

А.Л. ТУРЕВСКИЙ

РАЗРАБОТКА ПРОГРАММНЫХ СРЕДСТВ ПОДДЕРЖКИ КОНЦЕПТУАЛЬНОГО МОДЕЛИРОВАНИЯ СЛАБОСТРУКТУРИРОВАННЫХ ПРОБЛЕМНЫХ ОБЛАСТЕЙ

Индустрия производства программного обеспечения на протяжении уже более четверти века является одной из наиболее быстрорастущих отраслей промышленности. Обладая выгодным соотношением доли необходимых интеллектуальных и материальных затрат, данная сфера позволяет наиболее полно реализовать интеллектуальный потенциал и деловые способности человека, не требуя значительного начального капитала. Постоянно увеличивающийся спрос на вычислительную технику и программное обеспечение привлекают в эту сферу множество крупных и мелких производителей, желающих реализовать свои идеи и получить доход. Эти и другие факторы привели к возникновению в отрасли информационных технологий множества несовместимых средств представления данных и знаний, что еще больше усугубляет наблюдающуюся в последние десятилетия проблему лавинообразного роста их количества.

Область управления информационными ресурсами находится на переднем крае современных исследований и в силу этого в ней ощущается недостаток основательных теоретических наработок. Объектом исследования являются наиболее общие закономерности построения информационных структур реальной и формализованной действительности, которые до последнего времени являлись, в основном, объектом изучения философии, психологии и других гуманитарных наук. Следует отметить их значительные достижения в области изучения реальных систем управления информацией (в первую очередь - человеческого интеллекта [1, 2] и др.), а также общих принципов целостности картины проблемной области (ПО). Наиболее ярко последнее направление изложено в [3]. К сожалению, эти идеи только в последние годы нашли своё применение в практических разработках. Также на этом этапе были получены ценные результаты в области оптимизации взаимодействия человека и внешней среды (см., например, [4] и [5]), послужившие основой для выработки принципов оптимального вмешательства человека в процессы управления информацией ([6] и др.). При этом параллельно и, как правило, независимо, развивались основанные на технических задачах и средствах принципы построения автоматизированных информационных систем. В результате была создана солидная теоретическая и практическая база [7, 8], однако слабо приспособленная для описания сложных компонентов реальной действительности.

Следующим важным этапом развития исследований стала попытка имитации интеллектуальных принципов управления информацией традиционными математическими и техническими средствами. Исследования в области кибернетики [9] и смежных дисциплин, а также дальнейшие разработки в различных областях искусственного интеллекта ([10-12] и др.), продолжающиеся и сейчас, дали значительные частные результаты, однако не смогли решить главной задачи - создать техническую систему, способную заменить человека при осуществлении интеллектуальной деятельности в сложных, слабо формализуемых областях. Фактически, интеллектуальные системы "научились" работать с относительно простыми проблемными областями, но оказались не в силах справиться с более сложными, где потребность в их применении ощущается намного острее. В то же время значительные усилия по приближению средств машинного моделирования к структуре реальной действительности были предприняты и в прикладных областях. В частности, возникла и нашла широкое применение концепция объектно-ориентированного подхода к разработке и созданию программных систем [13, 14], позволившая значительно повысить адекватность машинного моделирования.

В ходе дальнейшего развития исследований была выявлена причина такого положения. Было выдвинуто предположение ([15-18] и др.) о необходимости принципиально иного подхода к рассмотрению информационных потоков в ПО и их представления в "сознании" (как человеческом, так и, в переносном смысле, машинном) познающего субъекта. Основой такого подхода, по-видимому, может стать выдвинутая еще в начале века выдающимся украинским учёным В.И. Вернадским концепция ноосферы [19], обосновывающая необходимость учета на современном этапе гармоничного и неразрывного сочетания и взаимодействия всех существующих в природной среде факторов: как гео- и биологического или космического характера, так и возникших вследствие деятельности человека, воспринимая последние как закономерную часть природных процессов. Основой нового подхода, получившего название "системологического", является понимание системы с учётом как её места и назначения в структуре системы ПО, так и её внутреннего устройства, рассматриваемых в диалектическом единстве и взаимозависимости. Следует отметить, что в силу своей новизны системологический подход нашел своё применение лишь в небольшой части современных разработок (см. [20]), и потребуется, по-видимому, длительное время для его повсеместного внедрения.

В свете новых подходов особую важность приобретают вопросы адекватного понимания и, следовательно, моделирования структуры рассматриваемой ПО. Отсутствие наработанного математического аппарата описания такой структуры и нетрадиционная для точных наук постановка

самого вопроса сдерживают развитие разработок в этой области. Однако большой интерес, проявляемый в последнее время к этой проблеме со стороны различных научных направлений (см., например, [21]), позволяет надеяться на скорые положительные сдвиги.

Также представляют значительный интерес работы в области изучения способов и форм представления накопленных знаний о ПО.

В ходе развития исследований в области искусственного интеллекта вопросы представления знаний и данных получили множество интересных алгоритмических находок, однако всегда были одним из главных препятствий на пути развития интеллектуальных систем. В частности, представляет интерес традиционный для современных исследований реляционный подход к представлению информации, отличающийся наличием в его основе мощного математического аппарата [7, 8]. С другой стороны, теория искусственного интеллекта достигла значительных результатов в приближении структуры представления знаний к их форме в сознании познающего субъекта [10, 11, 15, 16]. При этом недостаточное внимание уделялось форме представления информации в ЭВМ. На современном этапе ставится задача совмещения этих подходов в целях повышения эффективности их использования.

В связи с этим в последнее время все активнее проявляется необходимость создания многофункциональных систем управления знаниями. Потребности интенсивного информационного обмена требуют ускорения процесса усвоения знаний как человеком, так и автоматизированными устройствами с учетом совместимости формы ее представления в различных средах. Кроме того, в современных условиях никакая система, даже полностью соответствующая предъявляемым к ней на момент создания запросам, не может рассчитывать на продолжительную и эффективную эксплуатацию (являющуюся непременным условием ее экономической окупаемости), если в ней отсутствуют средства адаптации к меняющимся условиям окружающей действительности.

Для решения этих проблем необходима реализация новых средств управления информацией, способных справиться со всем её современным многообразием.

Работа посвящена исследованиям в области средств представления знаний сложной структуры, а также решению вопросов практической реализации программных систем, реализующих такие средства, в автоматизированных системах моделирования на основе ЭВМ традиционной архитектуры.

В процессе разработки большое внимание было уделено обеспечению переносимости оболочки, достигаемой использованием, по возможности, функций ОС, не зависящих от текущей аппаратной платформы. При разработке использовались современные инструментальные средства и

идеологии проектирования и программирования, учитывались требования к эргономическим характеристикам рабочего места оператора и стандартизации системных интерфейсов.

Как было отмечено ранее, принципы и средства представления данных и знаний в интеллектуальных системах должны обладать общностью и независимостью от характера представляемой действительности и его физической либо абстрактно-умозрительной природы.

Для выполнения указанных требований предлагается схема представления моделей ПО, характеризующаяся общностью и минимальной привязкой к предполагаемой структуре и средствам моделирования, и способная представлять различные современные средства и формы формализации знаний.

При проведении разработок, прежде всего, учитывались условия:

1) универсальности – разрабатываемые принципы и средства должны быть применимы для моделирования произвольных проблемных областей независимо от их внутренней структуры;

2) независимости – функционирование систем, построенных на основании разрабатываемой технологии, не должно зависеть от текущего аппаратного и программного окружения;

3) совместимости – модели, созданные в рамках данной разработки, должны одинаково пониматься всеми поддерживающими её средствами в любой программной и аппаратной среде;

4) неограниченности – должны отсутствовать ограничения на структуру и размерность хранимой информации (за исключением, возможно, требований, обусловленных аппаратными рамками);

5) надежности – работа системы должна быть надежной и безошибочной;

6) защищенности – обеспечение защиты содержащейся в системе информации;

7) простоты – принципы концептуальной модели и практической реализации системы должны быть доступны пользователю средней квалификации в необходимых для его практической деятельности объемах;

8) доступности – характер взаимодействия с системой не должен оказывать значительного отрицательного влияния на процесс работы во временном и эргономическом аспектах.

Опишем формы реализации и использования моделей проблемных областей в рамках предложенной системы. Известно [11,12], что всякую систему управления базами знаний можно условно разделить на три уровня представления. На высшем, концептуальном, уровне рассматриваются средства, оперирующие когнитивными понятиями и категориями. Средний, логический, уровень, осуществляющий поддержку высшего,

включает процедуры работы с информационными структурами, представляющими знания, абстрагируясь от их физического представления. И, наконец, нижний, физический уровень, являющийся поддерживающим для логического уровня, осуществляет хранение и первичную обработку информации, представленной на физических носителях.

Разрабатываемые средства должны, прежде всего, обеспечивать единую форму взаимодействия с базой данных на логическом уровне, независимую как от характера физической и программной среды функционирования системы, задаваемой на физическом уровне её представления, так и от моделей, используемых при проектировании высшего, концептуального уровня. Стандартизация и унификация интерфейсов логического уровня позволит использовать в рамках одной интеллектуальной системы различные средства моделирования понятийных структур человека (например, фреймы и продукции), что может значительно повысить качество работы таких систем. Продуктивность такого подхода подтверждена опытом проектирования и разработки многих современных экспертных систем [11, 12, 23].

Другой важной задачей разработки является создание надежных и мощных процедур физического управления знаниями, способных на необходимом уровне обеспечить выполнение запросов средств высших уровней. При этом главной целью разработки должно стать устранение физических ограничений на пути и характер развития форм и методов представления знаний. На основе описанных требований представляется возможным предложить такие способы представления моделей ПО:

1) Любое подмножество моделируемой действительности может быть представлено в виде сетевой иерархической структуры, содержащей объекты четырёх описываемых далее видов. Данное утверждение основывается на выводах, независимо полученных в результате исследований психологов, лингвистов, философов и специалистов других дисциплин, показавших универсальный характер базовых элементов понятийного комплекса человека вне зависимости от среды и объекта познания. Обзор таких исследований был дан, в частности, в [24] и [25]. Там же можно найти ссылки на более подробные исследования.

2) Все сущности ПО, понимаемые как обособленные объекты рассмотрения (независимо от их внутренней структуры и физического содержания), могут быть представлены в виде классов, понимаемых как совокупность хранимой в базе информации об их характеристиках. Достаточность такого представления следует из достаточности для описания объекта характеристик его положения в системе понятий ПО (описанных в виде статичных свойств и динамических отношений с другими объектами), а также описания в некотором виде алгоритмов изменения таких свойств и связей во времени.

3) Описывающие классы характеристики делятся на три вида: свойства, связи и процедуры.

4) Свойство класса есть выраженная в числовом, символьном или каком-либо ином виде характеристика, описывающая признак, принадлежащий классу изначально, как сущности, вне связи с другими объектами и явлениями, и предполагаемый статичным во времени. Величина, характеризующая свойство, может иметь произвольную структуру и размерность.

5) Связь класса с другим классом есть однонаправленное отношение между классами, характеризующее имеющуюся между ними ассоциативную либо иную связь, отношение или другой признак, зависящий от свойств обоих классов.

6) Процедура есть алгоритм изменения состояния (свойств и т.п.) данного класса, а также других объектов вследствие некоторых возникших причин.

7) Класс, свойство, связь и процедура с точки зрения программной реализации являются родственными понятиями и наследуют от общего предка - абстрактного элемента данных - ряд общих свойств. К ним относятся:

а) уникальный идентификатор элемента данных, присваиваемый при его создании и не меняющийся всё время его существования;

б) идентификатор типа данных, представленный универсальным целым значением и содержащий информацию о принадлежности элемента к одному из четырёх типов (класс, свойство и т.д.) и подтипу (например: целое, строковое свойство) данных;

в) имя, понимаемое как индекс элемента таблицы атомарных значений, в которой содержатся универсальные строковые значения, не зависящие от текущих региональных установок;

г) идентификатор родительского элемента (для зависимых типов данных содержит идентификатор класса-владельца), набор признаков элемента данных (флаговое значение), характеризующих особенности представления их в этом элементе.

8) Предполагается достаточность определенных выше четырех типов структур для описания модели любой ПО, независимо от её физической сущности и внутреннего содержания.

Следует отметить сходство предложенной модели с некоторыми наиболее развитыми средствами представления знаний в интеллектуальных системах. Видимо, наиболее близкой к ней является фреймовая парадигма. Основным преимуществом разработанных средств перед имеющимися является возможность реализации на основе разработки большинства имеющихся форм представления знаний статического (фреймы, семантические сети) и динамического (продукции и др.) характера. Для них характерно отсутствие (в отличие от фреймовой модели)

строгого ограничения на разнообразие типов объектов моделирования и отношений между ними, а также (в отличие от иерархических систем) ограничения многообразия связей их нециклическим характером.

При проектировании системы был разработан алгоритмический язык специального назначения, текущая версия которого реализована в данной работе. Основные концепции предложенного языка:

1) Язык является процедурным средством общего назначения, расширенным для работы с данными и знаниями.

2) Все элементы данных языка являются полями записей базы и не могут существовать вне ее.

3) Обеспечивается базовый универсальный уровень интерфейса с пользователем без привязки к особенностям используемой аппаратной и программной среды.

4) Все ключевые слова традиционных языков заменены комбинациями специальных символов, что обеспечивает независимость от родного для пользователя естественного языка.

5) Базовым элементом языка является процедура записи базы данных/знаний, понимаемая как алгоритм изменения состояния (свойств и т.п.) данного класса, а также других объектов ПО вследствие некоторых возникших причин. В частности, событие как вид процедуры есть механизм реакции и адаптации объектов к некоторому изменению глобального состояния ПО, понимаемого как совокупность состояний всех составляющих его сущностей.

Рассмотрим физический уровень представления информации. Информационная база представляет собой обобщенную схему структуры моделируемой проблемной области, представленную в виде ориентированного графа. Узлами такого графа являются описания рассматриваемых в данном контексте объектов или понятий, сопровождаемых описаниями присущих им статических и динамических свойств, выраженных в структурно-количественном или процедурном виде, а дуги помечены значениями, описывающими типы имеющихся между узлами взаимосвязей.

Кроме того, в базе хранится дополнительная информация, позволяющая полнее формализовать свойства основных объектов. К ней относятся таблицы атомарных значений, переадресации, прав доступа, а также системные таблицы, предназначенные для хранения сведений о размещении информации в оперативной памяти и на внешних носителях.

Для более полного понимания структуры и возможностей системы вначале опишем форму представления в системе используемых ею элементарных типов данных. Базовым понятием системы, как и любой другой информационной компьютерной системы, является понятие числа, в частности, целого числа (на текущий момент понятие вещественного числа не имеет особой реализации в рамках системы и определяется исходя из требований аппаратного и программного окружения).

Целочисленные значения, а также глобальные идентификаторы элементов, флаговые значения и другие типы данных, сводимые к целому, представляются в следующем виде: поле размера - 4 байта, хранящие размер информационного поля в байтах, затем 1 байт признака знака (0 - отрицательное, любое другое значение - положительное) и информационные байты в количестве, определяемом полем размера. Биты в поле размера и информационном поле должны следовать в порядке "младшие-первыми". Максимальное количество информационных байтов составляет, таким образом, $4 \cdot 294 \cdot 967 \cdot 296$, что позволяет представлять числа в диапазоне приблизительно от $-10^{(10^9)}$ до $+10^{(10^9)}$. Такая размерность целочисленных значений практически снимает проблему фиксирования ситуации целочисленного переполнения и ограничений разрядной сетки, что позитивно скажется на общей надежности и эффективности программных систем. Кроме того, она упрощает представление в системе флаговых и системных значений, а также глобальных идентификаторов, способствующих снятию реальных ограничений на размер и количество хранимой в системе информации.

Основным элементом базы является таблица данных/знаний, хранящая основное информационное содержание системы. Она представлена в виде двухуровневого списка следующей структуры:

- 1 уровень: таблица пулов (наборов в терминах CODASYL) - в памяти представляется динамическим массивом описателей пулов, хранящих информацию об их состоянии и размещении в системе.

- 2 уровень: пул (набор в терминах CODASYL) - каждый пул хранит информацию о состоянии и размещении некоторого подмножества взаимосвязанных данных/знаний.

Двухуровневая система представления информации позволяет обеспечить группировку взаимосвязанных элементов данных/знаний в целях оптимизации временных характеристик доступа к ним.

Таблицы переадресации представляют собой списки соответствий глобальных идентификаторов элементов знаний, используемых при адресации в системе, и идентификаторов их физического положения в базе. С их помощью достигается возможность физического переупорядочения информации в системе без необходимости коррекции глобальных идентификаторов в ссылочных элементах базы.

Таблицы атомарных значений реализуют метод оперирования символической естественно-языковой информацией в компьютерных системах. Как известно, одним из существенных препятствий на пути распространения ПО был и остается языковой барьер. Системы, в которых не предусмотрены специальные средства поддержки локализованных версий, требуют значительных материальных и временных затрат для их приспособления к нуждам

иноязычных пользователей. В данной разработке предпринята попытка создать универсальный путь преодоления такого барьера.

Основой предложенной транслингвистической концепции является введение нового понятия строковой величины не в виде линейного набора информационных (8-, 7- или 16-битовых) значений, что традиционно для современных диалоговых и информационных систем, а в виде унифицированного смыслового комплекса (УСК), являющегося отражением соответствующего данному лингвистическому термину абстрактного и обобщенного понятия, существующего в сознании некоторой общности людей. В компьютерных системах УСК выступает в виде набора представлений соответствующего понятия на некотором множестве естественных либо формальных языков, вызывающих интерес у потенциального круга пользователей системы. Кроме того, возможно внесение в систему базового представления УСК на некотором универсальном языке представления знаний, обеспечивающего однозначность трактовки понятия независимо от контекстной и лингвистической ситуации. Однако разработка такой концептуальной модели является сложной комплексной задачей в области психологии, теории информации, других смежных дисциплин, и требует отдельных глубоких исследований. В свою очередь, представления данных на каждом языке должны обеспечивать возможность описания отдельными непересекающимися областями всего множества символьных наборов языков народов мира для облегчения идентификации использованного при создании сообщения набора символов. На данном этапе представляется возможным принять в виде универсальной системы кодирования символов 16-битовый набор UNICODE, используемый корпорацией Microsoft в ее операционной системе Windows NT.

Основной областью применения разработанной системы является автоматизация создания концептуальных классификационных моделей в экспертных системах различного назначения. Предоставляемые системой средства позволяют выражать в машинно-ориентированном виде сложные структуры объектов и отношений, характерные для таких моделей. Возможность применения системы как информационной основы интеллектуальных средств показана опытом её использования при построении системы концептуальных классификационных моделей в научно-учебной лаборатории Приобретения знаний. Другой важной сферой применения системы является использование её в учебном процессе вузов как наглядного образца объектно-ориентированной (а также фреймовой, продукционной и т.п.) системы, позволяющего углубить и закрепить знания студента, полученные в ходе изучения объектно-ориентированной методологии программирования и проектирования и дисциплин, относящихся к теоретическим и практическим вопросам искусственного интеллекта.

Ещё одной сферой применения является разработка коммерческих систем управления базами данных и знаний, ориентированных на использование в областях, где представляется нерациональным сведение информационных структур к реляционному типу. В частности, к таким областям относятся базы данных, содержащие слабоформализованную информацию, каждый элемент которой может иметь уникальную структуру и значительную размерность.

В процессе разработки системы были выделены два основных направления её дальнейшего развития. Первым из них является повышение интеллектуальных возможностей программного средства в сфере автоматизации труда специалиста по знаниям, работающего с системой. Второе необходимое направление развития дальнейших разработок - усовершенствование формы хранения и алгоритмов обработки классификационных данных в ЭВМ. Особого внимания требует развитие специализированных расширений разработанной системы - механизмов событийного управления информацией, новых подходов к организации пользовательского интерфейса на основе модели «равноправного взаимодействия», алгоритмов контроля целостности и восстанавливающей коррекции данных, создания гибкого языка запросов к базам данных и знаний, способного работать со слабоструктурированной естественно-языковой информацией, средств визуализации структур данных и знаний, алгоритмов разделения доступа пользователей к информационным ресурсам. Также должна быть обеспечена независимость программного интерфейса и формата хранения данных от аппаратного и программного окружения системы и сохранение принципа ее открытости, что позволит в дальнейшем обеспечить её сопряжение с приложениями - потребителями информации. Этому способствуют открытость и унификация программного интерфейса, стандартизация используемых системой средств ОС и внешней среды.

Большой интерес представляет проблема внедрения новых принципов в имеющиеся информационные системы. Учитывая универсальных характер предложенных решений, рационально внедрять такие решения как базовую модель нижнего уровня, способную обеспечить эмуляцию используемых современными системами средств работы с данными. При этом будет обеспечена как перспектива развития имеющихся систем, так и возможность их прозрачного сопряжения с новыми разработками на основе предложенной модели.

В результате работы были изучены средства представления в автоматизированных системах концептуальных моделей сложных слабоструктурированных ПО. Выбрана форма представления таких моделей, основанная на понятиях "класса", "свойства", "связи" и "процедуры".

Разработанные средства позволяют оперировать как статическими (подобными фреймам), так и динамическими (процедурными) объектами. Создана программная система поддержки формирования и ведения баз знаний и данных, реализующих такие модели. Учтены требования надежности, защищенности и эффективности программных средств. Разработанная система может найти применение в системах управления базами данных и знаний различного назначения.

Список литературы: 1. Шалютин С.М. Искусственный интеллект: гносеологический аспект. М.: Мысль, 1985. 199 с. 2. Лингвистическая прагматика и общение с ЭВМ. Издание АН СССР. М.: Наука, 1989. 142 с. 3. Вернадский В.И. Живое вещество, два синтеза космоса // В сборнике: Начало и вечность жизни. М.: Сов. Россия, 1989. 702 с. 4. Перешков В. П., Савинов В. А. Толковый словарь по информатике. М.: ФизС, 1991. 543 с. 5. Венда В.Ф. Инженерная психология и синтез систем отображения информации. М.: Машиностроение, 1982 г. 344 с. 6. Коуте Р., Вильямик П. Интерфейс "человек - компьютер". М.: Мир, 1990. 501 с. 7. Толковый словарь по искусственному интеллекту. М.: РИС, 1992. 254 с. 8. Мартин Дж. Организация баз данных в вычислительных системах. М.: Мир, 1978. 616 с. 9. Грей П. Логика, алгебра и базы данных. М.: Машиностроение, 1989. 368 с. 10. Винер Н. Кибернетика или управление и связь в живом и в машине. М.: Сов. радио, 1958. 11. Осуа С. и др. Приобретение знаний. М.: Мир, 1990. 303 с. 12. Осуа С. Обработка знаний. М.: Мир, 1989. 292 с. 13. Буш Г. Объектно-ориентированное проектирование с примерами применения. К.: Дialeктика, 1992. 519 с. 14. Страуструп Б. Язык программирования C++. К.: Дialeктика, 1992. Тт. 1, 2. 15. Гаврилова Т. А., Червицкая К. Р. Извлечение и структурирование знаний для экспертных систем. М.: Радио и связь, 1992. 200 с. 16. Мельников Г.П. Системология и языковые аспекты кибернетики. М. Сов. радио, 1978. 368 с. 17. Шрейвер Ю. А., Шаров А.А. Системы и модели. М.: Радио и связь, 1982. 152с. 18. Основы системологии: Учебное пособие / М.Ф. Бондаренко, Е.А. Соловьева, С.И. Маторин. К.: УМК ВО, 1996. 80 с. 19. Соловьева Е. А. Теория понятийных знаний. К.: УМК ВО, 1990. 80 с. 20. Вернадский В.И. Научная мысль, как планетарное явление. // В сборнике: Начало и вечность жизни. М.: Сов. Россия, 1989. 702 с. 21. Соловьева Е. А., Бондаренко М. Ф., Маторин С. И., Павлов П. Ф. Проектирование банков естественноязыковых знаний. К.: УМК ВО, 1992. 136 с. 22. Материалы симпозиума "Модели мира". // В сборнике: Новости искусственного интеллекта, 1996 № 1. М.: АПИ, 23. Попов А.С. и др. Статические и динамические экспертные системы. М.: РИС, 1992. 284 с. 24. Соловьева Е.А., Маторин С.И. О моделировании системы понятий деловой прозы. // В сборнике: «Базы данных и знаний в автоматизированных региональных системах». К.: Наук. думка, 1991. С. 164-177. 25. Бурлит А.С., Кузнецов В.И. Системная организация баз данных и баз знаний // В сборнике: «Базы данных и знаний в автоматизированных региональных системах». К.: Наук. думка, 1991. С. 149-156.

Поступила в редакцию 25.11.98