

**Распределенные системы САПР на основе XML
как эффективное средство организации современного
проектирования изделий**

Кулаков В.Н., Кочуров В.А.

Белорусский национальный технический университет

В настоящее время практически все большие программные системы являются распределенными: в них обработка информации сосредоточена не на одной вычислительной машине, а распределена между несколькими компьютерами. Важность этой темы осознают все. Действительно, начальный этап локальной компьютеризации уже пройден. Многие задачи малой и средней автоматизации решаются адекватными средствами на достаточно высоком технологическом уровне. Но вот задачи совершенно иного качества - задачи создания корпоративных информационных систем - нуждаются в осмыслении и анализе.

Сложность нынешнего этапа во многом предопределена традиционализмом и инерционностью мышления, выражающейся в попытке переноса средств и решений локальной автоматизации в мир распределенных систем. Этот мир живет по своим законам, которые требуют иных технологий. В большинстве же случаев преобладает стремление использовать знакомые, понятные, испробованные и поэтому родные средства для решения новых задач, принципиально отличающихся от того, чем приходилось заниматься раньше. Требуется новый подход – создание распределенных систем.

Распределенные системы обладают рядом характеристик, которые определяют требования при их проектировании. Подчеркнем три наиболее важные:

1. *Получение ресурсов.* Проблема получения электронных ресурсов из документальных источников является наиболее острой, особенно в областях знаний, которые содержат огромные массивы информационно-логических таблиц, текстуальных рекомендаций и эмпирических выводов (напр. Машиностроение).

2. *Идентификация ресурсов.* Ресурсы в распределенных системах располагаются на разных компьютерах, поэтому системе имен ресурсов следует продумать так, чтобы пользователи

могли без труда открывать необходимые им ресурсы и ссылаться на них. Примером может служить система URL(унифицированный указатель ресурсов), которая определяет имена Web-страниц.

3. *Коммуникация.* Универсальная работоспособность Internet и эффективная реализация протоколов TCP/IP в Internet для большинства распределенных систем служат примером наиболее эффективного способа организации взаимодействия между компьютерами. Однако в случаях разработки распределенных САПР, когда требуется особая производительность или надежность, возможно использование специализированных средств.

Остановлюсь подробнее на возможных методах решения первой задачи. Эволюция средств представления инженерных знаний происходит в направлении перехода от традиционных носителей информации к электронным.

Ключевыми вопросами при создании хранилищ инженерных знаний являются:

1. Способ представления знаний - комбинация из текста, формул, таблиц и рисунков (эскизов). Данный способ хранения можно назвать традиционным и наиболее предпочтительным для инженерных знаний. Возможны и другие методы представления данных, но выбранный способ обладает главным преимуществом – доступностью и понятностью конечному пользователю такой системы – инженеру.

2. Обеспечение формализации данных, которая позволяет одновременно работать с одним и тем же элементом данных, как человеку, так и программе. Формализация данных выступает как один из основных принципов построения любых современных систем, поскольку давно минули те времена, когда проекты создавались как “вещь в себе”. Формализация данных сегодня позволяет организовать доступ к информации на SQL - универсальном языке запросов, использовать для хранения данных СУБД большого масштаба - Oracle, SQL-Server и др., получать доступ к данным через Internet, строить на основе созданных баз распределенные приложения и т.д.

3. Обеспечение обратимости преобразования инженерных знаний в базу данных. Этот факт означает, что в любой момент мы можем получить исходную таблицу, текст либо рису-

нок без потерь какой-либо информации. Это очень важный аспект, так как программный комплекс, хранящий базу знаний, может одновременно предоставлять информацию пользователю в привычном для него виде, и обрабатывать эту же информацию в удобном для себя виде – внутреннем формате представления данных.

4. Создание развиваемой структуры данных, поддерживающей не только хранение и обеспечение доступа к данным, но и имеющей возможности к дополнению и расширению.

Относительно новой тенденцией в данном процессе является создание XML(Extensible Markup Language)-документов. Сегодня XML может использоваться в любых приложениях, которым нужна структурированная информация. Эта технология обладает целой совокупностью свойств, необходимых для разработки информационных систем, в том числе и распределенных. Отметим наиболее полезные для решения нашей задачи:

- Технология XML позволяет описывать данные произвольного типа и используется для представления специализированной информации, например химических, математических, физических формул, медицинских рецептов, нотных записей, и т.д. Это означает, что XML может служить не только мощным дополнением к HTML для представления "нестандартной" информации, но и средством создания информационных систем для такой информации.

- Процесс создания XML-документа очень прост и требует лишь базовых знаний HTML и понимания тех задач, которые нужно выполнить, используя XML в качестве языка разметки. Таким образом, у нас появляется уникальная возможность определять собственные команды, позволяющие им наиболее эффективно определять данные, содержащиеся в документе. Сам по себе XML не содержит никаких тэгов, предназначенных для разметки, он просто определяет порядок их создания. Набор тэгов может быть легко расширен.

- XML считается самой современной технологией копирования данных или передачи их из одной системы в другую. При помощи этого универсального языка разметки приложение может сгенерировать поток данных в формате XML, который сможет прочитать любое другое приложение, способное понимать XML-контекст. Более того, приложение, которое читает

файл или поток XML, может работать под управлением любой операционной системы, включая и операционные системы минимальной конфигурации, используемые в карманных калькуляторах.

Таким образом, язык XML полностью удовлетворяет условиям для создания распределенной информационной системы.

Ключевая составляющая информационной системы, построенной на основе концепции XML – пространство имен, которые будут использоваться при создании тегов. Создание унифицированного пространства имен является серьезным препятствием при проектировании распределенной системы для любой области деятельности. В самом деле, представьте, наличие нескольких групп (иногда даже конкурирующих), разрабатывающих распределенные системы для одной сферы деятельности, способно привести к такому плюрализму мнений, что создание единого подхода к решению таких задач станет невозможным. Созданные распределенные системы будут совершенно несовместимыми, Из-за различий в пространстве имен и отличии систем идентификации ресурсов никакой интеграции данных быть не может. Примеры таких ситуаций в области информационных технологий можно привести множество.

Полагаем, что в данной ситуации оправданно использование доказавшей собственную состоятельность классификации – УДК.

УДК является единой интегральной системой, а не суммой отраслевых, частных, локальных схем. Вся сумма человеческих знаний и практики рассматривается в УДК как некоторая общность взаимосвязанных, взаимозависимых понятий, которая подразделяется по единому принципу на классы, разделы, подразделы и т. д. по принципу их наиболее частого применения в практической деятельности человека.

Универсальная десятичная классификация в целом характеризуется несколькими основными свойствами. Само название говорит о двух из них: универсальности и десятичности. Кроме того, следует отметить многоаспектность, как заложенную в структуре основной таблицы, так и возникающую при применении вспомогательных таблиц и приемов образования индексов, несущих в себе элементы стандартизации. В многочисленных разделах этой системы упорядочено множество понятий по всем

отраслям знаний или деятельности. Иными словами, УДК охватывает весь универсум знаний. УДК универсальна и в применении. Благодаря обилию средств и приемов индексирования, легко сокращаемой дробности, она успешно применяется для систематизации и последующего поиска самых разнообразных источников информации.

Таким образом, используя систему УДК можно организовать эффективное и универсальное пространство имен XML для любой сферы деятельности человека, а используя коммуникационные возможности этой технологии можно построить эффективную распределенную систему.

В результате реализации обозначенных концепций на кафедре САПР БНТУ была разработана подсистема поддержки пространства имен XML на примере тематического раздела УДК “Обработка резанием (снятие стружки). Резка (разделительные операции без образования стружки). Дробление и измельчение. Обработка листового материала. Изготовление резьбы и т.д. Способы (технология), инструменты, машины и приспособления”

Эта подсистема представляет собой базу даннь. Фактическое хранилище имен тегов XML – таблица этой базы, содержащая следующие поля:

- Поле `UDKCode`. Данное поле хранит непосредственно сам код УДК, который используется исключительно для представления его в приложении, поддерживающем пространство имен.
- Поле `TagXML`. Ключевое поле данной таблицы хранит тег XML и тем самым обеспечивает создание непосредственно пространства имен. Важнейшее требование к данному полю – его уникальность обеспечивается не только разработчиком базы данных, но и контролируется программно в режиме администраторского использования базы при добавлении новых записей
- `InnerCode`. Поле, используемое системой поддержки пространства имен для преобразования совокупности записей (`Recordset`) в древовидную структуру. Данное поле является уникальным и соответствует десятичной системе Дьюи. Для идентификации записей используется десятичное основание и два поля для каждого уровня. Таким образом, общее ко-

личество дочерних узлов для любого понятия достигает 100. При добавлении новой записи обновление поля InnerCode происходит автоматически на основании родительского поля и количества уже имеющихся дочерних полей.

- DescriptionN. Поле, содержащее текстовое описание понятия. Данное поле используется прежде всего для понятийного поиска в базе данных

Таким образом, созданная база данных достаточно проста в реализации и позволяет создавать по своему подобию словари тегов из других разделов УДК. Это один из способов создания унифицированного пространства имен при построении сложных информационных распределенных систем

Совместное использование созданной системы поддержки пространства имен (сервера XML-тегов), сервера баз данных SQL Server и специально созданного расширения для текстового редактор MS Word позволило создать трехзвенную систему для автоматизированной обработки текстов, содержащих знания. Дальнейшее развитие этой идеи и возможно сочетание таких систем с системами искусственного интеллекта позволит создавать системы автоматического либо автоматизированного семантического анализа текстов.

Литература

1. Кочуров В.А. Особенности реализации баз данных и знаний автоматизированного проектирования методами современных информационных технологий.: - Мн, 2002.
2. Кочуров В.А. О проблеме принятия проектных решений в САПР.: - Мн, 2003.