

УДК 004.82

*Ирина Посадская***АРХИТЕКТУРА БАЗЫ ЗНАНИЙ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ОБУЧЕНИЯ***Ирина Посадська***АРХІТЕКТУРА БАЗИ ЗНАНЬ АВТОМАТИЗОВАНОЇ СИСТЕМИ НАВЧАННЯ***Iryna Posadska***KNOWLEDGE BASE ARCHITECTURE OF AUTOMATED EDUCATIONAL SYSTEM**

В статье рассматриваются достоинства и недостатки разных видов представлений знаний в автоматизированных системах обучения. Сравнительная характеристика видов представлений знаний была получена на базе экспертных оценок. Предложена и обоснована архитектура базы знаний знание-ориентированной автоматизированной системы обучения.

Ключевые слова: автоматизированная система обучения, знание, база знаний, формализация, предметная область.

Рис.: 1. Табл.: 1. Библ.: 10.

У статті розглянуто переваги і недоліки різних видів представлень знань в автоматизованих системах навчання. Порівняльна характеристика видів представлень знань була отримана на базі експертних оцінок. Запропоновано та обґрунтовано архітектуру бази знань знаннеорієнтованої автоматизованої системи навчання.

Ключові слова: автоматизована система навчання, знання, база знань, формалізація, предметна галузь.

Рис.: 1. Табл.: 1. Бібл.: 10.

The advantages and disadvantages of different types of knowledge representation in automated learning system were considered in this article. Comparative characteristics of the types of knowledge representations were obtained on the basis of expert estimations. Knowledge base architecture of the knowledge-based automated learning system was proposed and substantiated.

Key words: automated learning system, knowledge, knowledge base, formalization, domain.

Fig.: 1. Tabl.: 1. Bibl.: 10.

Постановка проблемы. Автоматизированные системы обучения (АСО) предназначены для автоматизации подготовки специалистов с участием или без участия преподавателя (тьютора), обеспечивают процесс обучение, подготовку учебных материалов, управление процессом обучения и контроль его результатов. Одной из разновидностей АСО являются системы, базирующиеся на знаниях (АСОз). Основными пользователями данных систем являются ученик, тьютор, эксперт в предметной области и инженер по знаниям [1]. База знаний является ядром АСОз, поэтому иногда такие системы называют знание-ориентированными [2].

База знаний выступает не только как «накопитель», а и как механизм управления между функциональными модулями АСОз; между самой системой и знаниями. При проектировании таких систем важнейшей задачей является создание такой архитектуры базы знаний, которая будет обладать возможностями работы с различными видами представлений знаний, при этом сохраняя их в актуальном виде.

Анализ последних исследований и публикаций. Согласно взглядам Ю. И. Клыкова, знания – это совокупность данных, фактов и правил вывода о мире, которые включают в себя информацию о свойствах объектов, закономерностях процессов, явлений, а также правил, которые используются этой информацией для принятия решений [3].

Знание в системах обучения рассматривается как знание о предмете (предметной области учебного курса) и определяется как уверенное понимание предмета, умение самостоятельно обращаться и разбираться в нем, а также использовать его для достижения намеченных целей [4].

База знаний (БЗ) – это совокупность средств, которые обеспечивают хранение, поиск и преобразование знаний.

Среди широкого спектра представления знаний о предметной области курса существуют естественно-языковые представления учебного курса, формализованные представления фрагментов предметной области в виде диаграмм UML [5], формализованные представления, базирующиеся на определенном исчислении, таком как исчисление

предикатов 1-го порядка [6], фреймы [7], производственные правила [8] и др. Однако каждый из видов имеет как свои достоинства, так и свои недостатки.

Целью статьи является описание и обоснование архитектуры базы знаний для знание-ориентированной автоматизированной системы обучения.

Изложение основного материала. Рассмотрим достоинства и недостатки каждого из указанных выше видов представлений знаний. **Естественно-языковые представления** ориентированы на человека (в плане ролей – на тьютора, студента, эксперта в предметной области). Текст учебного курса разбивается на разделы, параграфы, подпараграфы и т. д. Навигация между отдельными частями текстового представления осуществляется с помощью оглавления, и как во многих видах текстовых документов, имеется терминологический словарь.

Достоинством естественно-языкового (ЕЯ) текстово-графического представления предметной области курса является ориентация на человека, что обеспечивает использование его во всех автоматизированных обучающих системах.

Недостатками естественно-языковых представлений знаний учебного курса являются:

- отсутствие формализации в структуре естественного языка;
- наличие большого количества неопределенностей, синонимов, антонимов;
- структуризация текстового документа ограничивается его оглавлением и при наличии страничного разбиения текста в какой-то степени, терминологическим словарем;
- информация, включаемая в разделы, подразделы может не носить целостный характер, что не свойственно предметной области.

Формализованное UML-представление предметной области имеет графический интерфейс, понятный специалистам по знаниям, фиксированную семантику элементов языка, более строгую семантику языковых конструкций. Фактически, UML – это многоязыковая система, способная описывать как статику (классовые диаграммы), так и динамику предметной области (диаграммы последовательностей, диаграммы состояний, диаграммы активностей) [5]. Несмотря на солидный теоретический фундамент, для некоторых подязыков UML семантика его языковых конструкций оставляет желать лучшего. Тем не менее, с помощью UML можно добиться компактного описания предметной области. Опыт использования UML, приобретенный при объектно-ориентированном проектировании программных систем, связанный с преобразованием текстового описания системы в UML диаграммы, позволяет надеяться на возможность его использования как средства организации баз знаний в АСОЗ.

Формализация описания предметной области, базирующаяся на использовании определенных математических теорий, в сильной степени определяется свойствами этих теорий и достаточно далеки от текстовых описаний. Процесс такой формализации предметной области, отталкиваясь от понятных человеку текстовых описаний, является достаточно трудоемким и требует задействования экспертов в данных теориях. Например, построение динамической предметной области в виде сети Петри [9] или полумарковского процесса [10] при описании реальных систем является нетривиальной задачей. То же можно сказать о попытках использования для формализации представления предметной области различных вариаций математической логики, включая теорию предикатов первого порядка, различные системы на основе логических правил и т. д. Однако эти формализмы позволяют использовать при анализе формализованных представлений предметной области (ПО) развитый математический аппарат, включая механизмы логического вывода, имитационное моделирование и др. Сводная информация о возможностях различных вариантов представлений ПО, полученная на базе экспертных оценок, дана в таблице.

Сравнительные характеристики различных представлений знаний о предметной области

Сравнительные характеристики различных представлений знаний о предметной области	Степень близости к конечному пользователю	Уровень формализации	Возможности удобного описания структуры системы	Возможности удобного описания динамических свойств ПО
Естественно-языковые представления ПО учебного курса (УК)	5	0	1	1
UML	4	3	4	4
Логические исчисления в виде предикатов	1	5	4	3
Логические исчисления в виде правил	2	5	2	3
Фреймы	1	3	1	1

Из таблицы видно, что наилучшего представления знаний об учебном курсе не существует. Более того, имеется необходимость в АСОз использовать несколько его представлений в соответствии с целями использования:

- при работе тьютора и ученика – текстовые представления;
- при автоматизированной оценке знаний ученика – UML, логические исчисления;
- при адаптации учебного материала к уровню знаний ученика – текстовые представления, UML.

В таблице использованы два критерия метода представления знаний:

- близость научного сленга, на котором описывается учебный курс, к естественному языку (ЕЯ) обучаемых и лиц, подготавливающих учебный материал для использования в АСОз;
- возможность контролировать усвоение учебного материала.

Очевидна важность первого критерия: чем ближе сленг к родному естественному языку пользователя, тем больше возможностей для восприятия, анализа и препарирования учебного материала как для ученика, так и для лиц, готовящих информацию курса для АСОз.

Для оценки степени важности второго критерия необходимо помнить, что основным для контроля при обучении является проведение опросов по пройденному материалу (экзамен).

От степени покрытия вопросами предметной области учебного курса и от степени сложности вопросов напрямую зависит качество тестовой последовательности, предоставляемой экзаменуемому. Естественно, на каждый из вопросов может быть дан правильный или неправильный ответ, и на базе ответов на все вопросы тестовой последовательности формируется интегральная оценка степени усвоения учебного материала.

Можно заметить, что описанная выше процедура контроля чрезвычайно сходна с процессом тестирования программ. Здесь, как и в случае тестирования программ, для каждого элемента тестовой последовательности экзаменатору нужно или заранее знать ответ, или воспользоваться механизмом вычисления самого ответа, или любым другим образом уметь определять правильность ответа.

Вторым пунктом сходства процесса контроля знаний и процесса тестирования программ является наличие бесконечного числа тестовых последовательностей для контроля степени усвоения порции учебного материала. Поэтому в программной инженерии для тестирования программ выработаны стратегии формирования ограниченного числа тестовых последовательностей, позволяющие существенно снизить их количество при контроле степени усвоения учебного материала.

Заметим, что одной из популярных стратегий, используемых при white box тестировании, является требование покрытия графа программы или автоматной модели программы набором путей для тестовой последовательности, покрывающих граф программы или граф состояний автоматной модели.

Учитывая то, что, как правило, модель предметной области также представима в виде пары

$$G_{\text{ПО}} = \langle E, R \rangle,$$

где E – набор сущностей, R – набор отношений.

Для любого структурного представления предметной области, в том числе и UML при контроле знаний можно использовать стратегию покрытия, соответствующую тому или иному типу диаграмм.

Структурные представления предметной области позволяют сформулировать определение еще одного важного показателя качества системы контроля знаний – сложность тестов.

Тест будем называть *простым*, если ответ на него находит непосредственное отражение в структурном (графовом) представлении предметной области. Примерами могут быть вопросы о наличии сущностей и связей, вопросы об их характеристиках, ответы на которые прямым образом отражены в структуре предметной области учебного курса.

Тест будем называть *сложным*, если ответ на вопрос требует решения задачи над предметной областью.

Для структурных представлений предметной области это может быть задача нахождения кратчайшего пути, задача о разметке графа, задача проверки свойств графа и многое другое. Поскольку получение ответа на вопрос для сложного теста может потребовать привлечения алгоритмов различной степени специализации. Метод представления предметной области здесь играет существенную роль. Так, например, использование для этих целей логических исчислений дает возможность использовать механизмы универсальных решателей задач, в частности, одну из разновидностей метода резолюций.

И, наоборот, при естественно-языковом представлении предметной области единственным механизмом структуризации является разбиение учебного курса на разделы, подразделы, параграфы, подпараграфы и т. д., что находит свое отображение в оглавлении курса. Последнее предполагает организацию ссылок на ЕЯ текст курса. Не исключено при этом разбиение ЕЯ текста на страницы. Однако такой способ структуризации крайне ограничивает возможности автоматизации контроля.

Поскольку порция ЕЯ представления текста, помимо синтаксической структуры, не имеет другого способа структуризации, автоматический контроль сводится к вопросам типа «Как называется раздел, параграф» и т. д.?

Использование в некоторых АСОз, ориентированных на ЕЯ представление предметной области, вопросников, направленных на выявление смысла в порции учебного материала, не опирается на формализованное представление предметной области, а скорее отражает мнение эксперта в предметной области курса. Поэтому сложно судить о степени освоения учебного материала или о его сложности.

Подводя итог, можно отметить, что нет однозначного ответа на вопрос о лучшей форме представления предметной области для базы знаний АСОз. В последовательности представлений «ЕЯ → UML → Логическое исчисление» происходит уменьшение значений первого критерия и возрастание для второго.

Исходя из вышесказанного, в качестве архитектуры базы знаний АСОз можно использовать структуру, представленную на рисунке.

УК НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

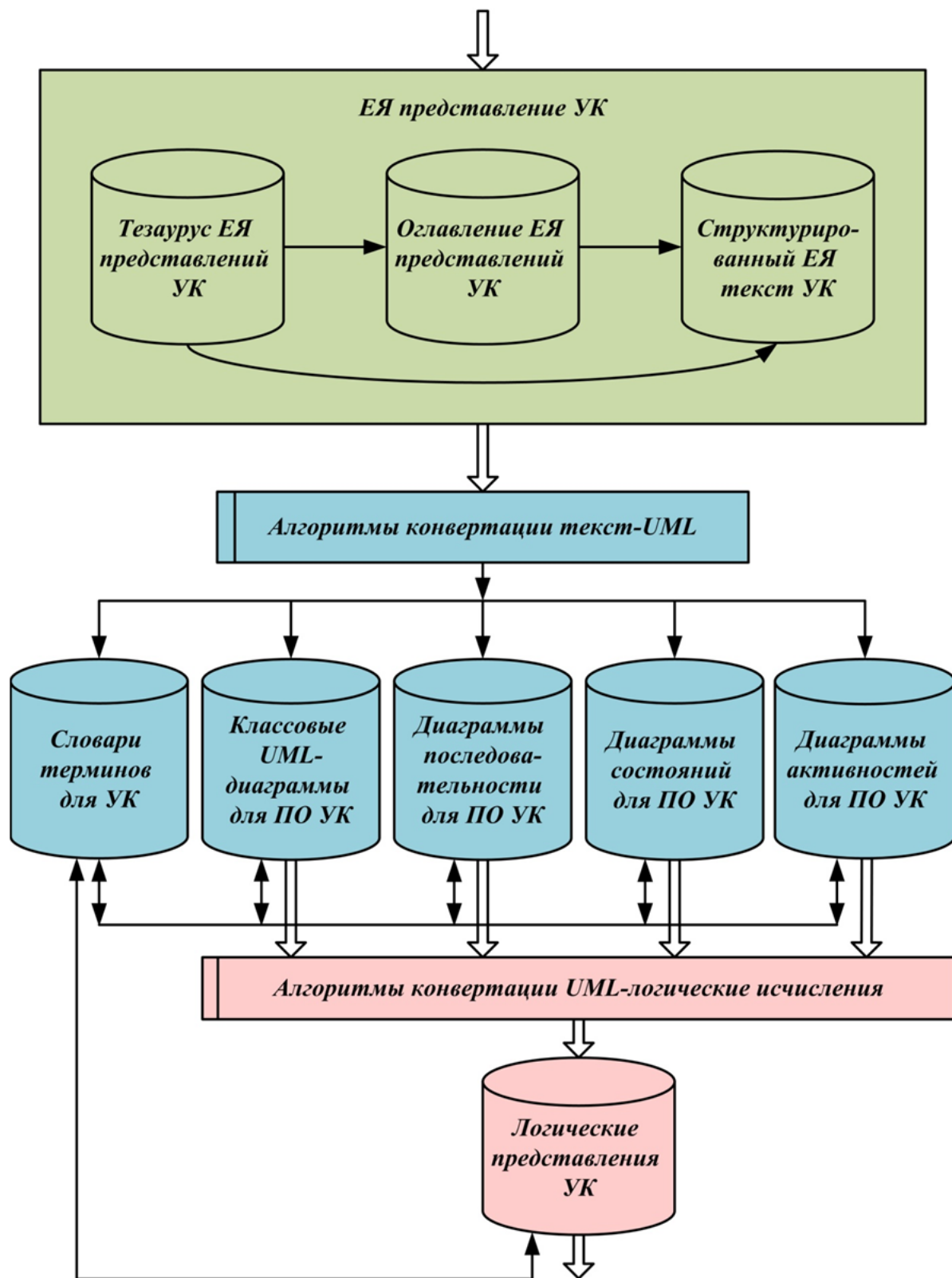


Рис. Архитектура БЗ АСОз

Особенности базы знаний знание-ориентированных систем обучения:

– БЗ АСО использует все три способа представления предметной области, что на рис. выделено разными цветами, представляющими собой слои. Первый слой предназ-

TECHNICAL SCIENCES AND TECHNOLOGIES

начен для ученика и тьютора, а в части формирования – экспертом в предметной области и инженером по знаниям, второй предназначен для работы инженера по знаниям, третий используется для создания вопросников при контроле знаний.

– Имеются средства автоматизированного формирования слоя структурного представления по данным слоя ЕЯ – текстового представления предметной области. Эти средства в основном используются экспертом в предметной области и инженером по знаниям.

– Имеются средства автоматического формирования слоя логического представления по информации слоя структурного представления.

– Имеются модули генерации простых вопросов и задач соответственно для 2-го и 3-го слоев.

– Имеются модули решения задач как на уровне второго слоя, так и универсальный решатель для 3-го слоя.

Выводы. Предложенная архитектура позволяет сохранять и эффективно использовать преимущества каждого из слоев представления знаний в общей системе. При данной архитектуре важную роль в обеспечении эффективности работы системы играют преобразователи «ЕЯ текст→UML» и «UML→Логические представления».

Такая архитектура дает возможность хранить знания в актуальном виде, не влияя на взаимодействие функциональных модулей АСОЗ.

Список использованных источников

1. Lytvynov V. Knowledge representation in the automated learning systems / Vitalii Lytvynov, Iryna Posadska // International Journal “Information Technologies & Knowledge”. – 2015. – Vol. 9, No. 1. – Pp. 34–43.

2. Литвинов В. В. Архитектура знание-ориентированной автоматизированной системы обучения / В. В. Литвинов, И. С. Посадская, М. В. Савельев // Технічні науки та технології. – 2016. – № 3 (5). – С. 122–130.

3. Клыков Ю. И. Банки данных для принятия решений / Ю. И. Клыков, Л. Н. Горьков. – М. : Сов. радио, 1980. – 208 с.

4. Шустов С. Б. Теория ресурсов и ресурсные кризисы: прошлое, настоящее и будущее: аналитический обзор / С. Б. Шустов. – Нижний Новгород : Нижний Новгород, 2009. – 163 с.

5. Bruce Powel Douglass. Real – Time UML. Second Edition. Developing Efficient Objects for Embedded Systems / Bruce Powel Douglass. – Wesley, 1999. – 238 p.

6. Лорьер Ж.-Л. Системы искусственного интеллекта : пер. с франц. / Ж.-Л. Лорьер. – М. : Мир, 1991. – 568 с.

7. Minsky M. A Framework for Representing Knowledge: Patrick Henry Winston / Marvin Minsky // The Psychology of Computer Vision. – McGraw-Hill : New York (U.S.A.), 1975. – 76 p.

8. Искусственный интеллект. Кн. 2: Модели и методы : справочник / под ред. Э. В. Попова. М. : Радио и связь, 1990. – 303 с.

9. Питерсон Дж. Теория сетей Петри и моделирование систем / Дж. Питерсон. – М. : Мир, 1984. – 264 с.

10. Королюк В. С. Процессы марковского восстановления в задачах надежности систем / В. С. Королюк, А. Ф. Турбин. – К. : Наук, думка, 1982. – 236 с.

References

1. Lytvynov, V., Posadska, I. (2015). Knowledge representation in the automated learning systems. *International Journal “Information Technologies & Knowledge”*, vol. 9, no. 1, pp. 34–43.

2. Litvinov, V.V., Posadskaya, I.S., Savelev, M.V. (2016). Arkhitektura znanie-orientirovannoi avtomatizirovannoi sistemy obucheniia [Architecture of knowledge-oriented automated system of learning]. *Tekhnichni nauki ta tekhnologii – Technical Sciences and Technologies*, no. 3 (5), pp. 122–130 (in Russian).

3. Klykov, Iu.I., Gorkov, L.N. (1980). *Banki dannykh dlia priniatiia reshenii [Databases for decision-making]*. Moscow: Sov. radio (in Russian).

4. Shustov, S.B. (2009). *Teoriia resursov i resursnye krizisy: proshloe, nastoiashchee i budushchee: analiticheskii obzor [The theory of resources and resource crises: past, present and future: an analytical review]*. Nizhnii Novgorod: Nizhnii Novgorod (in Russian).
5. Douglass, Bruce Powel (1999). *Real - Time UML. Second Edition. Developing Efficient Objects for Embedded Systems*. Wesley.
6. Lorier Zh.-L. (1991). *Sistemy iskusstvennogo intellekta [Artificial Intelligence Systems: Translation from French]* [Trans. from French]. Moscow: Mir (in Russian).
7. Minsky, Marvin (1975). *A Framework for Representing Knowledge: Patrick Henry Winston. The Psychology of Computer Vision. McGraw-Hill: New York (U.S.A.)*.
8. Popova, E.V. (ed.) (1990). *Iskusstvennyi intellekt. Kn. 2: Modeli i metody: spravochnik [Artificial Intelligence. Book 2: Models and Methods: A Handbook]*. Moscow: Radio i sviaz (in Russian).
9. Piterson, Dzh. (1984). *Teoriia setei Petri i modelirovanie sistem [The theory of Petri nets and modeling systems]*. Moscow: Mir (in Russian).
10. Korolyuk, V.S. & Turbin, A.F. (1982). *Protsessy markovskogo vosstanovleniya v zadachakh nadezhnosti sistem [Markov renewal processes in systems reliability problems]*. Kiev : Nauk. dumka (in Russian).

Посадская Ирина Сергеевна – ассистент кафедры информационных технологий и программной инженерии, Черниговский национальный технологический университет (ул. Шевченко, 95, г. Чернигов, 14027, Украина).

Посадська Ірина Сергіївна – асистент кафедри інформаційних технологій і програмної інженерії, Чернігівський національний технологічний університет (вул. Шевченка, 95, м. Чернігів, 14027, Україна).

Posadska Iryna – assistant of the Department of Information Technologies and Software Engineering, Chernihiv National University of Technology (95 Shevchenka Str., 14027, Chtrnihiv, Ukraine).

E-mail: irrkin@gmail.com

ORCID: <http://orcid.org/0000-0002-4905-2552>