

РАЗРАБОТКА МОДЕЛИ КОМПЕТЕНЦИЙ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ С ИСПОЛЬЗОВАНИЕМ ОНТОЛОГИЙ И БАЗ ЗНАНИЙ

Тимофеев А.Н.

Восточно- Сибирский государственный университет технологий и управления,
Улан-Удэ, Россия
89021632777@mail.ru

Аннотация. В статье рассматривается подход к построению модели компетенций в сфере информационных технологий. Анализ моделей компетенций, баз знаний, онтологий и алгоритмов автоматического построения онтологий показывает актуальность создания модели, с использованием автоматически дополняемой онтологии, наполняемой из баз знаний общего назначения.

Ключевые слова: модель компетенций, онтология, база знаний, Klink-2, информационные технологии.

Введение

Динамичное развитие информационных технологий, характеризующееся ускоряющейся сменой ведущих парадигм и трендов, повышением сложности процессов и систем и усилением специализации, обуславливает изменение требований к профессиональной подготовке ИТ-специалистов. Концептуально эти изменения можно охарактеризовать как актуализацию проблемы непрерывности образовательного процесса, его направленности и содержания на всех уровнях существующей образовательной системы [1] с одновременным переходом к компетентностно-ориентированной модели профессионального образования [2].

В условиях цифровизации актуализируется вопрос формализации методов построения компетентностных моделей. В связи с этим целью настоящей статьи является описание метода автоматизированного построения и использования модели компетенций с использованием онтологического подхода и баз знаний.

Для достижения поставленной цели необходимо рассмотреть существующие модели компетенций, онтологии и базы знаний. На основе проведенного анализа выработать методы формализации модели компетенций в виде онтологии, обогащения онтологии информацией из баз знаний и автоматизированного наполнения и актуализации полученной онтологической модели компетенций.

1. Модели компетенций, онтологии и базы знаний

За рубежом в экономически развитых государствах образовательные и профессиональные стандарты создаются на международном или государственном уровне, а также отдельными организациями и компаниями. Они могут описывать как требования к компетенциям, так и подход к «управлению талантами» [3]. Среди них можно выделить Computer Science Curricula, European Qualification Framework (EQF), European Competency Framework (E-CF), SFIA (Skills Framework for the Information Age), The Skills Framework for ICT, IBM Kenexa Talent Frameworks

В России разработаны Федеральные государственные образовательные стандарты (ФГОС), установленные для разных уровней образования, включая общее и профессиональное образование. Минтрудом России с середины 2010-х годов создано более 30 профессиональных стандартов в сфере связи, информационных и коммуникационных технологий.

Существующие модели компетенций характеризуются разнообразным подходом к систематизации изложенной в них информации и использованием формализмов, позволяющих структурировать эту информацию.

Указанный подход к построению модели хорошо описывает общие подходы, однако для выполнения конкретных трудовых функций требуется серьезное расширение объема понятий, связанных семиотическими и иерархическими отношениями.

Онтологии используются для описания онтологических обязательств для набора агентов, чтобы они могли общаться в области дискурса, не обязательно оперируя общепринятой теорией. Говорят, что агент совершает действия в онтологии, если его наблюдаемые действия согласуются с определениями в онтологии. Идея онтологических обязательств основана на предложенной А. Ньэллом [4] концепции уровня знаний. Уровень знаний — это уровень описания знаний агента, который не зависит от представления на уровне символов, используемого внутри агента.

Существуют как специализированные онтологии и базы знаний, так и онтологии общего назначения. Среди них можно выделить OWL-S, DOLCE (Descriptive Ontology for Linguistic and

Cognitive Engineering), Сус, Suggested Upper Merged Ontology (SUMO), General Formal Ontology (GFO), Computer Science Ontology (CSO).

Существующие онтологии достаточно полно описывают многие предметные области, в том числе область информационных технологий. Проработаны ручные, автоматические и комбинированные методы пополнения онтологий.

Ориентированность онтологий на использование механизмов логического вывода и достаточно общий характер определяют то, что в них слабо описаны функциональные отношения, задающие объем компетенций для конкретных ролей.

Согласно международному стандарту ISO/IEC/IEEE 24765-2010, Systems and software engineering — Vocabulary база знаний – это база данных, содержащая правила вывода и информацию о человеческом опыте и знаниях в некоторой предметной области. Среди баз знаний можно выделить DBpedia, YAGO, Google Knowledge Graph, Microsoft Concept Graph, ConceptNet, WordNet

Базы знаний являются важным источником структурированной и систематизированной информации. Активно пополняемые базы знаний могут служить источником данных для выявления понятий, уточняющих абстрактные понятия онтологий верхнего уровня и моделей компетенций.

2. Текущее состояние онтологического подхода

Онтологический подход к созданию и совершенствованию компетентностно-ориентированных моделей образования, в том числе в сфере информационных технологий, находит широкий отклик у исследователей.

Е.Ю. Благоев, И.А. Лещева, С.А. Щербан описывают алгоритм формирования «дополнительных» компетенций, на основе которого ими предложена усовершенствованная онтологическая модель компетентностно-ориентированного учебного плана.

Д. С. Конькова, Л. В. Курзаева, Ю. С. Лактионова, С. А. Чичиланова в своей работе описывают процесс построения онтологической фреймовой модели с использованием системы Protege на примере профессионального стандарта системного аналитика.

М. С. Гаспариан, С. А. Лебедев, Ю. Ф. Тельнов в своей статье рассматривают вопросы инжиниринга образовательных программ с использованием семантического моделирования информационно-образовательного пространства.

Е. Katis, Н. Kondylakis, G. Agathangelos и V. Kostas [5] рассматривают концептуализации образовательных структур знаний в академической среде и представляют методологию и процесс разработки образовательной онтологии.

N. Hubert, A. Brun и D. Monticolo [6] в своей работе создали онтологию EducOnto, которая направлена на моделирование университетских учебных планов и профилей студентов, и граф знаний EduKG.

M. Dascalu и соавторы описывают практики использования онтологий при разработке учебных программ (например, Болонской онтологии, онтологии учебных программ BBC или онтологии CCSO), и предлагают онтологическую модель разработки учебных программ в области инженерного образования.

L. Cassel G. Davies, R. Leblanc, L. Snyder, T. Heikki [7] в своей статье рассказывают о создании онтологии, описывающей сферу информационных технологий. Одной из целей разработки данной онтологии ставится ее использование в разработке учебных программ.

A. Salatino, T. Thanapalasingam, A. Mannocci, A. Birukou, F. Osborne, E. Motta [8] описывают онтологию Computer Science Ontology (CSO), варианты ее использования и историю развития. Это крупномасштабная, автоматически генерируемая онтология областей исследований, который включает в себя около 14 тысяч тем и 162 тысячи семантических связей. CSO создана с применением алгоритма Klink-2 на наборе данных из 16 млн. научных статей.

Алгоритм Klink-2 предложен F. Osborne и E. Motta [9] как развитие алгоритма Klink. Klink-2 принимает на вход набор научных метаданных: ключевых слов, их статистических или семантических связей с различными объектами, включая исследовательские работы, места проведения, авторов и организации. Результатом работы алгоритма является созданная или дополненная онтология OWL, описывающая три вида семантических отношений между темами исследования, определенными из набора ключевых слов, и другими входными данными.

Алгоритм Klink-2 относится к группе методов автоматического построения таксономий и онтологий. Среди сходных методов можно выделить естественно-языковые систему Text2Onto и подход Lexico-Syntactic Pattern Extraction (LSPE), OntoLearn, использующий иерархическую

агломеративную кластеризацию фреймворк TaxGen, такие статистические алгоритмы, как метод подстановки [10], GrowBag, PageRank, OntoLearn, алгоритмы OntoGen и OntoText.

Проведенный обзор показывает существенные достижения онтологического подхода к моделям компетенций и позволяет сформулировать некоторые проблемы в онтологическом подходе к компетентностным моделям и возможные пути их решения.

3. Выводы

Концептуально описывающие модели компетенций онтологий как правило являются оригинальными разработками, что влечет сложность их сопряжения с крупными онтологиями верхнего уровня и базами знаний общего назначения.

Формирование большинства онтологий происходит ручным способом с использованием эвристик, без детальной проработки вопросов автоматического построения.

С содержательной точки зрения онтологии описывают предметную область «Образование», некоторые сегменты предметной области «Информационные технологии» и особенности конкретного стандарта или подхода.

Использование баз знаний в моделях компетенций в некоторой степени проработано в различных исследованиях и применяется на практике. В отношении российских трудовых и образовательных стандартов использование открытых баз знаний описано слабо.

Поскольку поставленные выше проблемы и вопросы в различной степени решены за рубежом, актуальным является их решение применительно к отечественным профессиональным и образовательным стандартам таким как ФГОС и профессиональные стандарты Минтруда России. Для этого предлагается модель, которая позволяет гармонизировать требования стандартов и соотнести концепты сегментов специализированной онтологии с терминами широко применяемых баз знаний общего назначения.

Специализированная онтология связывается с общими онтологиями через отношение эквиваленции и с открытыми базами знаний (например DBPedia, Wikidata) отношениями «часть-целое» и «абстрактное-конкретное». Выявление данных отношений позволит определить объем понятий онтологии. Определение объема понятий придает модели практический характер, поскольку абстрактное «знание», «умение» или «навык» реализуется путем применения конкретных инструментов, разнообразие которых неуклонно растет [11].

Установление связей определяет возможность автоматического наполнения нужных сегментов базы знаний модели путем обработки открытых баз знаний, а также извлечения зависимостей на основе открытых репозиториях программного обеспечения.

4. Предлагаемая модель

Используя базы знаний и онтологии модель компетенций должна включать в себя концептуализацию требований к компетенциям, то есть онтологию, описывающую один из стандартов, рассмотренных выше. В модели должны быть определены источники дополнения онтологии и получения знаний предметной области, такими источниками являются базы знаний. Для этого должен быть определен алгоритм наполнения. Для достижения целей создания модели должен быть определен алгоритм верификации профессиональных компетенций с использованием, построенных онтологий и базы знаний. Таким образом предлагаемая модель имеет следующий вид верхнего уровня:

$$M = \langle S, O, B, A, V \rangle \quad (1)$$

S – модель компетенций, стандарт

O – онтология, соответствующая S , $O = \{o^S\} \cup \{o_1, o_2, \dots, o_n\} \cup \{o_1 \rightarrow o^S, o_2 \rightarrow o^S, \dots, o_n \rightarrow o^S\}$, где o^S – специализированная онтология на основе стандарта S , o_i – общая онтология, $o_i \rightarrow o^S$ – отношения эквиваленции понятий онтологии общей онтологии o_i и специализированной онтологии o^S .

B – база знаний в сфере ИТ, $B = \{b^S\} \cup \{b_1, b_2, \dots, b_n\} \cup \{b_1 \rightarrow b^S, b_2 \rightarrow b^S, \dots, b_n \rightarrow b^S\} \cup \{b_1 \rightarrow o_n, b_2 \rightarrow o_n, \dots, b_n \rightarrow o_n\}$, где b^S – специализированная база знаний модели на основе стандарта S , b_i – открытая база знаний, $b_i \rightarrow b^S$ – отношения элементов открытой базы знаний b_i и специализированной базы знаний b^S .

A – алгоритм пополнения базы знаний. Алгоритм пополнения заключается в том, что для каждого нового элемента базы знаний B устанавливается отношение R хотя бы с одним понятием онтологии O . Выявление отношений R производится двумя способами: вручную и с помощью адаптированного алгоритма Klink-2, принимающего на вход элементы $b_i \in B$, а на выходе дополняющие онтологию O

отношениями `skos:broaderGeneric` («убедительная подобласть»), `contributesTo` («полезно рассматривать»), `relatedEquivalent` («рассматриваются как эквиваленты»). Алгоритм включает следующие шаги:

- Каждая пара b_i и b_j количество общих связей которых с другими $b_k \cup o_t, b_k \in B, o_t \in O$ превышает пороговое значение p , анализируется на предмет возможности установления иерархических отношений между b_i и b_j , если это так, то устанавливаются отношения `skos:broaderGeneric` и `contributesTo`.
- Каждое b_i анализируется на предмет наличия множественности значений. Элементы b_i , которые кажутся неоднозначными, разбиваются на элементы $\{b_{1i}^1, b_{1i}^2, \dots, b_{1i}^n\}$ с уникальными значениями, которые затем сравниваются с другими b_i .
- Похожие b_i и b_j объединяются в новое b_k , а между b_i и b_j устанавливается отношение `relatedEquivalent`.
- Шаги 2 и 3 повторяются до тех пор, пока ни одного b_i не будут разделено или объединены. Затем удаляются циклы.

Для каждой пары b_i и b_j рассчитываются три метрики: $H_R(x, y)$, использующая семантический вариант метода подстановки, $T_R(x, y)$, использующая информацию о связи во времени и $S_R(x, y)$, оценивающая сходство. $H_R(x, y)$ и $T_R(x, y)$ используются для вывода отношений `skos:broaderGeneric` и `contributesTo`, $S_R(x, y)$ используется для вывода отношения `relatedEquivalent`. Эти метрики вычисляются для каждого семантического или статистического отношения R , связывающего b_i и b_j с элементами $b_k \cup o_t, b_k \in B, o_t \in O$.

Метрика $H_R(x, y)$ вычисляется по следующей формуле:

$$H_R(x, y) = \left(\frac{I_R(x, y)}{I_R(x, x)} - \frac{I_R(y, x)}{I_R(y, y)} \right) \times c_R(x, y) \times n(x, y). \quad (2)$$

Если отношение R является неколичественным, то $I_R(x, y)$ – это количество, связанных с x, y отношением R одновременно. Если отношение R является количественным, то $I_R(x, y)$ – это сумма минимальных значений, квантифицирующих отношения связанные с x, y .

$c_R(x, y)$ измеряет семантическое сходство x и y и вычисляется как коэффициент Отиаи для двух векторов, в которых каждый элемент представляет ключевое слово, которое имеет общие отношения с x и y .

$n(x, y)$ определяет сходство между двумя ключевыми словами и вычисляется как линейная комбинация строковых метрик, основанных на поиске общих подстрок, проценте одинаковых слов, количестве общих символов, наличии общих аббревиатур и т.д.

Если $H_R(b_i, b_j) \geq t_R$, то b_i является гипонимом b_j , если $H_R(b_i, b_j) \leq -t_R$ то b_i является гиперонимом b_j .

Метрика $T_R(x, y)$ – является вариацией метрики $H_R(x, y)$, в которой $I_R(x, y)$ рассчитывается путем взвешивания количества и интенсивности связей в каждом году в соответствии с длительностью от года появления. Вес вычисляется как $w(year, x) = (year - debut(x) + 1)^{-\gamma}$. Необходимо заметить, что расчет данной метрики затруднителен для баз знаний B .

Метрика $S_R(x, y)$ вычисляется по следующей формуле:

$$S_R(x, y) = \frac{c_R(x, y)}{\max(c_R^{super}(x, y), c_R^{sib}(x, y)) + 1} \quad (3)$$

V – алгоритм верификации индивидуальных наборов профессиональных компетенций и построения траектории профессионального развития.

5. Заключение

Предложенная адаптивная модель компетенций в сфере информационных технологий позволит проводить формальную верификацию индивидуальных наборов профессиональных компетенций, строить траектории профессионального развития с возможностью выбора альтернативных путей. Верификация наборов профессиональных компетенций поможет упростить выбор специалистами направления своего развития, в частности рекомендуя конкретные инструменты на основе схожести или отнесения их к абстрактным понятиям верхнего уровня.

Модель предполагает автоматизированное дополнение используемой в ней онтологии и базы знаний. Автоматизированное дополнение онтологии и базы знаний позволит поддерживать модель в

актуальном состоянии, что важно в условиях интенсивного развития области информационных технологий.

Предметом дальнейшего совершенствования является алгоритм автоматического дополнения модели, который должен учитывать возможность использования различных структурированных источников данных.

Литература

1. *Пушкарев, Ю. В.* Непрерывное образование в современных условиях: основные концептуальные подходы / Ю. В. Пушкарев, Е.А. Пушкарева // Вестник Новосибирского государственного педагогического университета. – 2015. – № 6(28). – С. 161-171..
2. Подходы к реализации онтологии компетенций и профессиональной подготовки IT-специалистов / Т.А. Ткалич, Е.Г. Гриневич // URL: <https://elib.bsu.by/bitstream/123456789/284425/1/233-238.pdf> (дата обращения 18.03.2023).
3. *Завьялова, Н. Б.* Автоматизация управления талантами в повышении эффективности компании / Н. Б. Завьялова, А.А. Киреева // Экономика, предпринимательство и право. – 2020. – Т. 10, № 5. – С. 1413-1424.
4. The Knowledge Level / A. Newell // URL: <http://cs.uns.edu.ar/~grs/InteligenciaArtificial/Allen%20Newell%20-%20The%20knowledge%20level.pdf> (дата обращения 02.05.2023).
5. Developing an Ontology for Curriculum & Syllabus / E. Katis, H. Kondylakis, A. Giannis, K. Vassilakis. // URL: https://www.researchgate.net/publication/326294216_Developing_an_Ontology_for_Curriculum_Syllabus (дата обращения 12.04.2023).
6. *Hubert N., Armelle B., Monticolo D.* New Ontology and Knowledge Graph for University Curriculum Recommendation. International Workshop on the Semantic Web, 2022.
7. Using a computing ontology as a foundation for curriculum development / L. Cassel, G. Davies, R. Leblanc, L. Snyder, T. Heikki // URL: https://www.researchgate.net/publication/228941636_Using_a_computing_ontology_as_a_foundation_for_curriculum_development%27 (дата обращения 12.04.2023)
8. The Computer Science Ontology: A Comprehensive Automatically-Generated Taxonomy of Research Areas. Data Intelligence / A. Salatino, T. Thanapalasingam, A. Mannocci, A. Birukou, F. Osborne, E. Motta // URL: https://www.researchgate.net/publication/345356286_The_Computer_Science_Ontology_A_Comprehensive_Automatically-Generated_Taxonomy_of_Research_Areas (дата обращения 12.04.2023)
9. Klink-2: Integrating Multiple Web Sources to Generate Semantic Topic Networks. / F. Osborne, E. Motta // URL: https://www.researchgate.net/publication/281650037_Klink-2_Integrating_Multiple_Web_Sources_to_Generate_Semantic_Topic_Networks (дата обращения 12.04.2023.)
10. *Sanderson M.*, Deriving concept hierarchies from text / M. Sanderson, B. Croft // In Proceedings of the SIGIR conference, pp. 206–213, 1999
11. *Тимофеев А. Н.* Семантическое моделирование: обзор процессов, инструментов, методов и знаний предметной области разработки программного обеспечения (Часть 1) / А.Н. Тимофеев, И.С. Евдокимова, Н.Б. Хаптахаева, А.А. Сенотрусова // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. – 2022. – № 12. – С. 85-91.