

Онтологическое моделирование потребностей в IT-специалистах на основе веб-данных

Р. А. Файзрахманов¹, Д. В. Яруллин²

Пермский национальный исследовательский политехнический университет

¹fayzrakhmanov@gmail.com, ²d.v.yarullin@ya.ru

Аннотация. В работе описывается подход к построению онтологической модели профессиональных навыков разработчика ПО, необходимых работодателям, на основе данных вакансий, опубликованных на сервисах-агрегаторах. Указывается на необходимость построения онтологической модели для анализа существующих потребностей экономики в IT-специалистах, обладающих требуемыми умениями. Проводится анализ существующих методов онтологического моделирования высококвалифицированных специалистов, выявляются их достоинства и недостатки. Описывается основной алгоритм извлечения данных из текстов вакансий и их структуризации при помощи денотатного графа. Приводится демонстрация работы алгоритма и проводится анализ полученной модели. Описываются дальнейшие перспективы развития представленного подхода.

Ключевые слова: онтология; денотатный граф; извлечение сущностей; онтологическое моделирование

I. ФОРМУЛИРОВКА ПРОБЛЕМЫ

Несмотря на то, что развитие IT-отрасли и подготовка высококвалифицированных кадров является одним из приоритетных направлений в экономике Российской Федерации [1], в настоящее время спрос на квалифицированных специалистов в значительной мере превышает предложение. Более 40 тысяч выпускников ежегодно [2] получает диплом по специальностям, связанным с информационными технологиями, но в отрасли сохраняется ситуация «кадрового голода». Очевидно, что данное положение дел связано, вероятнее всего, с несоответствием навыков выпускников вузов и требований работодателей. В связи с этим возникает необходимость анализа существующей потребности экономики в IT-специалистах и навыков, которыми они должны обладать.

Для решения задачи подобного рода необходима соответствующая технология, модель, систематизирующая и структурирующая компетенции востребованного в настоящее время IT-специалиста, фактически, создающая его профессиональный портрет.

Одним из возможных подходов к представлению данного портрета является онтология. Наиболее общее определение было дано Штрудером и др. [3] и основано на более ранних определениях Ушолда [4] и Грубера [5]. Исследователи определяют онтологию как «формальную явную спецификацию общей концептуализации».

Формальное свойство онтологии определяет ее машиночитаемый формат. Явная спецификация указывает на явно определенные понятия, свойства, ограничения и экземпляры объектов онтологии. Общая концептуализация — абстрактная модель некоторого явления, принимаемая некоторой группой людей.

В упрощенном виде онтологию можно описать как модель знаний, содержащую набор понятий, описывающих определенную предметную область и доступных для машинной обработки [6]. Данные понятия систематизированы при помощи иерархических отношений.

Онтологическая модель обеспечивает гибкость и системный подход к предметной области. Онтология представляет целостный взгляд на предметную область и позволяет восстановить недостающие логические связи [7]. Данное свойство онтологий является наиболее ценным для нас, поскольку позволяет на основе ограниченного множества навыков, указанных в текстах вакансий, описать полный профессиональный портрет специалиста, требующегося работодателю на эту позицию.

Стоит также заметить, что моделирование потребностей экономики в IT-специалистах неотрывно связано с вопросами их подготовки в высших учебных заведениях. Принимая во внимание важность этого аспекта проблемы, мы, тем не менее, в настоящей работе фокусируемся на возможных методах создания технологии, позволяющей оценить потребность в тех или иных компетенциях специалиста именно с точки зрения работодателей и, соответственно, экономики в целом.

II. СУЩЕСТВУЮЩИЕ ПОДХОДЫ

Как отмечают исследователи, оценка необходимых специалисту знаний, умений и навыков может проводиться на основе компетенций, прописанных в профессиональных стандартах. Как следствие, подготовка специалиста должна опираться на наиболее востребованные в настоящее время профессиональные стандарты. В качестве возможного подхода предлагается формирование тезауруса предметной области, построение и использование механизма определения необходимых компетентностей с использованием технологии онтологического моделирования и построения баз знаний.

В частности, В.В. Мартынов и др. [8] предлагают выделить при обработке требований работодателей следующие концепты:

- комплекс требований работодателей к системе обучения, проявляющийся в наборах квалификационных требований профессиональных стандартов;
- компетентностные модели выпускаемых специалистов;
- образовательная программа;
- обучающая система;
- заинтересованные лица процесса.

Необходимо отметить, что в предложенном подходе не раскрывается специфика «требований работодателя», наиболее важной, с нашей точки зрения, части онтологии.

Понятие компетентностной модели раскрывается в работе Т. М. Шамсутдиновой [9]. Как указывает исследователь, теоретическим и практическим вопросам реализации компетентностно-ориентированной модели образовательного процесса посвящены фундаментальные работы И. А. Зимней, Ю. Г. Татура, А. В. Хуторского, В. И. Байденко, С. Е. Шишова, Э. Ф. Зеера, Н. Ф. Ефремовой и многих других.

Предложенная Т. М. Шамсутдиновой модель на опирается объективную концепцию отношения сущностей «компетенции» – «учебные дисциплины». Онтологическая модель позволяет в данном случае выявить основные взаимосвязи в исследуемой области, дать их интерпретацию в виде семантической сети.

Предложенная модель позволяет выявить дисциплины в рамках программы подготовки, напрямую оказывающие влияние на формирование той или иной профессиональной компетенции, а также знаний, умений и навыков в рамках этой компетенции.

Заметим, что модель Т. М. Шамсутдиновой нельзя назвать полной: в ней не раскрываются конкретные умения и навыки, формируемые в рамках компетенции, и невозможно оценить степень влияния каждой отдельно взятой дисциплины на освоение соответствующих умений и навыков. Это затрудняет соотнесение результатов подготовки в рамках программы с требованиями, описанными в профессиональных стандартах и вакансиях работодателей.

Исследователь также отмечает, что оценка степени сформированности компетенций является сложной, многопараметрической задачей, зависящей от большого числа компонент; данная оценка является одним из основных факторов для формирования индивидуальной образовательной траектории обучения студентов. На наш взгляд, опора на знания, умения и навыки, требуемые на рынке труда, позволили бы упростить решение данной задачи и скорректировать «индивидуальную образовательную траекторию» соответствующим образом для повышения эффективности подготовки.

М. Бюссемакер и др. [10] предлагают опираться при создании онтологической модели на предметную область. Такая структура модели позволяет для каждой специальности в рамках предметной области получать список необходимых знаний, умений и навыков, а также сферы их применения. Кроме того, данная онтологическая модель напрямую связана с реальными производственными процессами и бизнес-логикой.

Формирование программы подготовки и, соответственно, знаний и умений осуществляется с учетом ряда заданных правил и ограничений, тем самым обеспечивая структуризацию профессиональных компетенций в соответствии с предметной областью. Это также способствует вертикальной и горизонтальной интеграции между отдельно взятыми элементами программы подготовки

Мы можем отметить, что подход к построению онтологической модели профессиональных компетенций на основе модели предметной области кажется нам достаточно перспективным, однако в работе не уделяется достаточного внимания методам получения актуальных данных от работодателей.

С нашей точки зрения, для модели, отражающей актуальную потребность экономики в IT-специалистах, необходимо использование информации, непосредственно публикуемой работодателями и регулярно обновляемой.

Обобщая вышесказанное, заметим, что исследователи, предлагая возможные методы построения онтологической модели специалиста, в большей степени ориентируются на процесс подготовки специалиста и не ставят перед собой задачу реализации модели, способной динамически корректировать и адаптировать профессиональный портрет на основе запроса экономики в настоящий момент. Этим, как мы полагаем, наш подход выгодно отличается от уже существующих разработок в данной области.

III. ПОСТРОЕНИЕ МОДЕЛИ НА ОСНОВЕ ВЕБ-ДАННЫХ

В качестве базовой сущности предлагаемой нами модели мы используем «навык». Под навыком мы понимаем фрагмент информации в рамках предметной области, владение которым позволяет решать определенные профессиональные задачи. В качестве примеров мы можем привести языки программирования (Python, Java, C++), отдельные библиотеки и фреймворки (Pandas, Angular JS), технологии и протоколы (UDP, контейнерная виртуализация), операционные системы (Linux, macOS), прикладные приложения (Adobe Photoshop). Именно навыки чаще всего указываются работодателями в требованиях к соискателю.

Каждый навык рассматривается в контексте иерархической модели предметной области: так, навык «React» будет являться подклассом «JavaScript», а сам «JavaScript», в свою очередь, войдет в более широкий класс «Языки программирования». Вертикальные и горизонтальные связи между различными элементами модели позволяют на основе 5-10 навыков, указанных в вакансии, выстроить целостный портрет IT-специалиста.

Нами было принято решение использовать в качестве основного источника данных о навыках тексты вакансий, опубликованные в открытом доступе на сайтах-агрегаторах. В числе наиболее популярных русскоязычных ресурсов — «HH.ru», «Мой круг», «Яндекс.Работа».

Для пилотного исследования мы сфокусировали наше внимание на вакансиях программистов и разработчиков ПО, опубликованных на агрегаторе «HH.ru». Поскольку мы предполагаем, что портрет IT-специалиста связан с экономикой региона, для каждого региона мы формировали собственный список навыков.

При помощи HeadHunter API [11] мы получили список кодов регионов, полные тексты вакансий, а также метаданные агрегатора. Используя библиотеку для обработки естественного языка RuMystem, мы извлекали из текстов вакансий навыки. Также для формирования списка использовалась информация из поля «ключевые навыки», если она была указана.

Полученные списки были подвергнуты частотному анализу, после чего по каждому региону сформирован список наиболее востребованных навыков для программистов и разработчиков (рис. 1).

Наиболее востребованные (т.е. наиболее частотные) навыки по регионам были также визуализированы на интерактивной карте, фрагменты которой приведены на рис. 2 и 3, что позволило наглядно увидеть различия в навыках, которые требуются соискателям-программистам в различных субъектах РФ.

```
"1317": {
  "1С программирование": 11,
  "1С: Комплексная автоматизация": 3,
  "1С: Розница": 1,
  "Обновление конфигурации 1С": 1,
  "Python": 15,
  "Git": 59,
  "MySQL": 18,
  "PostgreSQL": 19,
  "Linux": 22,
  "HTML5": 20,
  "PHP": 38,
  "HTML": 29,
  "CSS": 34,
  "jQuery": 18,
  "SQL": 39,
  "Ajax": 14,
  "PHP5": 3,
  "ООП": 35,
  "Bootstrap": 4,
  "Yii": 12,
  "CSS3": 16,
  "Java": 32,
  "Adobe Photoshop": 4
```

Рис. 1. Фрагмент списка денотатов, извлеченных из вакансий по запросу «программист» в Пермском крае, с указанием их частот.

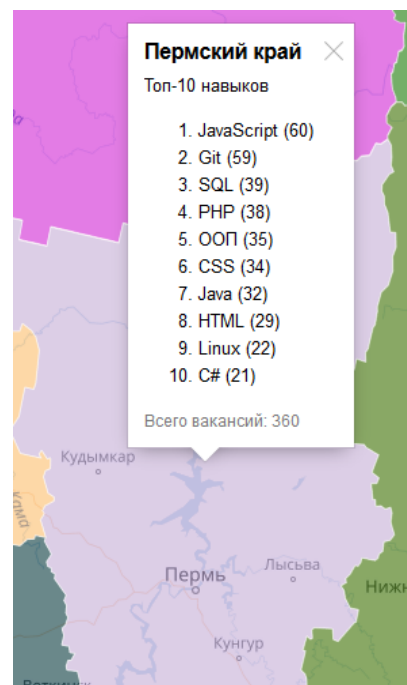


Рис. 2. Фрагмент интерактивной карты, демонстрирующей 10 наиболее часто указываемых в текстах вакансий навыков программистов в Пермском крае на основе данных агрегатора «HH.ru».

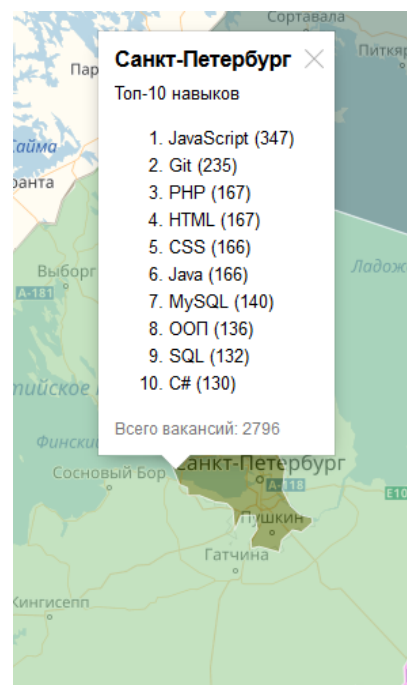


Рис. 3. Фрагмент интерактивной карты, демонстрирующей 10 наиболее часто указываемых в текстах вакансий навыков программистов в Санкт-Петербурге на основе данных агрегатора «HH.ru».

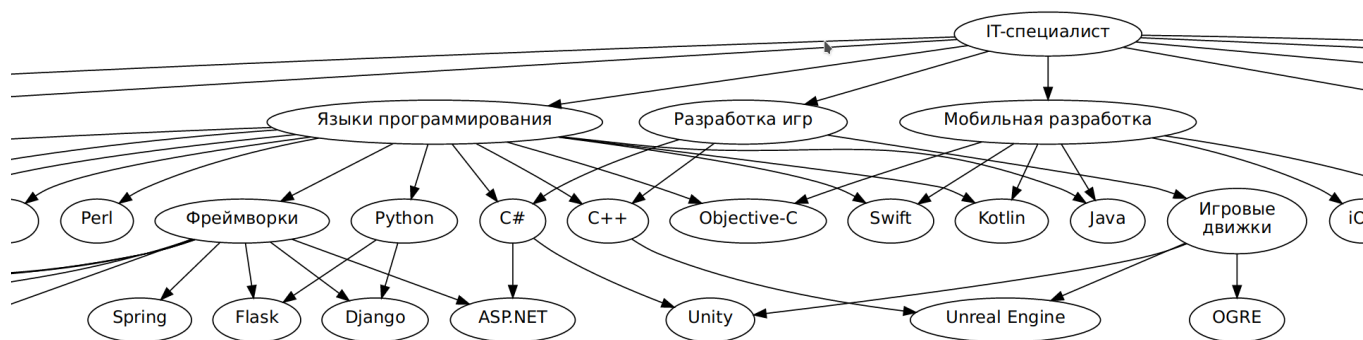


Рис. 4. Фрагмент иерархической структуры, визуализирующей полученную модель.

Данные навыков по каждому региону легли в основу онтологической модели, дополненной на основе знаний о предметной области. Все связи на данном этапе ограничены отношениями «класс–подкласс» (isA).

Для получения упрощенной модели предметной области в рамках пилотного исследования мы воспользовались методом автоматического построению денотатного графа гипертекста, описанным в [11]: базовая иерархия понятий (денотатов) предметной области выстраивалась на основе гиперссылок в текстах статей «Википедии», после чего осуществлялась доработка модели вручную. Итоговая модель была визуализирована в виде иерархической структуры (рис. 4).

IV. РЕЗУЛЬТАТЫ

В результате исследования было выявлено несколько возможных подходов к онтологическому моделированию профессиональных компетенций. Несмотря на наличие успешно функционирующих моделей, ни одна из рассмотренных нами онтологий не затрагивает в полной мере вопрос моделирования потребности рынка труда в определенных специалистах.

Мы предложили, опираясь на понятие навыка, построить подобную онтологическую модель на основе требований, непосредственно указываемых работодателями региона, интегрированных в более широкую модель предметной области.

Нам удалось построить упрощенный онтологический портрет программиста, используя данные из текстов вакансий, опубликованных на сайтах-агрегаторах, и знания о предметной области, извлеченные из веб-энциклопедий. Модель позволяет получить целостный портрет IT-специалиста на основе нескольких понятий.

Следующим шагом в развитии нашей работы станет, очевидно, усложнение иерархии отношений внутри модели для выявления более комплексных

закономерностей и повышении эффективности расчета потребностей экономики в IT-специалистах.

СПИСОК ЛИТЕРАТУРЫ

- [1] План мероприятий по направлению «Кадры и образование» Национальной программы «Цифровая экономика Российской Федерации» // Электронный ресурс. URL: <http://static.government.ru/media/files/k87YsCABuiyuLAjcWDFI LEh6itAirUX0.pdf>, дата обращения: 20.03.2019
- [2] Образование в цифрах: 2018 // Электронный ресурс. URL: <https://www.hse.ru/primarydata/oc2018>, дата обращения: 20.03.2019
- [3] Struder, R., Benjamins, V.R., Fensel, D. Knowledge engineering: principles and methods. Data Knowl. Eng. 1998. Вып. 25 (1). С. 161–197.
- [4] Uschold, M., Gruninger, M., 1996. Ontologies: principles, methods and applications. Knowl. Eng. Rev. 1996. Вып. 11 (2). С. 93–136.
- [5] Gruber, T.R. A translation approach to portable ontology specifications. Knowl. Acquis. 1993. Вып. 5 (2). С. 199–220.
- [6] Trokanas, N., Cecelja, F., Raafat, T. Semantic input/output matching for waste processing in industrial symbiosis. Comput. Chem. Eng. 2014. Вып. 66. С. 259–268.
- [7] О’Лири, Д. Управление корпоративными знаниями // Открытые системы. 1998. №4-5. С. 31-39.
- [8] Мартынов В.В., Филосова Е.И., Ширяев О.В. Организация подготовки и информационная поддержка реализации динамических образовательных программ, учитывающих требования работодателя // Управление экономикой: методы, модели, технологии четырнадцатая международная научная конференция: сборник научных трудов. 2014. С. 110-113.
- [9] Шамсутдинова Т. М. Управление формированием и оценкой профессиональных компетенций в электронных образовательных системах на основе когнитивных моделей // Вестник НГИЭИ. 2017. № 10 (77). С. 35–44.
- [10] Bussemaker M., Trokanas N., Cecelja F. An ontological approach to chemical engineering curriculum development // Computers and Chemical Engineering. 2017. Вып. 106. С. 927–941.
- [11] HeadHunter API: документация и библиотеки // Электронный ресурс. URL: <https://github.com/hhru/api>, дата обращения: 20.03.2019
- [12] Курушин Д.С., Леонов Е.Р., Соболева О.В. О возможном подходе к автоматическому построению денотатного графа гипертекста // Информационная структура текста. Сб.статей // РАН.ИНИОН. М., 2018. С. 113–118