



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

ПРОЕКТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ПОТРЕБНОСТИ В КАДРАХ НА ОСНОВЕ СЕМАНТИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

Ровба Е.А., Бойко В.К., Войтукевич Ю.А., Лявшук В.Е., Петров С.В.

*Гродненский государственный университет имени Янки Купалы,
г. Гродно, Республика Беларусь*

**rovba.ea@gmail.com; 50v.boiko@gmail.com; yury.vait@gmail.com; uliaushuk@gmail.com;
p2004r@gmail.com**

В работе рассмотрен опыт разработки системы поддержки принятия стратегических решений в области управления сферой профессионального образования Республики Беларусь, развитие которой зависит от знания долговременных тенденций спроса и предложения кадров на рынке труда. Разработка системы потребовала интеграции данных, накапливаемых в формализованных и неформальных системах знаний республиканских органов государственного управления, а также картографической информации из открытых сетевых источников. Задача решена на основе онтологического подхода путем построения семантических сетей.

Ключевые слова: онтология; профобразование; общегосударственные классификаторы.

Введение

В последние годы руководство Республики Беларусь (далее – РБ) все чаще отмечает нарастающую необходимость упорядочить и гармонизировать потоки информации об экономическом и социальном развитии регионов и отдельных территорий страны [Указ Президента, 2011]. Начало было положено в Директиве № 2 Президента Республики Беларусь от 27 декабря 2006 г. [Директивы Президента, 2011], которая определила пути улучшения работы с информацией органов госуправления и наделила особыми полномочиями по сбору статистической информации Национальный статистический комитет РБ (Белстат). Это позволило обеспечить сбор, обработку и хранение достоверной статинформации в одном месте и одновременно обеспечить доступ к ней заинтересованных структур госуправления и субъектов хозяйствования. Однако возможность доступа к информации является необходимым, но недостаточным условием для ее использования во благо экономического развития.

Министерство образования РБ систематически накапливает значения показателей и индикаторов, о состоянии процессов сферы образования. Но за большим количеством разноформатных данных организованных в таблицы по учреждениям образования и регионам тяжело увидеть полноту

картины развития системы образования на всех уровнях, тем более в динамике за несколько лет. Когда в условиях оптимизации госрасходов актуализировался вопрос повышения отдачи от сферы образования, то наиболее остро он встал перед системой профобразования, тесно связанной с реальным сектором экономики. Поэтому министерство инициировало разработку системы поддержки принятия стратегических решений на основе понимания своего потенциала подготовки кадров рабочих специальностей и информации о тенденциях процессов на региональных рынках труда, которая была бы привязана географически, статистически достоверна, визуально понятна и отражала сложность и взаимосвязь протекающих процессов. Авторы данной статьи разработали реализацию этой системы в рамках выполнения научно-исследовательского проекта «Разработка средств пространственно-временного моделирования для оценки и прогнозирования развития системы профессионального образования Республики Беларусь» (Кадастр профобразования). Перед проектом стояло две задачи. Во-первых, интегрировать информационные ресурсы, как накопленные, так и непрерывно поступающие на верхний уровень управления, для пространственно-временного моделирования, оценки и прогнозирования развития системы профобразования РБ. Во-вторых, построить многомерную модель данных, поступающих от учреждений образования с их локализацией на географической карте.

1. Этапы решения задачи

Первым этапом работы стала инвентаризация данных о деятельности учебных заведений профобразования; о потребности регионов в рабочих специальностях, их демографической и экономической ситуации. Второй этап состоял в формировании платформы интеграции данных системы профобразования из программных продуктов для формирования и сопровождения онтологий RDF (OWL), для интеграции в формат RDF разнородных данных, а также продуктов для хранения и обработки запросов к данным формата RDF(OWL). Третьим этапом стала разработка онтологий предметных областей, то есть семантическое описание сведений, распределенных в информационных базах системы профобразования, и понятий, описывающих деятельность системы профессионального образования на разных уровнях с учетом Международной стандартной классификации образования (МСКО 2011). На базе разработанных онтологий на четвертом этапе была реализована интеграция данных о географической привязке всех источников данных системы профобразования, с данными госстатистики о демографии и о реальном секторе экономики. Пятый этап заключался в организации хранения данных формата RDF(OWL) и реализации SPARQL-запросов о деятельности учебных заведений, со стороны OLAP-системы пользователя. На шестом этапе при помощи многомерных алгоритмов анализа, используемых в рамках методологии HASTD, были выявлены границы естественных группировок данных, связанных с потребностями рынка труда и возможностями их удовлетворения системой профессионального образования. На последнем этапе полученные результаты были объединены в виде информационной системы и электронных динамических карт-справочников (в части касающейся системы профессионального образования и ее связи с запросами рынка труда) на платформах MapServer и Google Earth для пространственно-временного моделирования, оценки и прогнозирования развития системы профессионального образования РБ. В качестве исходных данных для построения семантических сетей была отобрана информация из Общегосударственных классификаторов РБ, которая актуализируется Министерством экономики РБ, Национальным статистическим комитетом РБ, Министерством образования РБ; из единых квалификационных справочников (ЕКСД и ЕТКС) Министерства труда и соцзащиты РБ, международных классификаторов в области образования, а также сложившейся системы показателей оценки деятельности учреждений профобразования. Были проанализированы структура и содержание следующих документов:

- Единый квалификационный справочник должностей служащих (ЕКСД) (Постановление Минтруда и соцзащиты РБ № 32 от 30.03.2004);

- Единый тарифно-квалификационный справочник работ и профессий рабочих (ЕТКС) (постановление Минтруда и соцзащиты РБ от 30.03.2004 № 34);

- Международная стандартная классификация образования (МСКО-97, МСКО 2011 и МСКО 2013);

- Международная стандартная классификация образования и подготовки МСКО-Ф;

- Общегосударственные классификаторы ОКРБ 005-2006 и ОКРБ 005-2011 «Виды экономической деятельности» (ОКЭД);

- Общегосударственный классификатор ОКРБ 006-2009 «Профессии рабочих и должности служащих» (ОКПД);

- Общегосударственный классификатор ОКРБ 011-2009 «Специальности и квалификации» (ОКСК);

- Общегосударственный классификатор ОКРБ 014-2007 «Занятия» (ОКЗ).

Анализ представленной в сборниках статистической информации выявил проблемные места в структуре данных об экономическом развитии республики. Наиболее значимым, с точки зрения авторов настоящей статьи, является скачкообразное изменение структуры учета в разрезе видов экономической деятельности. Это связано с состоявшимся 01.01.2011 г. переходом органов статистики к формированию и предоставлению официальной статистической информации на основе ОКЭД.

ОКЭД был разработан на базе классификации видов экономической деятельности Европейского союза с введением дополнительного уровня классификации, учитывающего особенности белорусской экономики, и заменил действовавший до него Общесоюзный классификатор отраслей народного хозяйства (ОКОНХ). Переход с модели ОКОНХ на модель ОКЭД продолжается до сих пор, так как существует большое количество действующих документов республиканских органов госуправления, до сих пор использующих классификацию ОКОНХ. Для учета этого в разрабатываемой онтологии были использованы таблицы с переходными ключами не только из кодов одной модели в другую, но также из кодов ОКЭД версии 2011 г. в ОКЭД версии 2015. Обе таблицы переходных ключей переведены в цифровой формат для интеграции в разрабатываемую онтологию.

Перечисленные выше классификаторы представляют собой графовые базы данных, в некоторой степени связанные между собой. Однако формат их представления (.pdf, .doc) не позволяет напрямую сформировать из них базу данных проекта. Для приведения в требуемый формат использовался программный продукт Open Refine, с помощью которого содержание классификаторов денормализовано в плоские электронные таблицы, подготовленные к привязке в разрабатываемую онтологию. Выделенные при денормализации понятия включены в семантическую сеть,

описывающую первую версию целевой онтологии системы профобразования РБ.

В ряде случаев стандартных средств пакета Open Refine оказалось недостаточно и тогда была написана утилита для обработки сложных ситуаций «Парсер-комбинатор для преобразования текстового файла классификатора в xml-описание». Выделенные при денормализации понятия добавлены в семантическую сеть, описывающую первую версию целевой онтологии системы профобразования РБ.

При составлении семантической сети использовались возможности утилиты Grafviz [Электронный ресурс] – Режим доступа: <http://www.graphviz.org/>. Наглядность, достигаемая при использовании утилиты Grafviz, ограничена тем, что, во-первых, многомерный граф проецируется на плоскость, а, во-вторых, при большом количестве узлов сама плоскость вырастает до невообразимых размеров. Поэтому при построении онтологий понятий, содержащихся в классификаторах, представляющих собой документы размером более 250 страниц, использовался программный продукт Protege [Электронный ресурс] – Режим доступа : <http://protege.stanford.edu/>. Предоставленные в Protégé возможности визуализации упрощают работу по поддержанию классификаторов в актуальном (достоверном) состоянии, поиску совпадений и опечаток в кодировке, а также позволяют развивать семантическую сеть, дополняя ее новыми понятиями и их системами.

На базе информации и разработанных онтологий была построена семантическая сеть, описывающая потоки человеческих ресурсов между системой профессионального образования (ССУЗы и ПТО) и отраслями экономики РБ с проекцией на географию

регионов и демографическую ситуацию. Моделирование осуществлялось в среде программного продукта Protégé с помощью модуля Graphviz. Структура сети имеет сложное строение и для удобства она разбита на четыре модуля.

Первый онтологический модуль «Экономика» отражал связи и отношения между видами экономической деятельности, отраслями экономики, должностями, профессиями и занятиями в РБ, описывая иерархию связей между понятиями, определенными в справочниках ЕТКС и ЕКСД, а также классификаторах ОКОНХ, ОКЭД (версии ОКРБ 005-2006 и ОКРБ 005-2011), ОКПД и ОКЗ. Данный модуль необходим, ибо классификация отраслей экономики РБ неоднократно менялась, что приводило и к изменениям в составе показателей собранной и подвергаемой обработке информации.

Второй онтологический модуль «География» представлял собой проекции учебных заведений и видов экономической деятельности на территориально-административную структуру РБ. С его помощью произведена визуализация распределения на карте РБ учреждений системы профобразования, показателей экономического потенциала регионов и их демографической ситуации. В качестве источника картографической информации использовались данные открытых сервисов OSM (OpenStreetMap), Яндекс.Карты, Google Карты. Границы административных регионов получены из бесплатного (для некоммерческого использования) источника Global Administrative Areas (GADM). В этой базе доступны шейп-файлы трех уровней пространственного разрешения – на уровне страны, областей и районов. Хотя данные не имеют статуса официальных, но для целей проекта их точность и полнота признаны удовлетворительными.

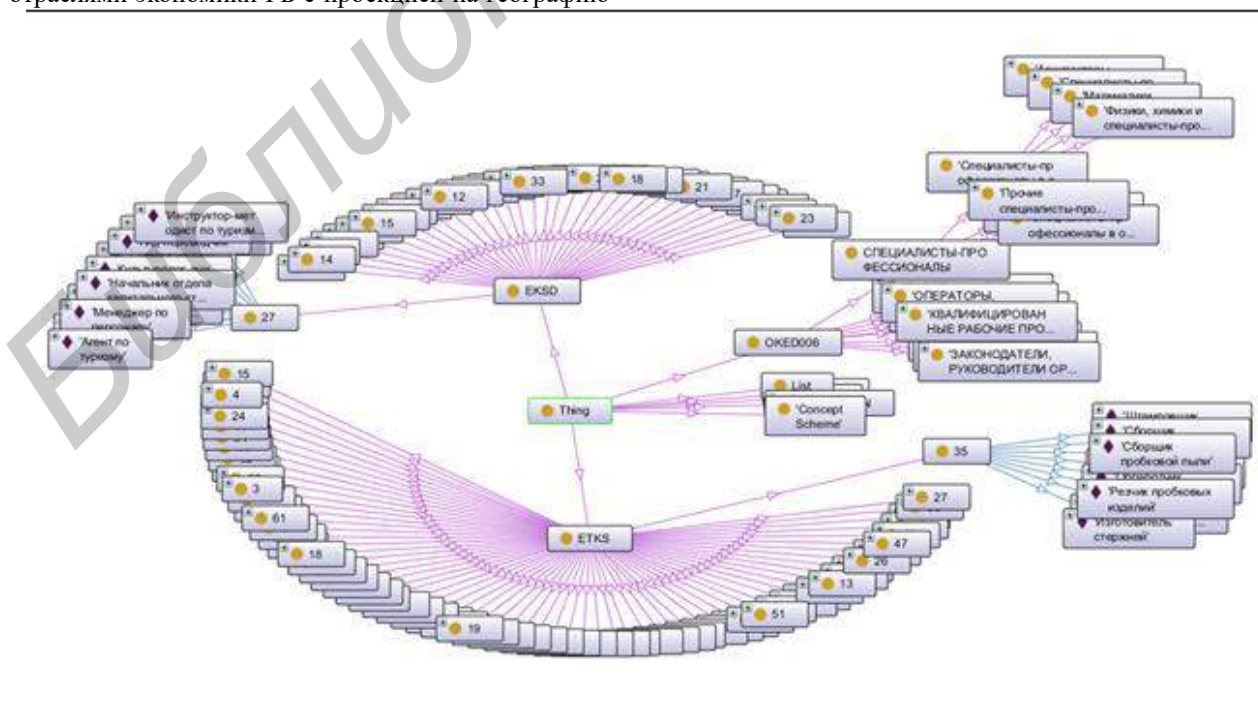


Рисунок 1 -- Граф онтологии, включающей ЕКСД, ЕТКС и ОКПД, выполненный в Protege

Третий онтологический модуль «Система профобразования» был проекцией перечня специальностей подготовки специалистов на виды экономической деятельности и региональную сеть учреждений профессионального образования. Он содержал распределение специальностей системы профессионального образования в соответствии с ОКСК по учреждениям образования, их юридические адреса для геопривязки, и историю изменения их изменений и ведомственной принадлежности, необходимую для корректной обработки информации.

Четвертый онтологический «Модуль данных» был проекцией динамических показателей экономической деятельности, демографической ситуации, результатов образовательной деятельности учебных заведений, возрастнополовой структуры и образовательного уровня трудоспособного населения на отрасли экономики и территориально-административную структуру РБ. Он содержал информацию о наборах и выпусках учебных заведений по специальностям, информацию о распределении видов экономической деятельности по регионам, о кадровом обеспечении отраслей, демографической ситуации в регионах.

Разработка перечисленных выше модулей семантической сети велась в среде программного продукта Protégé. Объединение модулей в единую онтологию выполнялось в среде программного продукта KARMA [Электронный ресурс] – Режим доступа :<http://www.isi.edu/integration/karma/>.

Заключение

В результате построения семантических сетей выявились информационные лакуны и нестыковки онтологий, лежащих в основе официальных государственных документов. Так, в комплексе экономических классификаторов проявилось отсутствие явной связи между описанием занятий работников на рабочих местах и видами экономической деятельности, реализуемой субъектами хозяйствования. Направления образования, задаваемые классификатором специальностей ОКРБ 011-2009, в ряде случаев опирались на термины из классификатора профессий ОКРБ 006-1996. Кроме этих проблемных мест построение онтологий позволило выявить недостаток практикуемого в настоящее время учета трудовых ресурсов. Он ведется по «условным поколениям» (возрастным группам), а не по реальным поколениям, что приводит к несопоставимости данных и проблемам оценки движения трудовых ресурсов. В результате в органах статистики отсутствует полнота информации о возрастном составе занятого населения и его квалификации. Перечисленные проблемные зоны, выявленные на этапе построения онтологий, с одной стороны усложнили разработку методик анализа и прогнозирования потребностей экономики в кадрах, но с другой стороны побудили к разработке ряда моделей, показывающих

относительное распределение трудовых ресурсов по регионам и отраслям экономики.

Тем не менее, среди рекомендаций, предложенных по итогам выполнения описываемого научно-исследовательского проекта, ключевыми были названы три предложения, сформулированные в процессе создания семантических сетей: разработать единую образовательную онтологию РБ, изменить систему учета трудовых ресурсов, перейдя на учет по дате рождения и квалификации (специальности) и обеспечить согласованную взаимосвязь национальных классификаторов.

Библиографический список

[Указ Президента, 2011] О некоторых вопросах сбора информации, не содержащейся в государственной статистической отчетности: Указ Президента Респ. Беларусь, 2 марта 2011 г., № 95 // Эталон – Беларусь [Электронный ресурс] / Нац. центр правовой информ. Респ. Беларусь. – Минск, 2015.

[Директивы Президента, 2011] Директивы Президента Республики Беларусь: сборник нормативных правовых актов. – Минск: Национальный центр правовой информации Республики Беларусь, 2011. – 38 с.

DESIGN OF INTELLECTUAL SYSTEM TO FORECAST STAFFING NEEDS BASED ON SEMANTIC KNOWLEDGE REPRESENTATION

Y.A. Rouba, V.K. Boika, Y.A. Vaitukevich,
U.J. Liaushuk, S.V. Piatrou,

*Yanka Kupala State University of Grodno,
Grodno, Belarus*

rovba.ea@gmail.com;
50v.boiko@gmail.com;
yury.vait@gmail.com;
uliaushuk@gmail.com;
p2004r@gmail.com

The paper describes the experience of designing of strategic decision support system in the field of vocational education of the Republic of Belarus. The system development depends on knowledge of long-term trends in staff supply and demand in the labor market. The design of the system required the integration of data accumulated in the formalized and informal knowledge systems of republican state bodies, as well as cartographic information from public online sources. The problem has been solved on the basis of ontological approach by building semantic networks.