

## О МЕТОДИКЕ СОЗДАНИЯ ВИРТУАЛЬНОГО КОНСУЛЬТАНТА С ИСПОЛЬЗОВАНИЕМ МИВАРНЫХ ТЕХНОЛОГИЙ



**Л.Е. Адамова**  
Заместитель директора  
ООО «Мивар», кандидат  
психологических наук,  
доцент



**Д.А. Протопопова**  
Младший научный  
сотрудник, ООО  
«Мивар»



**И.А. Шошев**  
Младший научный со-  
трудник, ООО  
«Мивар»

ООО «Мивар», Российская Федерация

E-mail: [l.adamova@mivar.ru](mailto:l.adamova@mivar.ru), [d.protopopova@mivar.ru](mailto:d.protopopova@mivar.ru), [i.shoshev@mivar.ru](mailto:i.shoshev@mivar.ru)

*Abstract.* The technique of developing a universal question-answering system "Mivar Virtual Advisor" is considered. The proposed method involves the representation of information about the subject area, texts, and queries (questions from users) as mivar knowledge representation model. Gained knowledge subsequently form the knowledge base of the virtual adviser. To construct such a framework the specialized information technology platform to understand the meaning of text in a natural language Tel!Mi is proposed. Such a system is able to accumulate knowledge from texts in natural Russian language and manage them.

*Введение.* Способность понимать и применять естественный язык является одним из основных аспектов исследований в области искусственного интеллекта. Успешная автоматизация такой способности выведет современные информационные технологии на принципиально новый уровень. Стоит отметить, что для понимания естественного языка недостаточно просто разбить предложения на части и найти их значения в словаре. Необходимо иметь обширные знания о предметной области и о контексте.

Большой интерес вызывают такие системы понимания естественного языка, как вопросно-ответные системы (Question-Answering systems). Такие системы, способные понимать вопросы и отвечать на них на естественном языке, в настоящее время получают все большее распространение [4-6]. Одним из перспективных направлений применения вопросно-ответных систем является создание виртуальных консультантов (ВК), основная цель которых состоит в консультировании пользователей веб-сайтов. На ВК можно возложить обязанности общения с посетителями, помощника в навигации по сайту, виртуального секретаря, сотрудника техподдержки или call-центра и т.д.

ВК относятся к системам, основанным на знаниях. В основе таких систем обычно лежит структурированная семантическая модель предметной области, по которой осуществляется поиск ответа на поставленный вопрос. Источником

информации для моделей могут служить тексты на естественном языке. Для хранения и накопления знаний в семантических моделях используются различные подходы, такие как, например, семантические сети, концептуальные зависимости, сценарии, фреймы, Resource Description Framework (RDF), миварная модель представления знаний и т.д. Для преобразования текста в такую модель разрабатываются специальные системы – семантические парсеры (semantic parsers).

Далее по построенной модели осуществляется поиск ответа на вопрос. Как известно [5-6], существуют способы автоматических ответов на вопросы посредством поиска и извлечения ответной фразы из больших коллекций текстовой информации (IR технологии). Такой подход опирается на огромное количество информации, доступной в текстовом виде в интернете, либо на специализированные текстовые коллекции. Получив вопрос от пользователя, технологии извлечения информации выбирают ответ, как кусок текста прямо из этих документов, руководствуясь текстами вопросов. При использовании данного подхода сначала обрабатывается вопрос, чтобы определить наиболее вероятный тип ответа, и формулируются запросы для поисковой системы. Поисковая система возвращает отранжированные документы, которые разбиты на подходящие отрывки. Кандидаты строк на ответы извлекаются из отрывков документов и ранжируются.

Такой метод оказался весьма успешными при поиске нужного документа, который релевантен запросу пользователя, в больших текстовых базах. Однако, очень часто пользователи хотят получить не целый документ, а краткий ответ на поставленный вопрос. Для поиска более точного ответа в документах необходимо использовать методы, которые более детально и разносторонне анализируют текстовую информацию.

В работе рассматривается методика создания виртуальных консультантов, основанная на знаниях. Для накопления знаний предлагается применять специализированную информационно-технологическую платформу для понимания смысла текста на естественном русском языке Tel!Mi, в основе которой лежит миварная модель представления знаний [1-3]. Используемая платформа дает возможность проводить морфологическую, синтаксическую и семантическую обработку текста, а также давать ответы на вопросы пользователей, основанные на знаниях [4-6]. Использование миварных технологий позволяет осуществлять понимание естественного русского языка и давать ответы на вопросы на более высоком уровне семантики. Во-первых, миварные технологии дают возможность обрабатывать сразу большие тексты, а не отдельные предложения. Во-вторых, миварные технологии позволяют на предварительных этапах обработки текста использовать традиционные синтаксические парсеры и другие методы традиционной математической лингвистики. В-третьих, миварный подход позволяет использовать сразу несколько статистически обученных традиционных парсеров и сравнивать полученные результаты.

*Методика создания виртуального консультанта.* Структура ВК может меняться в зависимости от используемого подхода и целей приложения. Но во всех таких системах можно выделить следующие обязательные компоненты: пользовательский интерфейс для общения с виртуальным консультантом, механизм обработки пользовательских вопросов и поиска ответов, база знаний ВК.

Предлагаемая в работе методика создания миварного виртуального консультанта делится на следующие этапы:

1 Анализ контекста предметной области. На данном этапе определяется принадлежность консультанта к конкретной предметной области. В зависимости от выбранной предметной области ВК будет обладать различным количеством знаний.

2 Создание и обучение базы знаний (БЗ) миварного виртуального консультанта. Данный этап является одним из основных в процессе создания ВК, так как именно от полноты базы знаний будет зависеть сможет ли ВК ответить на тот или иной запрос.

БЗ МКВ построена на концептах и сетях концептов. Под концептом понимается слово или словосочетание, которое отражает смысл термина или терминов. Сеть концептов - связи между концептами в картине мира [1]. Для формирования сети концептов необходимо обучить базу знаний ВК концептами, исходя из анализа предметной области и связать эти концепты между собой в единую сеть - картину мира. Если присутствует однозначность контекста по выбранной предметной области, то процесс формирования сети концептов значительно упрощается.

Если данных в БЗ консультанта, полученных на втором этапе недостаточно, то необходимо провести дополнительное обучение либо концептами, как на предыдущем этапе, либо текстами. Для обучения текстами необходимо провести графематический, морфологический, синтаксический и семантический анализ входящей информации. Под графематический анализом подразумевается разбиение полученного текста на абзацы, каждый абзац далее разбивается на предложения, предложение на словосочетания и слова. После этого требуется провести морфологический анализ слов в каждом предложении, полученном в результате графематического анализа. Итогом является получение массива словоформ с набором морфологических признаков (начальная форма, род, число и т.п.) в рамках каждого предложения. В процессе морфологического анализа Tel!Mi использует морфологический словарь, который содержит словоформы с набором морфологических признаков. На данный момент такой словарь содержит более 3 млн словоформ. Если в процессе морфологического анализа необходимая словоформа не найдена, предусмотрена возможность расширения морфологического словаря. Причем, предусмотрена автоматическая генерация словоформ с учётом правил русского языка (склонение, спряжение). После происходит синтаксический анализ, когда по предложению строится граф, связывающий слова в нём. Заключительным этапом анализа текста является

семантический анализ - определение смысла вводимой информации. Происходит обращение к имеющимся в БЗ МВК концептам и сетям концептов.

3 Проектирование интерфейса диалогового окна ВК. Выполнение данного этапа зависит от технического задания, в частности, от типа виртуального консультанта (персонализированный, неперсонализированный, технический). В случае персонализации ВК система Tel!Mi предусматривает обучение ответами на общие вопросы и на вопросы личного характера.

4 Разработка алгоритмов вывода ответа. Алгоритмы вывода ответа зависят от типа запроса пользователя (вопроса к консультанту). На первом этапе происходит анализ типа вопроса. После определения типа вопроса менеджер сценариев (блок алгоритмов вывода ответов) запускает соответствующий сценарий. Под сценарием понимается последовательность действий поиска ответа на вопрос конкретного типа. Итогом является вывод конкретного ответа в диалоговое окно с консультантом.

5 Тестирование. На этапе тестирования происходит проверка корректности данных хранящихся в базе МВК и отладка алгоритмов вывода ответа.

Данные этапы могут быть реализованы с использованием информационной платформы Tel!Mi, которая позволяет хранить большое количество информации, осмысленно систематизировать знания и управлять ими. Tel!Mi использует в своей основе миварную технологию представления знаний, включая пространство ВСО и миварные сети [1-3], что позволяет обрабатывать тексты на естественном русском языке. Стоит отметить, что Tel!Mi позволяет создавать ВСО-модель достаточно большого текста за приемлемое время. Например, время автоматической обработки Толкового словаря Ожегова составляет:  $0.05$  (среднее время обработки одной статьи) \*  $100\ 000$  словарных статей =  $5000$  секунд =  $83$  минуты. В настоящее время общий объем базы знаний, построенный с помощью Tel!Mi в формате модели ВСО, называемой "Картина мира", составляет  $160\ 000$  вершин на  $600\ 000$  ребер. Поиск ответа на вопрос пользователя в Tel!Mi выполняется на такой сверхбольшой модели ВСО менее, чем за минуту. Фактически решаются задачи поиска подграфа в графе и нахождения пути между вершинами графа.

Информационно-технологическая платформа позволила снизить адаптацию текстов для обучения виртуального консультанта. Поэтому миварный виртуальный консультант (МВК) способен работать с текстами на "живом" естественном языке. Кроме того, использование Tel!Mi позволяет создавать ВК с необходимым функционалом для совершенно разных предметных областей.

Использование миварных технологий для создания «виртуальных консультантов» может повысить качество работы контакт-центров и сервисов, ориентированных на массовый рынок. Общение пользователя с МВК осуществляется на контекстном уровне семантики. Созданные с помощью Tel!Mi «виртуальные консультанты» легко обучаемы: они способны понимать письменную человеческую речь, книжные тексты, адаптированные литературные произведения. На основании обученных знаний МВК способны

сами принимать решения в различных ситуациях общения с пользователем сервиса.

#### *Литература*

[1]. Варламов О.О. Основы многомерного информационного развивающегося (миварного) пространства представления данных и правил // Информационные технологии. 2003. № 5. pp. 42-47.

[2]. Варламов О.О. Роль и место миваров в компьютерных науках, системах искусственного интеллекта и информатике // Радиопромышленность. 2015. № 3, с. 10-27.

[3]. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. Москва. Радио и связь. 2002.

[4]. Люгер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е изд. Пер. с англ. – М.: Вильямс, 2005. — 864 с.

[5]. Hermjakob U., Novy E. H., Lin C. Y. Knowledge-based question answering //Proceedings of the Sixth World Multiconference on Systems, Cybernetics, and Informatics (SCI-2002). 2000.

[6]. Jurafsky D.,Martin J.H. Speech and language processing. – New Jersey: Pearson Prentice Hall, 2008. – 1031 с.