностью предложенного алгоритма является его способность работать в режиме разделения сложных компетенций на простые, когда строка с несколькими компетенциями делится на несколько строк с одной компетенцией в каждой. Такой режим необходим, например, когда в объявлении о работе несколько фактически раз-

личных компетенций перечислены в одной строке. Например, компетенция «Experience and knowledge of ETL Development» является сложной, так как совмещает в себе 2 типа компетенций – с одной стороны опыт, с другой – знания. Компетенция «Experience in Microsoft SQL Server 2008, Administration experience» также является сложной, так как в первой ее части до запятой подразумевается непосредственно опыт работы, а во второй части опыт администрирования. Аналогичных примеров в списке компетенций предостаточно.

Возможность алгоритма разделять сложные компетенции на простые повышает корректность результатов его работы. Ниже, в таблице приведены основные метрики качества работы алгоритма в двух режимах – с разделением сложных компетенций на простые и без разделения.

Оценки работы алгоритма классификации

№	<i>n</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>P</i>	<i>T</i>	<i>F</i>
I	126	119	7	0	1	94%	97%
II	189	175	14	0	1	93%	96%
III	270	225	45	0	1	83%	91%
IV	483	400	83	0	1	83%	91%

В таблице используются следующие обозначения: *n* – число компетенций; *a* – правильно распознанные компетенции; *b* – неправильно распознанные компетенции; *c* – нераспознанные компетенции; *P* – полнота распознавания; *T* – точность распознавания; *F* – F-метрика.

При этом оценивались два режима работы алгоритма для двух типов выборок – обучающей и контрольной: I – обучающая выборка (без разделения компетенций); II – обучающая выборка (с разделением компетенций); III – контрольная выборка (без разделения компетенций); IV – контрольная выборка (с разделением компетенций).

Из таблицы видно, что качество классификации в режиме без разделения и с разделением компетенций отличаются незначительно. Тем не менее разница между качеством обучающих и контрольных выборок меньше у алгоритма с разделением компетенций. Кроме того, общее число реальных «простых» компетенций после разделения составило 672 компетенции против 396 в начале, то есть выросло более, чем на 58%!

На основании приведенных оценок результаты работы данного алгоритма можно считать вполне приемлемыми.

Результаты исследования и их обсуждение

В результате работы предложенного алгоритма была получена многомерная

модель востребованных компетенций IT-отрасли, которая также может быть представлена в виде онтологии (структурированного перечня).

Для презентации были выбраны результаты классификации работы алгоритма в режиме разделения компетенций на объединенном множестве из обучающей и контрольной выборки, мощность которого составила 575 компетенций (таблица).

На рис. 1 приведена диаграмма структурного распределения различных типов компетенций по рассматриваемым профессиям.

В представленной структуре, несмотря на явное разнообразие по профессиям в сред-

нем, заметна преобладающая роль двух типов «Experience» и «Knowledge» – 46%. Менее значимым из всех является «Skills» – 8%.

На рис. 2 показано распределение значений параметра измерения LEVEL (уровней владения компетенциями) по типу компетенций.

Видно, что самый большой процент для всех типов – это начальный уровень владения компетенцией – Min. Максимальная доля уровня Min достигается для компетенции типа «Experience» – 81%. И наоборот, максимальная доля самого высшего уровня Max – 36% относится к компетенции «Skills».

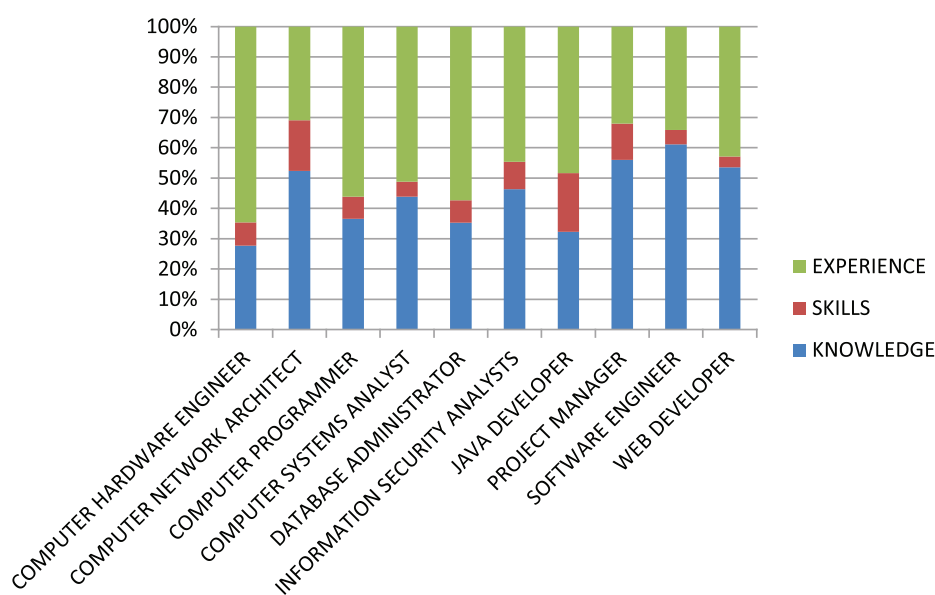


Рис. 1. Распределение структуры типов компетенций по профессиям IT-сферы

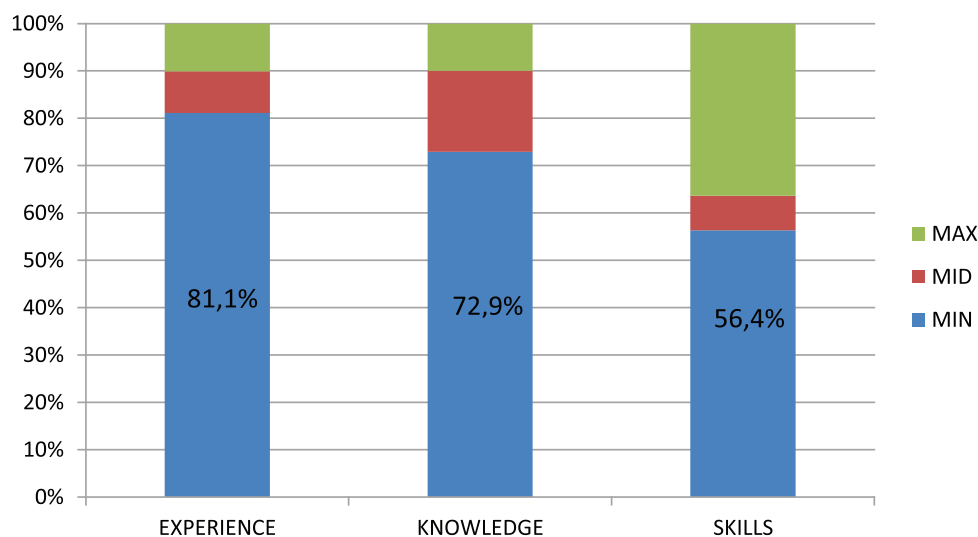


Рис. 2. Распределение структуры уровней владения компетенциями по типу

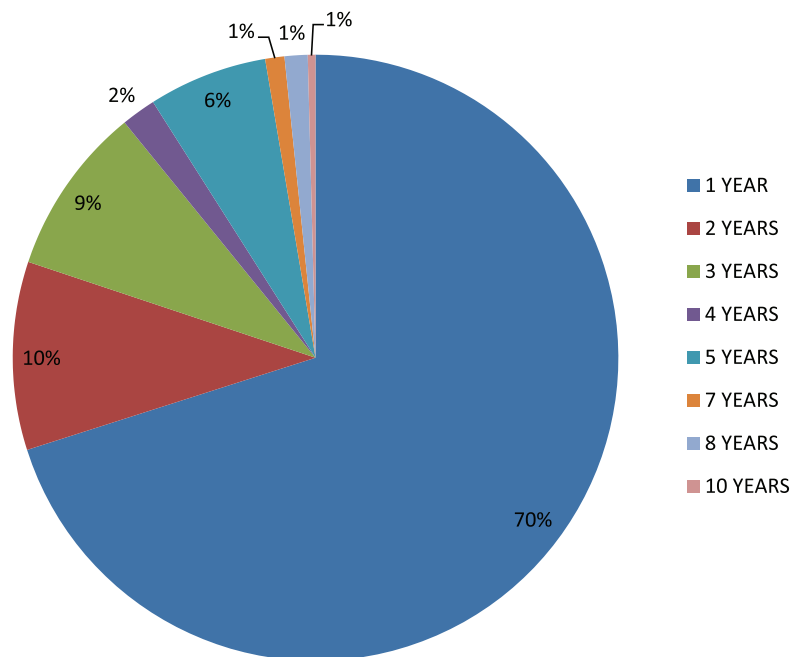


Рис. 3. Распределение требований к опыту работы по компетенциям

Компетенция типа «Experience» имеет свою особенность, которая присуща только ей – интерпретацию ее уровней будет правильно проводить не по лексическим значениям, а по соответствующим им целочисленным, так как в данном случае они несут смысл лет опыта работы. Более корректное представление компетенции типа «Experience» приведено на рис. 3.

Из рисунка видно, что распределение структуры опыта по годам идет в обратном порядке: около 70% объявлений не указывает число лет опыта работы, что интерпретируется как минимум 1 год; около 10% требуют наличие опыта не менее 2 лет; 9% не менее 3 лет и т.д. Максимальным требованием к опыту работы является значение не менее 10 лет.

Стоит отметить еще одну особенность классификации компетенций – распределение компетенций по классу Edu, с требованиями о наличии соответствующего образования (диплома, сертификата и т.д.) по типам Type. Наличие образования является не менее важным критерием при отборе претендентов на вакантные места. Согласно нашим исследованиям, 83% всех имеющихся требований о наличии диплома или сертификата относятся к типу Knowledge. Причем 95% из них встречается при минимальном уровне владения компетенцией, т.е. при LEVEL = Min. Остальные 17% всех требований об образовании принадлежат типу Experience. В типе Skills такие требования полностью отсутствуют.

Статистический анализ полученной нами модели позволил выявить особенности компетентной структуры для каждой из 10 профессий, получить количественные характеристики для используемых параметров классификации, а также приобрести новое знание о потребностях работодателей ИТ-отрасли.

Полученные результаты открывают в будущем потенциальную возможность сравнения компетенций работодателей с компетенциями дисциплин в учебных планах, что является следующим важным этапом проводимой серии исследований.

Список литературы

1. Бурденко Е.В. Рынок образовательных услуг в трансформируемой экономике: автореф. дис.. канд. эконом. наук. – Москва, 2004. – С. 22.
2. Мушкетова Н.С. Маркетинговое взаимодействие рынка труда и рынка образовательных услуг / Философия социальных коммуникаций. – 2013. – № 3 (24). – С. 88–99.
3. Варфоломеев А.Г., Питухин Е.А., Тулаева А.И. Концепция информационной системы управления учебными планами // Университетское управление: практика и анализ. – 2016. – № 5(105). – С. 122–132.
4. Top 10 jobs for computer science majors. URL: <https://www.thebalance.com/top-jobs-for-computer-science-majors-2059634> (дата обращения: 17.03.2017).
5. Pitukhin E., Varfolomeyev A., Tulaeva A. Job Advertisements Analysis for Curricula Management: The Competency Approach // ICERI2016 Proceedings. 9th International Conference of Education, Research and Innovation, 14–16 November 2016, Seville, Spain. IATED. P. 2026–2035.
6. Codd E.F. «Is Your DBMS Really Relational?» ComputerWorld, № 14, October 1985.