

ТЕОРИЯ ФОРМАЛИЗАЦИИ ЗНАНИЙ ПРЕДМЕТНОЙ ОБЛАСТИ ДЛЯ ПОИСКОВОГО ИССЛЕДОВАНИЯ

О.Б. ПОПОВА

*Кубанский государственный технологический университет,
350002, Российская Федерация, г. Краснодар, ул. Московская, 2,
электронная почта: popova_ob@mail.ru*

Сегодня существует много формальных моделей описания знаний, используемых для соответствующей аппаратно-программной реализации прикладных задач. Они не подходят при формализации предметной области поискового исследования. Автор для получения новой формальной модели описания знаний предлагает использовать стандартные технологии формального описания знаний и определения, но внести изменения в процедуры и инструментарий технологии под предметную область поискового исследования. В результате получены новые методы: представления знаний и организации знаний, которые приближены к естественному интеллекту, используя свойственное ему структурирование знаний. В искусственной форме автором были использованы современные информационные технологии формально-языкового описания знаний – продукционные правила, формализация области знаний в объектном виде, реализация когнитивности, структурность знаний, объединённых в иерархическую структуру, где используется метод классификации и теория энтропии. Чтобы приблизить их к естественному интеллекту, разработана новая теория получения «бинарного дерева системы вопросов и ответов», включающая правила получения узлов данной структуры, три метода расстановки ключей, преобразующих её для программной реализации. Предлагается современная концепция хранения и распространения знаний в формально-языковой форме.

Ключевые слова: формализация знаний, предметная область, поисковое исследование, представление знаний, организация знаний.

Сегодня существует огромное количество формальных моделей описания знаний, которые применяются для соответствующей аппаратно-программной реализации решаемой задачи. Представление знаний разделяют на декларативные и процедурные, среди которых процедурные знания позволяют создавать более эффективные интеллектуальные информационные системы. Так как они содержат исходные состояния и знания по существу решаемой проблемы, которые используются в дальнейшем анализе и принятии решений. Всё это «экономит память при хранении знаний, но ухудшает возможности пополнения баз знаний» [1]. Существует структурное и параметрическое представление знаний. Где структурное представление следует понимать как постепенное теоретическое накопление знаний, описанных через множество

фактов или объектов. Предполагается, что с течением времени проблемная область конкретизируется за счёт изменения структуры знаний. Структура здесь изменяется эволюционно или адаптивно [1]. Использование параметрического представления знаний предполагает моделирование знаний, полученных опытным путём. Конкретизация знаний здесь при решении текущей задачи производится изменением соответствующих параметров выбора, которые позволяют принять нужное решение. Сейчас наиболее эффективным считается совместное комбинирование этих представлений, так как совместно используется теоретическое и опытное накопление знаний о проблемной области.

Сейчас существуют три модели описания, представляющие «системы знаний с помощью разного математического аппарата для конкретного формального описания и построения процедуры решения задачи» [1, с. 30]. Это реляционные, объектные и ассоциативные модели описания. Каждая модель описания имеет свой перечень способов реализации. Самыми эффективными технологиями решения из известных способов реализации этих моделей можно считать – использование продукционных правил, формализация области знаний в объектном виде, реализация когнитивности, структурность знаний в нейронной форме (набор "связанных локальных нейросетей, объединённых в иерархическую структуру" [1, с. 33]), где используются методы классификации и теория энтропии.

Указанные выше модели используются для реализации алгоритмов методов решения различных прикладных задач прогнозирования, планирования и управления. В этих задачах используется своя технология представления предметной области, которая совершенно не подходит для представления предметной области для поискового исследования [2], так как там используется совершенно иной алгоритм решения – «определение перспективности работы над темой, отыскивание путей решения научных задач», тогда как в прикладных задачах производится решение задачи в найденном направлении. Так как прикладное исследование [2] – это «применение новых знаний для

достижения практических целей, решения конкретных задач». Поэтому в поисковых исследованиях сейчас в основном применяются вспомогательные информационные технологии для эффективной работы естественного интеллекта над задачей поиска наиболее эффективного метода решения научной задачи. К ним относятся: автоматизированные системы научных исследований, которые позволяют эффективно проводить эксперименты и работать с полученными экспериментальными данными; поисковые системы; информационные системы обработки текста; различные базы научного цитирования; электронные издания с системой удалённого поиска, рецензирования и публикации статей. Алгоритм поискового исследования традиционен и обязателен, результатом которого являются хорошие обзорные статьи по выбранной области научного исследования. Выбрав предмет исследования из такой обзорной статьи можно не беспокоиться за перспективность темы исследования и выбранного пути решения научной задачи. Сейчас существует принятое естественно-языковое описание области научного исследования, которое используется в научных журналах. Прочитав статью, естественный интеллект способен структурировать материал статьи, представив область в удобном для понимания виде. Процесс представления и понимания завершается осознанным выбором. Сегодня эти процессы широко изучаются в психологии, которая ищет пути усиления человеческого интеллекта через создания методик визуализации и структурирования знаний. Автор предлагает использовать методики эффективного структурирования знаний в представлении предметной области для поискового исследования. Предлагаемый автором алгоритм структурирования соответствует процессам, происходящим в естественном интеллекте, поэтому он может быть использован в процессе создания новой формальной модели описания знаний, для которой будет применена новая программная реализация. Эффективность такой реализации должна приближаться к эффективности естественного интеллекта. Для получения новой формальной модели описания знаний автором использована стандартная технология и необходимые определения.

Так, предметная область – это «сфера деятельности человека, выделенная и описанная в соответствии с некоторыми целями» [3]. Применительно к поисковому исследованию, *деятельность человека* – поиск пути решения, что выражается в нахождении метода решения из множества известных методов решения или создание своего. *Некоторой целью* поискового исследования является решение текущей научной задачи, для которой ищется метод решения. «Описание предметной области включает: *предметы* (объекты, процессы, явления); *отношения* между выделенными предметами и / или их частями; *взаимодействия* между предметами, их частями и отношениями, возникшие в результате осуществления деятельности человека» [3]. Теперь, чтобы правильно описать предметную область, необходимо определить предмет для поискового исследования. Автор предлагает выбрать *предметом поискового исследования метод решения*, так как такой подход позволит определить наиболее подходящий метод решения задачи, который: может принадлежать смежной научной области исследования; может появиться недавно; ранее не применялся, но более эффективен, чем принятый. Если предметом поискового исследования выбрать научную задачу, то исследование из поискового вида перейдет к виду прикладной научной задачи или к виду научной разработки [2].

Сведения о предметах представляются в знаковой форме [3]. Теперь, дадим определение знака – это «символ наших идей». Идея – «мысленный прообраз какого-либо действия, предмета, явления, принципа, выделяющий его основные, главные и существенные черты» [4]. Сегодня для описания символов широко используется в статьях, книгах и учебниках естественно-языковая знаковая система. Следовательно, собирая сведения об известных методах решения из научных статей и другой требуемой литературы, можно найти для каждого из методов решения их *главные и существенные черты*, которые определяют *отношения* между методами решения и их *взаимодействие*. Главными и существенными чертами каждого метода будем считать определённые свойства решаемой задачи, ради решения которой создавался

данный метод, так как появление задачи с новыми свойствами предопределяет необходимость появления нового метода решения *задачи такого типа*, то есть с определёнными свойствами, отличающими её от других задач. Естественно предположить, что некоторые группы методов решения будут обладать рядом схожих свойств, чем-то они будут отличаться. Всё это определит их взаимодействие с определёнными типами задач, для которых они могут быть использованы как наиболее подходящие.

Автор предлагает свой тип знаковой системы для описания собранных сведений о методах решений задач взятых из естественно-языковой системы описания. Пусть собранные сведения – это исходные данные, которые необходимо «опредметить» [2]. Другими словами, необходимо провести «автоформализацию знаний – придание владельцем знаний (автором) им некоторой формы» [2]. Так, формализованные знания предметной области для поискового исследования получают свой искусственно-языковой образ. Используем для получения образа формально-языковой дифференцированный метод формализации знаний. Искусственная форма организации знаний такой предметной области должна обладать следующими свойствами: ориентированность на определённые естественные формы; существенная зависимость от носителя образа; моделирование способности субъекта и его деятельности [2]. Как видим, искусственная форма должна визуально повторять структурную форму представления знаний естественным интеллектом и позволять выбрать наиболее эффективный метод решения задачи среди всех известных методов решения.

Для получения искусственной формы автором были использованы указанные выше современные и эффективные информационные технологии формально-языкового описания знаний – продукционные правила, формализация области знаний в объектном виде, реализация когнитивности, структурность знаний, объединённых в иерархическую структуру, где используется метод классификации и теория энтропии. Все они должны быть приближены к естественному интеллекту, поэтому их общие процедуры и

инструментарий технологии соответствующим образом изменён автором под предметную область для поискового исследования. В результате автором получен ряд новых методов: метод представления знаний и метод организации знаний, которые приближены к естественному интеллекту и используют свойственное ему структурирование знаний.

В инструментарий технологии формально-языкового описания знаний вводится новая теория получения бинарного дерева системы вопросов и ответов [5], которая используется в новом методе усиления интеллекта, новом методе представления дерева принятия решений и новом методе классификации [6-8], которые получены автором и приближены к естественному интеллекту. В этой теории предлагаются свои правила получения узлов дерева [9 – 11], в которых «опредмечиваются» *исходные данные*. Листами дерева являются *цели поискового исследования* – наиболее подходящий метод решения текущей задачи. Путь по дереву – это *процесс нахождения решения при поисковом исследовании, свойственный естественному интеллекту*.

В новом методе представления знаний предметной области для поискового исследования автор использует новый метод классификации, в котором используется бинарное дерево системы вопросов и ответов со своими правилами получения узлов дерева. Построение указанного дерева использует изменённые положения из теории информации, которые затрагивают следующие понятия: энтропия и знание. Автор вводит следующее предположение [8], которое соответствует реальной действительности. Естественному интеллекту характерно оперировать не битами, а теми объёмами знаний, которые он усвоил, то есть понял и, следовательно, смог соответствующим образом структурировать. Тогда, в соответствии с основными положениями из теории информации и определениями энтропии можно считать, что прирост знаний Z на определённую величину ΔZ позволит уменьшить энтропию S на некоторую величину ΔS . Это возможно, если знание ΔZ отличается от знания Z каким-то свойством C . Свойство C можно считать

атрибутом. Следовательно, в узлах дерева расположены *главные и существенные черты* методов решения. Структура построенного дерева определяет *отношения и взаимодействия* между всеми методами решения и задачами, требующими решения [8].

Новый метод организации знаний предметной области для поискового исследования, разработанный автором, который приближен к естественному интеллекту и использует свойственное ему структурирование знаний, содержит следующий алгоритм действий. Сначала создаются знания в формально-языковой форме – в виде бинарного дерева системы вопросов и ответов, которое представляет собой первичное представление знаний [8] удобное для визуализации всей предметной области, её проверки, тестирования и верификации. Формирование такого дерева требует сбор исходных данных – поиск актуальной литературы по заданной области научного исследования, сбор сведений о методах решения из текста, представленного в естественно-языковой знаковой форме. Последовательно выделяются *главные и существенные черты* методов решения, используя новый метод классификации, приближенный к естественному интеллекту. Потом производятся необходимые действия для хранения и распространения знаний в формально-языковой форме.

Для дальнейшего хранения знаний в формально-языковой форме в новом методе организации знаний предметной области для поискового исследования используется следующий инструментарий для программной реализации, который разработал автор. Где инструментарий – «часть информационной технологии, метод представления знаний и средства их поддержки» [3]. Сначала производится перевод первичного представления знаний (эскиза дерева) в абстрактный тип данных «бинарное дерево системы вопросов и ответов» для программной реализации на языке высокого уровня. С этой целью автором разработаны и использованы три метода расстановки ключей в дереве [5]. Первый метод расстановки ключей позволяет преобразовать эскиз дерева, которое занимает до 34 листов машинописного текста [5] или 25 листов

рукописного текста, на котором так же расставляются ключи, к компактному виду. Всё дерево, узлы которого заменены ключами, может быть представлено на одном рисунке. Такая замена позволяет проверить правильность полученной структуры дерева и её связей. Другими словами, проверить соотношения и взаимодействия между выявленными предметами. Тщательно проверяются переходы по дереву, которые образуют ответы на вопросы в узлах дерева. Где ответ «да» соответствует переходу вниз влево по дереву, ответ «нет» соответствует переходу вниз вправо по дереву. После проверки производится расстановка ключей в соответствии с алгоритмом второго метода на преобразованном дереве. Значения ключей с заданным шагом, например, равным 20, расставляются таким образом, что бы получилось отсортированное бинарное дерево. Такой метод расстановки ключей необходим для использования эффективной реализации операторов, которые работают с абстрактной структурой данных «бинарное дерево системы вопросов и ответов». Их можно теперь реализовать рекурсивным образом, при этом сама структура дерева реализована через указатели. Таким образом, полученная абстрактная структура данных имеет динамическую структуру. Для организации удобного хранения исходных данных, их загрузки в дерево и внесения новых элементов в структуру дерева, если обнаружен новый метод решения, был разработан третий метод расстановки ключей в дерево. Его алгоритм расстановки ключей такой – с шагом 1 расставляются ключи, начиная с корня дерева, потом каждый раз опускаясь на уровень ниже, посещают сначала все промежуточные узлы дерева слева на право, потом, то же самое делают со всеми листьями дерева, опускаясь к самому последнему уровню. Все ключи, полученные тремя методами, сводятся на одном рисунке, изображающем структуру «бинарного дерева системы вопросов и ответов». Автор предлагает следующий метод их визуализации. В кружках, которые обозначают все узлы дерева, располагают ключи по первому методу расстановки. Слева от кружков другим цветом, например, зелёным, изображают ключи по второму методу расстановки. Тогда справа красным цветом

указывают ключи по третьему методу расстановки. Такая визуализация всех методов расстановки помогает составлять программный код на любом языке и тестировать полученный программный продукт.

Распространения знаний в формально-языковой форме в новом методе организации знаний предметной области для поискового исследования предполагает использование языка высокого уровня, который является *free cross-platform*, использует *free compiler* и создаёт свободно распространяемый код под открытой лицензией, например, *GNU General Public License*, *GNU Lesser General Public License*. Полученное программное обеспечение может быть расположено на веб-сервисе для хостинга IT-проектов и их совместной разработки. Предложенный инструментарий позволяет создать программный продукт, который может быть использован:

- в учебных заведениях для изучения методов решения;
- в проектных организациях, для решения прикладных задач;
- учёными для решения научных задач.

Используемые концепции и цели разработки программного продукта создают следующие перспективы его развития:

- можно ссылаться на результаты выполнения программы или на предметную область, используемую в программе, если в программе использовать так же ссылки на литературу;
- изменение предметной области совместно с другими учёными, используя сетевые технологии.

Теория формализации знаний предметной области для поискового исследования была опробована автором на практическом примере [5-15]. Была выполнена формализация следующей предметной области – методы оптимизации, которая была использована в программе «Оптимэль» [12], помогающей выбрать наиболее подходящий оптимизационный метод для решения текущей оптимизационной задачи.

Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда № 16-03-00382 от 18.02.2016 в рамках темы

“Мониторинг исследовательской деятельности образовательных учреждений в условиях информационного общества”.

ЛИТЕРАТУРА

1. Станкевич, Л. А. Интеллектуальные системы и технологии: учебник и практикум для бакалавриата и магистратуры / Л. А. Станкевич. – М.: Издательство Юрайт, 2016. – 397 с. – [Электронный ресурс] – <https://www.biblio-online.ru/viewer/750B2832-EE3F-4E58-9F52-037C8E916959#page/1>
2. Научное исследование. Материал из Википедии – свободной энциклопедии, 2016. - https://ru.wikipedia.org/wiki/Научное_исследование
3. Формализация знаний в интеллектуальных системах. Организация знаний: Методы и решения в системах организации знаний, 2016. - <http://it-claim.ru/Education/Course/Knowledge/Files/L21.pdf>
4. Идея. Материал из Википедии – свободной энциклопедии, 2016. - <https://ru.wikipedia.org/wiki/Идея>
5. Бинарное дерево выбора знания из области знания, используя систему вопросов и ответов. Теория и практика: монография / О.Б. Попова, Б.К. Попов, В.И. Ключко; ФБГОУ ВПО «Кубан. гос. технол. ун-т». – Краснодар: Издательский Дом – Юг, 2013 – 166 с.
6. Попова О.Б. Новые метод усиления интеллекта и способ представления дерева принятия решений, которые приближены к естественному интеллекту // Научные труды Кубанского государственного технологического университета. – 2016. – №1, с. 38-47; URL: <http://ntk.kubstu.ru/file/779>.
7. Попова О.Б. Современный классификатор для предметной области // Научные труды Кубанского государственного технологического университета. – 2016. – №13, с. 100-111; URL: <http://ntk.kubstu.ru/file/1259>.
8. Попова О.Б. Энтропия в дереве принятия решений, получение которого приближено к естественному интеллекту // Научные труды Кубанского государственного технологического университета. – 2016. – №12, с. 156-170; URL: <http://ntk.kubstu.ru/file/1226>.
9. Попова О.Б., Попов Б.К., Ключко В.И. Получение корня бинарного дерева системы вопросов и ответов // Современные проблемы науки и образования. – 2013. – № 3; URL: <http://www.science-education.ru/109-9146>.
10. Попова О.Б., Попов Б.К., Ключко В.И. Правила получения элементов бинарного дерева системы вопросов и ответов // Фундаментальные исследования. – 2013. – №6-1. – С. 55-59; URL: <http://www.fundamental-research.ru/ru/article/view?id=31413>
11. Popova O., Popov B., Karandey V., Evseeva M. Intelligence amplification via language of choice description as a mathematical object (binary tree of question-answer system) // Procedia – Social and Behavioral Sciences. – 2015. – V. 214. – С. 897–905; URL: <http://www.sciencedirect.com/science/article/pii/S1877042815061030>.

12. Попова О.Б., Попов Б.К. «Оптимэль». Свидетельство о государственной регистрации программы №2012615868 от 27.06.2012.

13. Системный анализ процесса выбора метода оптимизации информационной системы: монография / О.Б. Попова, Б.К. Попов, В.И. Ключко; ФБГОУ ВПО «Кубан. гос. технол. ун-т». – Краснодар: Издательский Дом – Юг, 2012 – 135 с.

14. Попова О.Б. Системный подход к исследованию процесса оптимизации. Деп. в ВИНТИ №83-В2010 от 17.02.2010.

15. Попова О.Б., Попов Б.К., Ключко В.И. Проблема сокращения времени выбора методов управления большими системами (БС) // Современные проблемы науки и образования. – 2013. – №1. – С. 163; URL: <http://www.science-education.ru/ru/article/view?id=8371>.

REFERENCES

1. Stankevich, L. A. Intellektual'nye sistemy i tehnologii: uchebnik i praktikum dlja bakalavriata i magistratury / L. A. Stankevich. – М.: Izdatel'stvo Jurajt, 2016. – 397 s. – [Jelektronnyj resurs] – <https://www.biblio-online.ru/viewer/750B2832-EE3F-4E58-9F52-037C8E916959#page/1>

2. Nauchnoe issledovanie. Material iz Vikipedii – svobodnoj jenciklopedii, 2016. – https://ru.wikipedia.org/wiki/Научное_исследование

3. Formalizacija znanij v intellektual'nyh sistemah. Organizacija znanij: Metody i reshenija v sistemah organizacii znanij, 2016. – <http://it-claim.ru/Education/Course/Knowledge/Files/L21.pdf>

4. Ideja. Material iz Vikipedii – svobodnoj jenciklopedii, 2016. – <https://ru.wikipedia.org/wiki/Идея>

5. Popova O.B., Popov B.K., Kljuchko V.I. Binarnoe derevo vybora znanija iz oblasti znanija, ispol'zuja sistemu voprosov i otvetov. Teorija i praktika: monografija (A binary tree of the knowledge of the selection field of knowledge, using a system of questions and answers. Theory and practice: a monograph). – Krasnodar: OOO «Izdatel'skij Dom-Jug», FGBOU VPO «KubGTU», 2013. – 166 s.

6. Popova O.B. Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta (Proceedings of the Kuban State University of Technology), 2016, №1, S. 38-47; URL: <http://ntk.kubstu.ru/file/779>.

7. Popova O.B. Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta (Proceedings of the Kuban State University of Technology), 2016, №13, S. 100-111; URL: <http://ntk.kubstu.ru/file/1259>.

8. Popova O.B. Nauchnye trudy Kubanskogo gosudarstvennogo tekhnologicheskogo universiteta (Proceedings of the Kuban State University of Technology), 2016, №12, S. 156-170; URL: <http://ntk.kubstu.ru/file/1226>.

9. Popova O.B., Popov B.K., Kljuchko V.I. Sovremennye problemy nauki i obrazovanija (Modern problems of science and education), 2013, № 3; URL: <http://www.science-education.ru/109-9146>.

10. Popova O.B., Popov B.K., Kljuchko V.I. Fundamental'nye issledovanija (fundamental research), 2013, №6-1, S. 55-59; URL: <http://www.fundamental-research.ru/ru/article/view?id=31413>.

11. Popova O., Popov B., Karandey V., Evseeva M. Intelligence amplification via language of choice description as a mathematical object (binary tree of question-answer system). *Procedia – Social and Behavioral Sciences*, 2015, V. 214, C. 897–905; URL: <http://www.sciencedirect.com/science/article/pii/S1877042815061030>.

12. Popova O.B., Popov B.K. Svidetel'stvo o gosudarstvennoj registracii programmy (State registration certificate program) № 2012615868 ot 27.06.2012

13. Popova O.B., Popov B.K., Kljuchko V.I. The system analysis of the process to choose the method of optimization of the information system: monograph (The system analysis of the process to choose the method of optimization of the information system: monograph). – Krasnodar: OOO «Izdatel'skij Dom-Jug», FGBOU VPO «KubGTU», 2012. – 135 s.

14. Popova O.B. Sistemnyj podhod k issledovaniju processa optimizacii. Dep. v VINITI №83-V2010 ot 17.02.2010.

15. Popova O.B., Popov B.K., Kljuchko V.I. Sovremennye problemy nauki i obrazovanija (Modern problems of science and education), 2013, №1, S. 163; URL: <http://www.science-education.ru/ru/article/view?id=8371>.

THE THEORY OF FORMALIZATION OF THE SUBJECT DOMAIN FOR EXPLORATORY RESEARCH

O.B. POPOVA

*Kuban State Technological University,
2, Moskovskaya st., Krasnodar, Russian Federation, 350002,
e-mail: popova_ob@mail.ru*

Today there are many models of formal description of knowledge used for the appropriate hardware and software implementation of applications. They do not fit in the formalization of subject area exploratory research. Author for obtaining the new model of the formal description of the knowledge proposes to use the standard technology of the formal description of knowledge and determination, but make changes to the procedures and tools technology into subject area exploratory research. As a result, are obtained new methods: knowledge representation and organization of knowledge, which are close to the natural intelligence, using his inherent structuring knowledge. The artificial form of the author of modern information technologies formal language description of knowledge were used - production rules, formalization of knowledge in the area of the object form, the implementation of cognitive, structural knowledge, united in a hierarchical structure, where the method of classification and theory of entropy is used. In order to bring them closer to the natural intelligence, developed a new theory of receipt of the «binary tree of questions-answers system», which includes the rules for obtaining a given structure nodes, three methods of arrangement of keys, converting it to a software implementation. It is proposed to the modern concept of knowledge storage and distribution in the form of a formal language.

Key words: formalization of knowledge, subject area, exploratory study, knowledge representation, knowledge organization.