

# ПРЕДИСЛОВИЕ

В настоящее время наиболее развитые страны мира находятся на завершающей стадии индустриального этапа развития общества и перехода к следующему — информационному — этапу развития общества. Широкое применение современных средств доступа к информационно-вычислительным ресурсам открыло принципиально иные возможности использования информационных технологий и вычислительной техники. В условиях современного динамичного развития общества, усложнения технической и социальной инфраструктуры информация становится таким же стратегическим ресурсом, как традиционные материальные и энергетические ресурсы. Современные информационные технологии, позволяющие создавать, хранить, перерабатывать и обеспечивать эффективные способы представления информационных ресурсов потребителю, стали важным фактором жизни общества и средством повышения эффективности управления всеми сферами общественной деятельности. От уровня использования информации сегодня зависят успешное экономическое развитие и конкурентоспособность региона как на внутреннем, так и на внешнем рынке. Информация является важнейшим стратегическим ресурсом, и наибольший экономический и социальный успех сегодня сопутствует тем странам, которые активно используют современные средства компьютерных коммуникаций и сетей, информационных технологий и систем управления информационными ресурсами. Перенесенные на электронные носители информационные ресурсы приобретают качественно новое состояние и становятся активными. Доступная для оперативного воспроизводства средствами компьютерной обработки информация является важнейшим фактором социального развития общества.

Осознание мировым сообществом роли информации как стратегического ресурса стимулировало разработки новых технологий как для получения и переработки больших объемов информации, так и для ее хранения и предоставления пользователям. Стремительное развитие глобальных информационных и вычислительных сетей ведет к изменению фундаментальных парадигм обработки данных, которые можно охарактеризовать как переход к распределенным ресурсам и созданию инфраструктуры для свободного доступа к ним. На первое место в современном информационном обществе выходят технологии использования распределенных информационно-вычислительных ресурсов. В западной литературе несколько лет назад даже появился новый термин “*GRID-технологии*” — технологии создания и использования распределенных информационно-вычислительных ресурсов.

На самом деле идея использования распределенных ресурсов далеко не нова. Она была сформулирована еще на заре компьютерной эры — здесь можно вспомнить Вэннивера Буша<sup>1</sup>, который в своей книге “*As We May Think*” (1945) описал концепцию гипертекста. В дальнейшем идея распределенной обработки информации стала активно развиваться с появлением первых суперЭВМ второго поколения — это советская (российская) ЭВМ БЭСМ-6 и американская ILLIAC-IV. Именно на этих машинах впервые были поставлены эксперименты и создано соответствующее программное обеспечение по использованию распределенных вычислительных ресурсов. В дальнейшем технология массивного счета была перенесена и на “мэинфреймы” третьего поколения. Недаром одним из

---

<sup>1</sup>Вэннивер Буш (Vannevar Bush) (1890–1974) — основатель Национального научного фонда (NSF) США, создатель “дифференциального анализатора” (1930) — релейной машины, способной решать дифференциальные уравнения; руководитель работ по созданию первых ЭВМ.

лозунгов создателей технологий GRID является следующий: *“Вперед (назад) к мэинфреймам”*.

Второй этап развития технологий использования распределенных ресурсов можно связать с появлением сети Интернет. В 1989 году в CERN’е перед сотрудниками лаборатории ECP (Electronics & Computing for Physics) была поставлена задача разработать систему, позволяющую обеспечить доступ к любым сведениям, содержащимся в сети, и объединение всех имеющихся в ней источников информации для сотрудников, приезжающих из различных уголков мира. Иными словами, создать универсальную технологию доступа к распределенной разнородной информации. Таким образом, родилась технология “World Wide Web” (“Всемирная паутина”). Создаваемая система должна не только поддерживать произвольные гипертекстовые ссылки, но предоставлять поиск по распределенным в сети базам данным, но что самое главное — быть многопользовательской, платформенно-независимой системой. Дальнейшее развитие технологии WWW немного подкорректировало начальные установки, и из системы доступа к разнородным ресурсам она превратилась во всемирную “презентационную” систему.

Развитие концепции предоставления удобного и разветвленного поиска ресурсов, распределенных по сети Интернет, привело к созданию протокола Wide Area Information Server (WAIS), замыкающего триаду технологических средств, которую возглавляют Gopher и WWW. Данный протокол обеспечивает функционально полный набор информационных инструментов для Internet. Если Gopher (или сейчас более модный его собрат LDAP) — это виртуальная файловая система, где в качестве ее элементов могут использоваться не только файлы различных форматов и каталоги, но и виртуальные объекты в виде поисковых критериев, а WWW — распределенная мультимедийная система, то WAIS реализует концепцию распределенной информационно-поисковой системы и распределенной системы доступа к ресурсам. Система базируется на универсальном протоколе WAIS, в основе которого лежит американский стандарт Z39.50.

Стандарт Z39.50 — один из протоколов семейства OSI<sup>2</sup>, который описывает прикладной уровень взаимодействия распределенных информационно-поисковых систем<sup>3</sup>. Протокол определяет механизм информационного обмена в процессе обработки поисковых запросов и протокол обмена данными в системах, осуществляющих поиск. Основная область применения протокола в настоящее время — это библиотечные системы и системы научнотехнической информации. Однако область применения протокола значительно шире перечисленных приложений: он может использоваться в информационно-поисковых системах общего назначения. При разработке протокола подразумевалось, что он будет описывать порядок обмена информацией между пользователями информационной системы и ее ядром через сеть передачи данных. При этом сами системы могут управлять данными, используя разные модели и различные языки манипулирования этими данными. Таким образом, информационно-поисковую систему можно построить на основе любой системы управления данными, будь это обычная файловая система или объектно-ориентированная СУБД.

Для разработки распределенных информационных систем сегодня используются различные технологии (RPC, DCOM, RMI, ODBC, JDBC, CORBA, SOAP, eXML, WSDL, WSFL, UDDI и др.). Большая часть этих технологий по своей сути ориентирована не на

---

<sup>2</sup>OSI — Open System Interconnection.

<sup>3</sup>В отечественной литературе за термином “Information Retrieval System” закреплено понятие “информационно-поисковая система”. Стандарт Z39.50 посвящен организации механизма поиска в компьютерных сетях распределенных информационных ресурсов с использованием ключевых слов.

работу с информационными ресурсами, а на сетевое взаимодействие программ и распределенные вычисления в гетерогенных средах. Однако ни одна из упомянутых технологий не обеспечивает универсальных способов работы с информационными ресурсами, т.е. необходимого уровня абстрагирования от конкретных систем и платформ при сохранении высокой степени функциональности и предоставления механизма доступа к данным.

В основе протоколов семейства OSI находится поддержка метаданных — формализованных знаний о внутренней структуре и поведении информационных и вычислительных ресурсов (или документов): структурированные сведения о ресурсе, представляющие его свойства (атрибуты) и функции, т.е. информация, предназначенная для анализа, дизайна, развития, обработки и использования документов<sup>4</sup>. Понятие метаданных уже давно и успешно используют в таких контекстах, как информационные хранилища и системы аналитической обработки данных, электронный документооборот и управление потоками работ, управление знаниями. Этот стандарт позволяет описывать классы объектов, представляющие типы информационных документов, атрибуты и отношения между различными типами документов и ресурсов.

В моделях распределенных вычислений метаданные являются основным структурным и функциональным описанием ресурса, использование которого и обеспечивает унифицированный доступ. В рамках стандартов OSI уже разработан целый набор схем метаданных для описания документов.

Основная задача, поставленная разработчиками GRID, — это работа по объединению всех ресурсов сети Интернет в единую интегрированную среду распределенных ресурсов (GRID, ИРИС<sup>5</sup>), которая составит информационно-вычислительную инфраструктуру будущего. GRID можно определить как исходно распределенную систему, которая сводит воедино данные, вычислительные мощности и ресурсы для представления данных. Единый интерфейс должен предоставлять доступ ко всем необходимым ресурсам так, словно мы имеем дело с одним огромным “метакомпьютером”. Все задачи, как традиционные для обычных компьютеров (управление процессами, памятью, файловой системой, вводом/выводом и пр.), так и принципиально новые/старые (учет, контроль, способ доступа и распределение ресурсов, обеспечение безопасности, совместная работа над набором данных в реальном масштабе времени и пр.), должен решать специализированный комплекс программного обеспечения на базе соответствующей аппаратной инфраструктуры.

Разработчики технологии GRID (термин не является аббревиатурой, это английское слово “сеть, решетка”<sup>6</sup>) составляют отдельное сообщество, которое использует собственную терминологию<sup>7</sup>. Уже из названия можно увидеть, что они делают заявку на “второй Интернет” — подход концепции GRID претендует на глобальность.

---

<sup>4</sup>Под понятием “документ” мы понимаем следующее: в информационном пространстве все информационные ресурсы: события, факты и любые другие сущности реального или виртуального мира — существуют только в форме некоторых информационных объектов. В этом смысле информационный ресурс, снабженный метаописанием, превращается в документ, который является основным “объектом”, с которым оперирует любая информационная система. Таким образом, документ — это структурированное описание реальной сущности (объекта, факта или понятия, включая алгоритмы и программы). Описания составляют информационное наполнение системы.

<sup>5</sup>ИРИС — Интегрированная распределенная информационная система Сибирского отделения РАН.

<sup>6</sup>Правда, иногда этот термин расшифровывают следующим образом: “Global Resource Internet Distributed”.

<sup>7</sup>Это составляет некоторые трудности для сопряжения с концепцией OSI, работающий на открытых международных стандартах.

Предлагаемый Вашему вниманию специальный выпуск журнала “Вычислительные технологии” включает в себя тексты отобранных докладов, сделанных на IX рабочем совещании по электронным публикациям “EL-Pub2004”.

Совещание было посвящено активно развиваемым информационно-телекоммуникационным технологиям работы в глобальной компьютерной сети Интернет и задачам создания новых вычислительных и информационных ресурсов для науки. Перечисленные направления являются фундаментальными и становятся все более актуальными в настоящее время. Программа рабочего совещания предусматривала работу следующих секций:

1. Технология GRID как основа информационно-вычислительной инфраструктуры будущего. Технологии представления разнородной информации в удобном для конечного пользователя виде.

2. Разработка корпоративных стандартов хранения и представления информации и метаданных. Безопасность информационных систем и защита информации.

3. Научно-образовательные распределенные информационные системы, электронные публикации, библиотеки и коллекции.

4. Применение GRID-технологии для построения интегрированной информационно-вычислительной среды Западно-Сибирского региона.

Цель проведения этого научного мероприятия заключалась в обмене информацией и обсуждении технологий разработки распределенных информационных систем и технологии GRID, объединяющей все ресурсы сети Интернет в единую интегрированную среду распределенных ресурсов, как основы информационно-вычислительной инфраструктуры будущего. Были рассмотрены вопросы направлений виртуального единства пользователей распределенных информационных систем, фундаментальная парадигма обработки данных, вопросы обеспечения (и унификации) доступа к любым данным, содержащимся в сети, концепции предоставления удобного и разветвленного доступа к информационно-вычислительным ресурсам и перспективы их развития. В работе совещания приняли участие ведущие ученые и специалисты из научных центров Москвы, Самары, Уфы, Тюмени, Омска, Новосибирска, Барнаула, Красноярска, Иркутска, Владивостока.

*Чл.-корр. РАН А.М. Федотов*