



OSTIS-2015

(Open Semantic Technologies for Intelligent Systems)

UDK 681.3.

РАЗРАБОТКА СРЕДСТВ ПЕРСОНИФИКАЦИИ ИНТЕЛЛЕКТУАЛЬНЫХ WEB-ПРИЛОЖЕНИЙ

Рогушина Ю.В.

Институт программных систем НАН Украины, Киев, Украина

ladamandraka2010@gmail.com

Предложен основанный на онтологическом анализе подход к персонификации интеллектуальных Web-приложений, который направлен на извлечение и использование знаний об информационных потребностях и особенностях восприятия информации пользователем. В качестве примера рассматривается семантический поиск в Web.

Ключевые слова: онтология, знания, семантический поиск, персонификация, самообучение.

Введение

В последнее время большинство прикладных разработок в сфере ИТ непосредственно связано с развитием интеллектуализации и персонификации различных приложений. В развитых интеллектуальных информационных системах (ИИС) поддерживается интеллектуальный адаптивный интерфейс с пользователем, который для более эффективного взаимодействия способен учитывать как контекст взаимодействия, так и особенности конкретного пользователя. Другим важным аспектом семантических технологий проектирования интеллектуальных систем является персонификация контента, предлагаемого различным пользователям.

Для этого возникает необходимость получения и обработки знаний – как о самих пользователях, так и о специфике той предметной области (ПрО), к которой относится решаемая пользователем при помощи ИИС конкретная задача.

Постановка задачи

Для того, чтобы разрабатывать эффективные средства персонификации интеллектуальных информационных систем, необходимо определить, какие именно сведения о пользователе могут улучшить функционирование ИИС; из каких источников можно извлечь эти сведения; с помощью каких методов доступные сведения могут быть преобразованы в пригодные для автоматической обработки знания; в какой форме их представить, чтобы облегчить их повторное использование; и как именно структурировать получаемую информацию

о пользователе с учетом предметной области решаемой задачи. Для решения этих вопросов предлагается использовать онтологическую модель информационного взаимодействия пользователя с ИИС, ориентированную на привлечение знаний из внешних онтологий, тезаурусов и распределенных ресурсов Web.

Источники персонифицированных сведений о пользователе ИИС

Рассмотрим подробнее, из каких источников ИИС может получать персонифицированные сведения о пользователях.

В первую очередь, это те сведения, которые пользователь сам явным образом сообщает о себе ИИС – при регистрации или затем в процессе взаимодействия с системой. С одной стороны, это наиболее простой для ИИС способ получить такие сведения, так как при этом основная часть усилий приходится на того, кто хочет воспользоваться ИИС. В таком случае разработчикам надо лишь предусмотреть, какая именно информация может быть полезна для адаптации ИИС к потребностям пользователя, и обеспечить механизм для ее ввода – соответствующие анкеты, тесты и т.п. к недостаткам этого способа можно отнести его трудоемкость для пользователя и недостаточную достоверность сведений, предоставляемых пользователем (например, пользователь может зависить свою квалификацию или неправильно понять вопрос).

Более объективные сведения о пользователе можно получить в процессе его тестирования. В частности, тесты позволяют определить психофизиологические особенности пользователя,

связанные со способностью к восприятию информации (например, экстраверт и интроверт по-разному оценивают различные формы представления информации), его скорость восприятия информации, способность к пониманию сложных информационных объектов (например, различные критерии легкости чтения) и эрудицию. Но при таком подходе пользователю придется потратить достаточно много времени для предоставления этих сведений. Это целесообразно только в том случае, если пользователь намерен одновременно взаимодействовать с ИИС либо если для него очень важно получить от ИИС результат, учитывающий его персональные особенности.

При персонификации ИИС важно правильно оценивать и учитывать для дальнейшей работы психофизиологические свойства пользователя, влияющие на его способность к восприятию информации

В большинстве ИИС эти свойства не используются, поэтому рассмотрим ее более детально. Сейчас существует много психологических тестов, позволяющих достаточно четко классифицировать личность (например, тест Айзенка, ММРІ), а затем на основе такой классификации охарактеризовать (с точки зрения формы представления информации, а не собственно контента) те ИР, которые в большей степени могут удовлетворить такого пользователя. Например, достаточно распространена классификация личностей по скорости возбуждения и торможения, в соответствии с которой выделяют четыре типа – холерик, сангвиник, флегматик и меланхолик, подразделение людей на экстравертов и интровертов и т.д. Социопсихофизиологические исследования обеспечиваются рядом компьютерных систем, которые используются для получения индивидуального адаптационного социопсихофизиологического портрета человека в различных сферах его жизнедеятельности и рассматривают проблемы создания для пользователя комфортного индивидуального адаптационно-информационного пространства [Кузьмина и др., 2012].

Такие индивидуально-типологические свойства, как сила, подвижность, уравновешенность нервных процессов, тип реагирования, эмоциональный фон, скоростные свойства анализаторных систем, особенности высших психических функций (восприятие, внимание, память, мышление, интеллект); поведение в стрессе, тип оптимальной работоспособности, могут в значительной мере определить выбор пользователя при восприятии информации – и, соответственно, при оценке предложенных ИР.

Например, для экстраверта положительными чертами ИР являются наличие иллюстраций, использование коротких простых предложений, эклектичный дизайн, применение ярких цветов для выделения важных элементов и т.д., а для интроверта такими положительными чертами

являются сложная структурированность текста, сложноподчиненные предложения, пастельная цветовая палитра. Поэтому можно предположить, что близкие по контенту ИР с различным оформлением по-разному будут оценены экстравертами и интровертами. Включение в число анализируемых при индуктивном выводе свойств психофизиологических параметров пользователя позволит находить более точные обобщающие правила.

Еще один источник персонифицированной информации – социальный Web. Сведения о пользователе могут быть извлечены из различных социальных сетей, Wiki-ресурсов, блогов и т.д. кроме того, ряд сведений о пользователе может быть извлечен из профилей пользователя в других ИИС. Это связано с тем, что даже при решении совершенно разных задач основные свойства личности пользователя (характер, темперамент, скорость реакции) остаются неизменными (или изменяться во времени пренебрежимо медленно), и потому допустимо повторное использование накопленных знаний в других системах. Кроме того, анализ того, какими именно приложениями оперирует пользователь и какие запросы к поисковым системам он выполняет, тоже может расширить представления ИИС о пользователе.

Но в различных ИИС похожие характеристики пользователя могут описываться различными терминами. Для обеспечения интероперабельности персональных данных можно использовать онтологический подход, интегрируя сведения на основе сопоставления онтологий различных ИИС.

Важным источником персонифицированных сведений о пользователе может послужить опыт его взаимодействия с ИИС. Именно в этих сведениях неявным образом содержатся различные предпочтения и требования пользователя, конкретизированные именно для данной ИИС. Именно поэтому извлеченные из таких сведений знания наиболее полезны для повышения эффективности работы ИИС. Например, если пользователь регулярно отказывается от каких-то рекомендаций ИИС, имеющих ряд общих свойств, то выявление этих свойств позволит в дальнейшем не предлагать ему такие рекомендации, что сэкономит и время пользователя, и вычислительные ресурсы ИИС. Так, если при покупке одежды в Интернет-магазине пользователь всегда отказывается от красных и зеленых товаров, то в результате анализа ИИС перестанет предлагать ему предметы таких цветов, даже если они соответствуют явно заданным условиям пользователя о размере и цене. Чтобы извлечение таких сведений осуществлялось эффективно, важно, с одной стороны, разработать в ИИС эффективные методы самообучения, а с другой – предусмотреть удачное структурирование сведений об обрабатываемых информационных объектах (если у товара отсутствует параметр «цвет», то даже самые совершенные методы самообучения не смогут

выделить правила, относящиеся к группе предметов зеленого цвета).

Именно онтологии могут послужить основой для такого структурирования. Значительную часть терминов, используемых для описания обрабатываемых в ИИС объектов, можно извлечь из онтологии соответствующей Про, а знания о связях между различными элементами – из онтологии задачи.

Основные элементы взаимодействия ИИС с пользователем

В наиболее общем виде элементы ИИС относительно взаимодействия с пользователем можно разделить на следующие составляющие:

- интерфейс ИИС с пользователем;
- контент, предоставляемый пользователю;
- хранилище, в котором накапливается опыт взаимодействия ИИС с пользователем;
- модель информационного взаимодействия ИИС с пользователем, обеспечивающая структурное представление знаний об основных элементах ИИС, необходимых для обеспечения потребностей пользователя.

Во всех этих составляющих для персонификации можно применять онтологическое представление знаний – как для внутренних знаний ИИС, закладываемых в нее при разработке, так и для тех знаний, которые импортируются ИИС из внешней среды или извлекаются ИИС из собственного опыта взаимодействия с пользователями.

Так как ориентация на конкретного пользователя и адаптация к его персональным возможностям и потребностям в значительной мере могут повысить эффективность работы ИИС, то целесообразно во всех этих составляющих использовать информационную модель пользователя и усовершенствовать ее на основе самообучения и поиска новых знаний.

Персонификация интерфейса ИИС

В [Курзанцева, 2010] отмечается, что что сейчас много исследований (как украинских, так и иностранных ученых) посвящены по проблемам адаптации информационных систем к пользователям. Одним из наиболее распространенных методов решения этой проблемы является использование модели пользователя, причем увеличение сложности такой модели обеспечивает более высокую степень адаптации, что вызывает рост стоимости создания модели. В частности, такие подходы широко применяют при разработке поисковых и рекомендующих систем.

Модель стратегии построения интерфейса можно представить следующим образом: каждому пользователю предлагается единственная оптимальная конфигурация системы и сервисов, которая выводится из множества ответов

пользователя на вопросы интерфейса.

Если пользователь известен системе (режим 2), то подсистемой моделирования пользователя проводится анализ его последнего сеанса работы с системой и при соответствующих предположениях происходит корректировка модели пользователя.

В [Яковлев и др., 2006] рассматривается использование онтологий при создании модели пользователя ИИС с адаптивным интерфейсом. Данный подход позволяет учесть совокупность понятий и связей между ними, имеющих место при взаимодействии пользователя с ИИС.

Решение задачи построения модели пользователя в предлагаемой постановке представляет собой выявление совокупности тех характеристик индивидуального пользователя (или классов пользователей), которые обеспечивают ИИС способностью приспосабливаться к потребностям, возможностям и целям пользователя. В целом при создании такой модели необходимо учитывать следующие процессы аспекты взаимодействия пользователя с ИИС: 1) накопление в ИИС персональной информации о пользователе; 2) наблюдение за пользователем в процессе работы; 3) определение персональных особенностей (свойств) пользователя; 4) прогнозирование действий пользователя на основе его свойств; 4) проверка предположений на согласованность; 5) использование в работе ИИС разработанных предположений.

При этом сама модель пользователя представлена в виде метаонтологии, содержащей такие понятия, как «информационная система», «требования к интерфейсу», «пользователь», «методы взаимодействия пользователя с системой», «методы формирования модели пользователя» и т.д.

На втором уровне в онтологии присутствуют свойства этих понятий. Например, для понятия «пользователь» это такие характеристики, как

- демографические характеристики;
- знания пользователя об ИИС;
- опыт пользователя при работе с ИИС;
- предпочтения пользователя в интерфейсе;
- цели и планы при решении задачи;
- заинтересованность в результатах;
- когнитивные особенности;
- специфические свойства пользователя (ограничения психологического, физического характера и т. д.).

Часто в качестве средства персонификации интеллектуальных информационных систем (ИИС) рассматривают их адаптивный интеллектуальный интерфейс.

В частности, в [Анохин, 2014] отмечается противоречивых подходов к организации адаптивного человеко-машинного интерфейса в современной литературе. Это обусловлено сложностью ИИС: с одной стороны, сложность

может стать причиной неожиданного и труднопредсказуемого поведения системы, а с другой стороны, именно благодаря своей сложности ИИС способна изменять свое поведение для адаптации к изменяющимся условиям.

Сложность системы представляет собой неформальное понятие, но можно выделить такие его признаки, как отсутствие формального математического описания; сложность наблюдения и управления вследствие большого числа процессов; независимость от субъекта; нестационарность, выражающаяся в динамическом изменении параметров; невозпроизводимость экспериментов.

Традиционно сложность систем описывают через их структурную и функциональную сложность: структурная сложность определяется связями между компонентами системы, а функциональная – динамическим поведением системы.

Наиболее простой мерой структурной сложности системы является количество входящих в нее элементов (если элементы системы образуют иерархическую структуру, то мера сложности может учитывать и количество уровней иерархии). Кроме того, в качестве меры сложности системы может служить и количество связей между ее элементами.

Такие меры достаточно просто использовать, если в основе ИИС лежит онтология: сложность системы зависит от количества терминов онтологии и количества отношений между ними.

Оценить функциональную сложность системы можно также на основании количества ее состояний. Однако определить количество реально достижимых состояний ИИС очень сложно.

Неотъемлемым признаком сложной системы, является адаптивность ее поведения, т.е. возможность целенаправленного изменения ее параметров и структуры в процессе функционирования. В [Rothrock et al., 2002] рассматриваются такие параметры интерфейса, которые могут изменяться в процессе адаптации, как содержание представляемой пользователю информации; форма представления информации и ведения диалога; распределение задач между человеком и машиной.

Адаптивный интерфейс ИИС позволяет системе самостоятельно, в результате анализа состояния и решаемых задач в текущей ситуации, настраивать отображение предоставляемой пользователю информации и доступных пользователю действий, соответствующих его текущим целям и возможностям. Текущая ситуация и решаемая задача являются факторами, порождаемыми объектом управления и средой деятельности.

Адаптивность интерфейса проявляется в фильтрации и упорядочении контента, предлагаемого пользователю, – как в зависимости

от предпочтений самого пользователя, так и под воздействием контекста и внешних факторов (например, продукт выгодно купить именно в данный момент) [Langley, 1999]).

При описании формализованного адаптивного интерфейса возникает методологическая проблема, связанная с выбором критерия успешной адаптации, обеспечивающей более надежную и эффективную работу пользователя. В [Курзанцева, 2011] предлагается в качестве такого критерия использовать время поиска пользователем необходимого элемента, а в [Ходаков и др., 2003] – все усилия пользователя, включая моторные действия и запоминание информации.

Персонализация контента ИИС

Однако интеллектуальный адаптивный интерфейс является лишь одной из составляющих персонализации работы ИИС и касается в первую очередь выбора способов подачи информации пользователю. Другими важными компонентами персонализации является отбор интересных и полезных для пользователя сведений и выбор способов и методов их обработки.

Использование внешних, динамично обновляемых знаний в современных приложениях, особенно – в Web-ориентированных – является на сегодняшний день всем очевидной необходимостью.

Наиболее популярны при этом онтологии. Но их использование имеет как ряд достоинств, так и недостатков. Онтологии целесообразно использовать в ИИС с изменяющейся структурой БЗ – в тех случаях, когда разработчику заранее не известна не только структура знаний, но и свойства этой структуры, и поэтому из априорных сведений о них нельзя выбрать некое более простое средство представления знаний,

Достоинства онтологий широко известны и не требуют дополнительного обоснования:

- интероперабельности и возможность многократного использования в различных приложениях;
- теоретический базис в виде дескриптивных логик, обеспечивающий гарантии выводимости и разрешимости на онтологии;
- наличие большого количества уже созданных ранее онтологий;
- многочисленные языки представления, стандарты и программные средства для обработки онтологий.

Недостатки онтологий менее очевидны, но их необходимо в первую очередь учитывать при создании реальных приложений:

- в общем случае онтологии сложны в обработке и требуют большого времени и больших вычислительных ресурсов для установления выводимости или невыводимости того или иного факта на онтологии;

- значительная часть реальных онтологий (в отличие от «игрушечных» примеров) слишком громоздки для того, чтобы пользователь, не являющийся ее разработчиком, мог свободно вносить в нее изменения даже при достаточных знаниях о ПрО;

- проблема масштабируемости – при увеличении количества понятий в онтологии и связей между ними время обработки увеличивается настолько, что перестает удовлетворять потребностям реальных задач.

Кроме того, следует отметить, что в большинстве исследованиях, связанных с онтологическим представлением знаний, их полные выразительные возможности просто не используются, но так как их обработка все же предусмотрена, то сложность вычислений неоправданно увеличивается.

В то же время достаточно сложно спрогнозировать, какие именно знания о ПрО окажутся существенными при создании той или иной информационной системы, использующей соответствующую онтологию.

Поэтому представляется целесообразным перейти к двухуровневой модели знаний ПрО.

На первом уровне – модель ПрО – знания ПрО моделируются с помощью онтологии $O_{ПрО} = \langle X_{ПрО}, R_{ПрО}, F_{ПрО} \rangle$, где X – множество концептов, R – множество отношений между концептами, F – функции интерпретации концептов из множества X и отношений из R (или набора каким-то образом интегрированных онтологий), на которые не накладывается практически никаких ограничений за рамками дескриптивных логик.

На втором уровне – модель задачи – используется только некоторое подмножество модели ПрО, представляющее собой ее проекцию на стоящую перед пользователем проблему. Например, ряд связей может быть отброшен или заменен каким-то одним отношением «между терминами существует связь», могут быть оставлены только термины, входящие в фиксированное подмножество и связанные с ними, оставлены только иерархические связи и т.п.

Таки образом формируется значительно меньшая и более простая в обработке онтология, которую уже можно более четко отнести к конкретной дескриптивной логике.

Частным случаем такой модели задачи является используемый при семантическом поиске онтологический тезаурус.

Онтологический тезаурус представляет собой функцию от онтологии ПрО.

При поиске информации, в частности, при поиске знаний в Web, большое значение приобретает персонификация поиска, т.е. учет личных особенностей пользователя,

осуществляющего поиск. Корреляции типа личности и предпочтениям, отдаваемым различным формам представления информации, посвящено много научных исследований.

Однако в них предпочтение отдается выявлению связей между свойствами пользователя и формальными характеристиками информационного ресурса.

Но интересно установить связи между свойствами пользователя и тем, каким структурам знаний он отдает предпочтение. Практически в любой ПрО, даже высоко формализованной, можно построить онтологию, отражающие ее структуру, существенно различающиеся как по количеству различных отношений между понятиями, в том числе иерархических, так и по количеству свойств каждого из понятий.

Методы пополнения знаний, используемых для персонификации

Одно из базовых направлений интеллектуализации систем обработки информации связано с их способностью к самообучению. Таким образом, представляется целесообразным включать в состав ИИС подсистему, предназначенную для извлечения правил и закономерностей из накопленного опыта. Пополнение знаний, используемых для персонификации, может осуществляться путем анализа опыта взаимодействия ИИС с пользователями при помощи различных методов Data Mining.

Для этого можно применять различные методы индуктивного и традуктивного извлечения знаний из данных, например, ID3, ACLS, CART и их различные модификации, ориентированные на обработку данных с различными свойствами (количественных, качественных, динамически изменяющихся и т.д.) [Rogushina et al., 2012].

При этом важно отметить, что подобные методы, как правило, предназначены для классификации, а не для кластеризации, и потому они позволяют выявлять закономерности только относительно тех понятий, связанных с пользователем, и их свойств, которые явным образом заданы при описании структуры знаний ИИС – например, при помощи набора соответствующих онтологий. Поэтому эффективность подобных методов самообучения в значительной мере зависит от удачного выбора таких онтологий.

Архитектура ИИС, обеспечивающая персонификацию

В наиболее общем виде архитектура ИИС, обеспечивающая персонифицированное взаимодействие с пользователями в информационном пространстве Web, описывает связи между информационными потребностями

пользователя, которые описываются через стоящую перед ним задачу, и теми информационными объектами (ИО) и информационными ресурсами (ИР), которые ИИС предоставляет пользователю в процессе решения этой задачи.

Во время работы с ИИС в ее базе данных накапливаются сведения о пользователях, и затем на основе их анализа формируются специфические правила, предназначенные для учета персональных особенностей и потребностей пользователя.

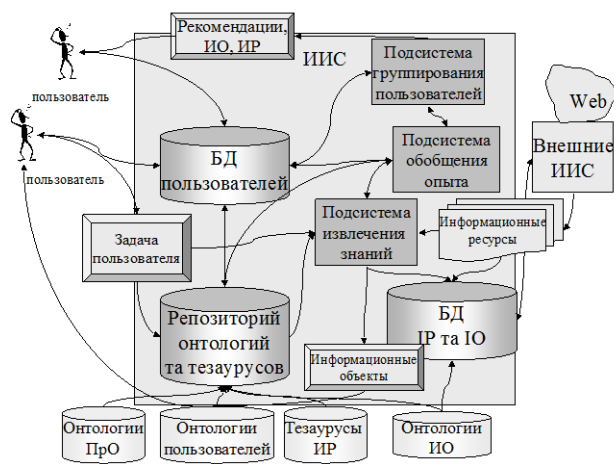


Рисунок – Обобщенная архитектура ИИС

Важным аспектом предложенной выше архитектуры является использование внешних источников знаний – репозитория онтологий и тезаурусов, а также других ИИС, формирующих персонифицированные знания о пользователях в интероперабельной форме. Такой подход, в частности, применялся при разработке знание-ориентированной поисково-рекомендующей системы МАИПС, архитектура которой более подробно описана в [Рогущина, 2014].

Заключение

Предложенный в работе подход к персонификации семантических технологий проектирования ИИС обеспечивает более полный учет индивидуальных особенностей пользователя, а использование онтологий позволяет повторно использовать знания, накопленные ранее в других приложениях.

Библиографический список

- [Кузьмина и др., 2012] Кузьмина К.И., Оноприенко В.Н., Козак Н.С., Семик Т.М., Андон Т.А. Семейная медицина сегодня и проблема ее дальнейшей интеллектуализации с помощью информационных технологий и компьютерных систем // Теорія і практика управління соціальними системами, № 2, 2012. – С.56-67.
- [Курзанцева, 2010] Курзанцева Л.И. О построении интеллектуального адаптивного интерфейса на базе агентной технологии для компьютерных систем широкого назначения // Інформаційні технології та комп'ютерна інженерія, № No1 (17), 2010. – С.16-20.
- [Яковлев и др., 2006] Яковлев Ю.С., Курзанцева Л.И. // О применении онтологии для построения модели пользователя информационных систем // , Комп'ютерні засоби, мережі та системи. 2006, No 5. – С.109-116.
- [Анохин, 2014] Анохин А.Н. Адаптивный интерфейс для

операторов сложных систем, 2014. – <http://vspu2014.ipu.ru/proceedings/prcdn/6345.pdf>.

[Rothrock et al., 2002] Rothrock L., Koubek R., Fuchs F. et al. Review and reappraisal of adaptive interfaces: toward biologically inspired paradigms // Theoretical Issues in Ergonomics Science. 2002. Vol. 3, No. 1. P. 47-84.

[Langley, 1999] Langley P. User modeling in adaptive interfaces // Proceedings of the Seventh International Conference on User Modeling, Banff, Alberta, Canada, June 20-24, 1999. Springer, 1999. P. 357-370.

[Курзанцева, 2011] Курзанцева Л.И. Методика комплексного исследования адаптивного человеко-машинного интерфейса // Математичні машинні системи, № 4. 2011. – С.69-77.

[Ходаков и др., 2003] Ходаков В.Е., Ходаков Д.В. Адаптивный пользовательский интерфейс: проблемы построения // Автоматика. Автоматизация. Электротехнические комплексы и системы, № 1(11), 2003. – С. 45-57.

[Rogushina et al., 2012] Rogushina J., Gladun A. Ontology-based competency analyses in new research domains // Journal of Computing and Information Technology. V.20, N. 4, 2012. – P.277-293.

[Рогущина, 2014] Рогущина Ю.В. Знание-ориентированные средства поддержки семантического поиска в Web // Матеріали IV міжнародної науково-технічної конференції «Открытые семантические технологии проектирования интеллектуальных систем» OSTIS-2014, Минск, БГУИР, 2014. – С.339-352.

KNOWLEDGE-ORIENTED MEANS OF SEMANTIC SEARCH INTO THE WEB

Rogushina J.

Institute of software systems of National Academy of Sciences Ukraine, Kiev, Ukraine

ladamandraka2010@gmail.com

For developing of effective personification means of intelligent information systems (IIS) it is necessary to describe what user information can improve the functioning of the IIS; the information sources; available methods; the form of knowledge presentation that facilitates its reuse; and the structure of user and domain knowledge. Ontological model of information interaction between user and IIS designed to attract knowledge from external ontologies, thesauri and distributed Web resources was proposed.

Keywords: semantic search, ontological model, thesaurus, inductive inference, recommending systems.