

В. Е. Асрибеков, И. Ю. Никольская, А. А. Погребков, А. Г. Попов,
Н. С. Авраменко, Л. Ю. Вермишева, Т. А. Диеспелова, Ю. Е. Сливкина

Концептуальное обоснование информационной он-лайн службы по физико-математическим наукам и информатике как альтернативной теоретической базы для разработки концепции GRID-интеграции научной распределенной информационно-вычислительной среды для интернет-ресурсов*

Рассматриваются концептуальные проблемы GRID-интеграции для многокомпонентного и многосвязного комплекса научно-информационных интернет-ресурсов для распределенной информационно-вычислительной среды с фактическим моделированием 3-х компонентного комплекса (ЭИС+издательства+пользователи). Осуществляется выбор формы представлений научно-информационных периферических GRID-инфраструктур в рамках теории преобразований корневых деревьев в "клеточном" пространстве.

1. ВВЕДЕНИЕ

Проблема управления распределенной системой информационных тематических запросов пользователей-исследователей, с одной стороны, и журнальных ресурсов физических издательств, с другой стороны, не могла быть решена на основе линейного принципа построения Реферативного журнала с однократным помещением реферата под так называемой главной рубрикой. Для решения этой задачи в 70-х годах прошлого столетия стала издаваться Сигнальная информация по физике с многократным размещением описаний статей по нелинейному принципу, т. е. одновременно как по разным разделам физики, так и по множественным аспектам статьи внутри данного раздела. Результат интеллектуального анализа каждой статьи несколькими экспертами-физиками (на основе 17 древовидных 6-уровневых рубрикаторов разделов, проблемно переплетенных друг с другом; общее число рубрик более 8 тыс.) выражается с помощью сети кодов рубрик. Интеллектуальный "сетевой образ" в виде закодированного квазиреферата или сложного гиперграфа является базой механизма управления тематическими ресурсами, распределенными в различных статьях и журналах (термин GRID-технология управления ресурсами появился позднее).

А. Критическая ситуация, возникшая к настоящему моменту в связи с катастрофическим сокращением в 25–30 раз поступлений бумажных научных журналов (в т. ч. подписных, опаздывающих, как минимум, на полгода), особенно по физике, математике, вычислительным и другим фундаментальным наукам диктует необходимость перехода в случае фундаментальных наук к использованию

специалистами только журнальных интернет-ресурсов, поскольку в области фундаментальных наук:

- практически все основные западные научные журналы издаются, прежде всего, в электронной форме;
- практически все основные русские научные журналы переводятся на английский язык и обе версии, прежде всего в электронной форме, издаются одновременно, т. е. "языковой барьер" для Запада ликвидирован;
- все электронные научные журналы представлены в Интернете через сайты издательств и через службы OJS (Online Journal Services) с опережением бумажной версии на 1–1,5 мес. (доступ бесплатный к библиографии и, как правило, к аннотациям авторов, но к статьям платный, хотя для информационных центров возможно исключение).

Б. Доступ к любым журнальным интернет-ресурсам для удовлетворения информационных нужд профессиональных физиков, математиков, информатиков и др. специалистов из смежных дисциплин и фундаментальных направлений исследований, включая тошkie технологии, должен осуществляться тотально, оперативно и точно-адресно с ориентацией предшествующей обработкой интернет-ресурсов на возможный поиск современной динамичной проблематики, поскольку на сайтах издательств уже содержатся одновременно библиография (подлежащая глубокой систематизации), аннотации авторов и полные тексты статей.

Для этой цели не могут быть использованы ни УДК (отвергнутая в зарубежных информационных центрах по физике как архаизм еще в начале 70-х годов прошлого столетия), ни разнотипные не

*Статья публикуется в виде, представленном авторами. Редакция не несет ответственности за ее содержание и форму представления.

“нацеленные” на актуальную проблематику — библиотечные классификаторы и спецификаторы, ни ГОСТированные не очень дробные (для 1-кратного помещения рефератов по линейному принципу!) консервативные рубрикаторы традиционного РЖ.

Адекватный научно-информационный метод заключается в иерархическом способе выявления и фиксации “проблемных модулей” в сформированных доменах понятий, образующих виртуальные “сетевые образы” (впоследствии названные GRID-образами), которые обеспечивают прямой “выход” на точноадресные информационные ресурсы новой проблематики.

Как уже говорилось, задача управления распределенной системой информационных тематических запросов физиков-пользователей, с одной стороны, и многопроблемных статей из журнальных ресурсов физических издательств, с другой стороны, не могла быть решена на основе линейного принципа РЖ по физике с 1-кратным помещением реферата под так называемой главной рубрикой.

Для решения этой задачи, как уже отмечалось, стала издаваться Сигнальная информация по физике с многократным размещением библиографии статей по нелинейному принципу, т. е. одновременно как по разным разделам физики, так и по множественным аспектам статьи внутри данного раздела. Именно возникающий “сетевой образ” статьи в виде иерархически закодированной ее проблематики или сложного гиперграфа является базой механизма управления проблемно-ориентированными ресурсами уже в рамках GRID-технологии в адекватном “клеточном пространстве”.

Создавая GRID-базу для переплетенных иерархических информационных систем-деревьев для каждого из основных разделов физики (деление по объектам и по явлениям), можно далее надстраивать этот неплоский граф или соответствующую GRID-базу до древоподобного гиперграфа путем введения новых виртуальных корневых вершин, отражающих появление новых “проблемных модулей”, и одновременно формирования модернизированных динамических доменов с виртуализацией прежних понятий, т. е. уже обновленных научно-информационных ресурсов, управляемых вычислительными кластерами.

В. Имея в виду что переплетенные иерархические информационные системы образуют GRID-инфраструктуры, в рамках интернет-идеологии может быть целесообразна интеграция научной распределенной информационно-вычислительной среды в рамках единой или согласованной мировой GRID-инфраструктуры. В области физики попытки согласования инфраструктур на языке рубрикаторов РФ и Западных стран (Великобритания - INSPEC, США - PACS, Международная версия ICSP) были осуществлены еще в 1998 г. в рамках ICSTI.

Таким образом, GRID-интеграция научно-информационных ресурсов в форме:

- 1) детально виртуализированных и регулярно активизируемых проблемных рубрикаторов по физике и информатике;
- 2) проблемно-обработанных, глубоко систематизированных по рубрикаторам статей из электронных журналов, предоставляемых он-лайн через интернет-сайты автономных издательств и центров по физике и информатике;
- 3) информационных запросов интернет-пользователей Электронных информационных служб (ЭИС), специалистов по физике и информатике создаст единый научно-информационный

комплекс, изображаемый системой иерархических корневых гиперграфов и фактически образующий распределенную информационно-вычислительную среду. Такая ранжированная в рамках графо-образований 1.-3. GRID-система будет многокомпонентной и многосвязной и должна включать в себя как блоки виртуализации научно-информационных интернет-ресурсов на базе перманентно виртуализируемых рубрикаторов, так и необходимые вычислительные кластеры и соответствующие, технологически адекватные, пакеты управления вычислительными ресурсами. При этом реализации схемы GRID-интеграции на основе аппарата системного построения иерархических корневых гиперграфов, наряду с концептуальными разработками предусматривает.

а) усовершенствование способов представления информации по физике и информатике в Электронных информационных службах,

б) детализацию и максимальную автономность системы “скачивания” и нормализации данных из интернет-источников по ряду ведущих профильных журналов, прежде всего, издательств ACM, IEEE, AIP, APS, ELSEVIER и Springer.

Как уже указывалось, результат интеллектуального анализа каждой статьи несколькими экспертами-физиками выражается с помощью сети кодов рубрик или квазирефератов статьи. Именно интеллектуальный “сетевой образ” статьи в виде иерархически закодированного квазиреферата или сложного корневого гиперграфа является основой механизма управления тематическими ресурсами, распределенными в различных статьях и журналах. Таким образом, построение единого графо-комплекса “ЭИС+издательства научных журналов+интернет-пользователи — специалисты по физике и информатике (встраиваемые в создаваемую модель “безбумажного” информационного обеспечения научной литературы)”, который использует как генерируемые в блоки виртуализации научно-информационных ресурсов, связывающих “скачивание” электронных журналов и отдельных электронных статей из издательств, с одной стороны, и контингент исследователей физиков и информатиков, с другой стороны, становится центральной задачей для данной многоуровневой GRID-системы, компактно генерируемой путем наложения древоподобных GRID-инфраструктур. В архитектуре подобной GRID-системы, обеспечивающей точноадресный доступ к информационным “сетевым образам” статей (с низким уровнем шума), возникают дополнительные ассоциативные связи между иерархическими блоками рубрикатора, которые могут быть отражены в сопровождающем предметном указателе как для простого тематического поиска статей, так и для установления проблемно-ориентированных виртуальных связей с “выходом” в итоге на полнотекстовые оригиналы релевантных статей. При этом, с одной стороны, информационные тематические запросы пользователей-исследователей формируются в электронной форме в процессе навигации по гипертексту рубрикатора, а с другой стороны, можно также активно использовать непосредственно тексты статей из журнальных интернет-ресурсов издательства по физике и информатике, т. е. механизм управления тематическими ресурсами заложен в самом построении этой научной специфически распределенной информационно-вычислительной GRID-среды.

2. КОНЦЕПТУАЛЬНЫЕ ПРОБЛЕМЫ GRID-ИНТЕГРАЦИИ ДЛЯ МНОГОКОМПОНЕНТНОГО И МНОГОСВЯЗНОГО КОМПЛЕКСА НАУЧНО-ИНФОРМАЦИОННЫХ ИНТЕРНЕТ-РЕСУРСОВ КАК НАУЧНОЙ РАСПРЕДЕЛИТЕЛЬНОЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ С ФАКТИЧЕСКИМ ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ GRID

Кардинально новая научно-информационная программа, максимально отвечающая требованиям полного электронно-информационного обеспечения исследователей через телекоммуникационные интернет-каналы, теоретически (в начале 70-х годов) и практически (в начале 90-х годов) реализовала один из перспективных способов интеллектуальной виртуализации структурированных научно-информационных ресурсов в области физики на основе некоторой версии “сетевой” технологии, названной впоследствии обобщенно “технологией GRID” и используемой в настоящее время для обычной виртуализации более частных прототипов ресурсов, в том числе стандартных вычислительных ресурсов. Естественно, что структурированные научно-информационные ресурсы, в отличие от стандартных вычислительных ресурсов, должны включать в себя обязательные интеллектуальные элементы (так называемые коды “узлов” на языке теории графов, интеллектуально модернизированных и актуализированных рубрикаторов GRID-системы). Эти интеллектуальные элементы, подвергнутые процедуре виртуализации, отражают закодированные этапы текущего научно-творческого процесса с мониторингом построения новых теоретических моделей, с разработкой существенно новой экспериментальной аппаратуры и, наконец, с последующей генерацией тематически новых глубоко систематизированных информационных массивов силами активно функционирующих специалистов-экспертов, с учетом ранее проведенных виртуализаций. Отметим еще раз, что основная идея новой научно-информационной программы 70-90-х годов прошлого века заключалась в иерархическом способе выявления и фиксации “проблемных модулей” в общепринятой системе понятий глубоко структурированной физики, в виртуализированном рубрикаторе которой динамически формировались домены понятий, образующие виртуальные “сетевые образы” статей, обеспечивающие прямой “выход” на точно адресные научно-информационные ресурсы повои проблематики.

Таким образом, научная распределенная информационно-вычислительная среда в рассматриваемой 3-х компонентной GRID-системе: “ЭИС+издательства научных журналов+интернет-пользователи” должна включать в себя как блоки виртуализации научно-информационных интернет-ресурсов (на базе перманентно актуализируемых, с введением дополнительных тематических блоков, проблемно-виртуализированных рубрикаторов), так и вычислительные кластеры и пакеты управления вычислительными ресурсами с созданием максимально распределенных полей ссылочных структур. В подобной среде будут фигурировать, помимо обработанных научно-информационных ресурсов, “проецированных” на рубрикаторы (по физике и по

информатике), также автономные издательства журналов и широкий контингент интернет-пользователей, который встраивается в создаваемую модель “безбумажного” информационного обеспечения научной литературой. Оптимальная реализация указанной 3-х компонентной “интегрирующей” GRID-системы, использующей генерируемые в блоки виртуализации научно-информационные ресурсы и запросы контингентов исследователей-пользователей “GRID-преобразованной” научной информации (“скачанной” из издательств), решает поставленную задачу.

2.1. Базис комплекса научно-информационных ресурсов на языке аппарата теории деревьев

Интеграция GRID-инфраструктур в форме виртуализованных иерархических рубрикаторов (по физике и информатике) требует постулирования “клеточного” пространства дискретной математики, базис которого может быть реализован на основе остовных и корневых деревьев. При этом адекватная содержанию статьи непосредственная интеграция GRID-инфраструктур порождает наборы однородных графов в виде правильных платоновых тел или тел Архимеда, отвечающих сложному, в нашем случае 3-х компонентному, многосвязному комплексу научно-информационных ресурсов.

Соответственно, указанные однородные графы распадаются на автономные подграфы отдельные блоки 3-х компонентного комплекса, базирующиеся на разных профилях многоаспектной и многопрофильной статьи. Разумеется, профильный поиск осуществляется по разным аспектам в рамках одного из остовных или корневых деревьев, однако ЭИС позволяет устанавливать все гиперсвязи между профилями статьи и определять их смысловой гиперграф.

2.2. Моделирование 3-х компонентного многосвязного комплекса научно-информационных интернет-ресурсов в виде распределенной информационно-вычислительной среды с помощью целочисленных распределений Планка (при наличии “понятия” complexity)

Естественно, что среди трех компонентов комплекса научно-информационных интернет-ресурсов основную роль играет исходный иерархический рубрикатор, на который “проецируются” методами GRID-интеграции (см. раздел 4.4) две остальные компоненты комплекса из издательства журналов и с сайта ЭИС (проблемно-переработанные точноадресные ресурсы), что автоматически приводит к распределенному целочисленному образованию, описываемому с помощью целочисленных распределений Планка (с включением “понятия” complexity и с отказом от вероятностного подхода при использовании римановских суперпозиционных сумм). При этом, если сайты издательств журналов содержат заголовки, резюме и полные тексты оригинальных статей (включая, возможно, индексы их систематизации по западным рубрикаторам), то на сайте ЭИС может быть осуществлен точноадресный тематический поиск статьи с последующим доступом к ее полному оригиналу на сайте соответствующего издательства.

2.3. Гипотеза Улама для деревьев как алгоритм восстановления периферических проблемных рубрик в иерархических рубриках

Используя известный изоморфизм любого дерева совокупности поддеревьев с вычлененной одной периферической вершиной, отвечающей в рубрике проблемной рубрике, т. е. определяя процедуру восстановления дерева из специфической совокупности поддеревьев с удаленной периферической вершиной (гипотеза Улама), можно подойти к решению проблемы создания распределенной ссылочной структуры в рамках концепции GRID-интеграции. Этот путь, в известной мере, может “вывести” на задачу генерирования проблемных рубрик, исходя из статистических закономерностей восстановления деревьев.

2.4. Иерархические древоподобные рубрикаторы в качестве основной компоненты научно-информационных интернет-ресурсов для реализации теоремы перечисления Пойа и построения распределенной информационно-вычислительной среды

Количественные характеристики распределенной информационно-вычислительной среды могут быть установлены с помощью теоремы перечисления Пойа для типовых деревьев с фиксированным количеством вершин, что, в конечном счете, приводит к наборам целочисленных распределений Планка для разделов иерархических древоподобных рубрикаторов, которые могут быть гомоморфно трансформированы в любые иные версии, включая западные рубрикаторы.

3. ВЫБОР ФОРМЫ ПРЕДСТАВЛЕНИЯ НАУЧНО-ИНФОРМАЦИОННЫХ GRID-ИНФРАСТРУКТУР, В ТОМ ЧИСЛЕ ИЕРАРХИЧЕСКИХ ДРЕВОПОДОБНЫХ РУБРИКАТОРОВ, В РАМКАХ ДИСКРЕТНОЙ МАТЕМАТИКИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ТЕОРИИ ГРАФОВ

Различные аспекты реализации научно-информационных иерархических древоподобных рубрикаторов могут быть наиболее естественно представлены именно в рамках концепции GRID-технологий в широком смысле (виртуализация информационных ресурсов разнообразной природы, образование проблемных научно-информационных GRID-инфраструктур с различной степенью иерархического упорядочивания, способы GRID-интеграции виртуальных информационных GRID-инфраструктур и т. п.). С другой стороны, введение этого GRID-представления означает постулирование более адекватной для данного варианта аксиоматики дискретной математики, положенной в основу одного из разделов теории графов, а именно теории корневых и остовных деревьев, использованных еще в 1847 г. Г. Кирхгоффом для анализа электрических цепей. В случае исследования иерархических древоподобных рубрикаторов

целесообразно применение аппарата теории корневых деревьев с рассмотрением процедуры вершинообразования при учете степени (degree) и мощности (power) вершин. Фактографические “таблицы” цифровых данных и понятийные иерархические факсетки (в виде отдельных типовых проблемных блоков) располагаются именно на периферических рубриках — “висячих” вершинах.

3.1. Дискретная проблематизация континуальных (текстовых) информационных массивов статей как результат их виртуальной систематизации на основе иерархических графо-язычных GRID-рубрикатов

Процесс систематизации “континуального” содержания статей на базе детальных проблемных GRID-рубрикатов непосредственно приводит к дискретным научно-информационным массивам (“проблематизация” статей).

Специфика научно-информационных массивов в области физико-математических наук и информатики определяется, прежде всего, строгой логикой тематического ранжирования вводимых понятий и характеризующих их дискретных величин, как элементов дискретных проблемных многообразий, и неизменно иерархическим строем их упорядочивания как в форме однозначного подчинения (“сильная” иерархия), так и многозначного соподчинения (“слабая” иерархия). Наиболее естественно обе формы иерархии изображаются графоструктурами с корнями (“центрами” иерархии) или гиперграфами с множеством корней, каждый из которых определяет свой блок иерархии или свою GRID-инфраструктуру с образованием, в частности, иерархических древоподобных графоструктур или локальных рубрикатов (в том числе научно-информационных).

Согласно Б. Риману, образование понятия дискретной величины возможно лишь в том случае, если ему предпослано некоторое более общее понятие, находящееся с ним, согласно вышесказанному, в отношении иерархии. С другой стороны, в дискретной математике имеет место прерывное многообразие однородных дискретных величин, где количественное сравнение осуществляется посредством счета (в случае непрерывных величин – посредством измерения).

Только после конструирования такого понятийного строя или таких опорных аксиоматических понятий, упорядоченных в GRID-инфраструктуры или в графо-блоки научно-информационных GRID-рубрикатов, можно развернуть именно дискретную демокритовскую схему GRID-представлений или дискретную проблематику указанных наук на языке теории графов. При этом автоматически решается задача “выключения” дискретных проблемных “частей” при сохранении понятия “целого”, что нарушает положение Демокрита о сведении “целого” к сумме “частей”.

Характерно также, что уже в случае обычной GRID-технологии управления стандартными распределенными вычислительными ресурсами надлежащий способ виртуализации ресурсов адекватен фактически переходу к графо-языку в соответствующем “клеточном” пространстве.

3.2. Интеграция проблемных научно-информационных GRID-инфраструктур как простое объединение тематических иерархических блоков виртуализированных рубрикаторов, заполняемых перманентно текущей информацией с образованием распределенной информационно-вычислительной среды

Прямая сумма детально виртуализированных проблемных рубрикаторов, регулярно актуализируемых на сайте ЭИС и проблемно обработанных, глубоко систематизированных по рубрикаторам статей из электронных интернет-журналов, предоставляемых ВИНТИ, создаст тематически локальный фрагмент распределенной информационно-вычислительной среды для осуществления точного адресного информационного поиска (библиографического, но с возможностью доступа к полнотекстовой статье), в соответствии с профессиональной организацией информационных запросов интернет-пользователей ЭИС, специалистов по физике и информатике.

Такая простая схема объединения 3-х компонентов многосвязного комплекса полностью обосновывает концепцию GRID-интеграции научной распределенной информационно-вычислительной среды с фактическим использованием технологии GRID для интернет-ресурсов в области физико-математических наук и информатики. В области физики эта схема успешно функционирует более 12 лет.

3.3. Проблемная модернизация 18 разделов древоподобных рубрикаторов по физике с образованием актуализированных иерархических GRID-инфраструктур, содержащих прямые алгоритмы для их графо-визуализации, для графо-интеграции с мировыми GRID-инфраструктурами и для поисковой графо-навигации по ним

Регулярная систематизация статей по всем основным аспектам и одновременно по всем необходимым профилям иерархического рубрикатора по физике (соответствующих проблемно-виртуализированным GRID-инфраструктурам) выявляет, в процессе мониторинга обработки журналов, новую актуальную тематику, указывающую на направления проблемной модернизации всех 18 разделов, выставленных в настоящее время в ЭИС Ф: Физика элементарных частиц. Теория полей. Физика атомных ядер. Структура и свойства ядер. Ядерные реакции. Ядерная и релятивистская астрофизика. Физика атомов. Физика молекул. Физика плазмы. Физика твердых тел (Электрические, Магнитные и Оптические свойства). Физика ядерных реакторов. Акустика. Радиофизика. Методика и техника ядерно-физического эксперимента. Электрооптика.

Как уже известно, в соответствии с целенаправленным развитием концепции виртуализированных (проблемных) рубрикаторов, основанных на принципах “сильной” и “слабой” иерархии и предназначенных для макроупорядочивания дименов понятий в результате текущей информационной обработки новых журнальных статей, можно

определять в каждом конкретном тематическом интервале оптимальную глубину древоподобного рубрикатора, имея в виду его наиболее целесообразную четырехъярусную структуру:

А. Основ раздел (физика или информатика) — корень дерева

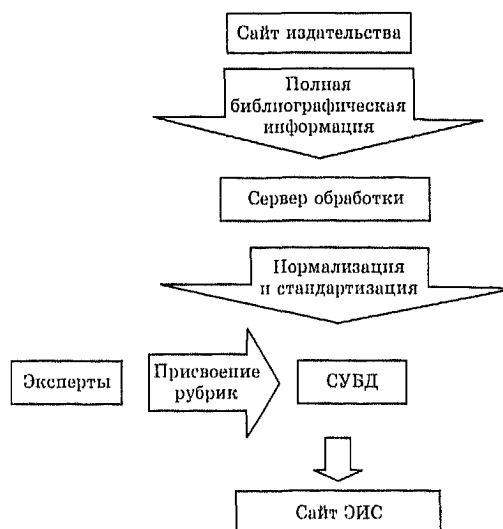
В. Стабильные традиционные подразделы (тематические блоки физики или информатики) и их рубрики и подрубрики с соподчинением по схеме “сильной” иерархии — боковые внутренние стволы и главные ветви “дерева”.

С. Динамические проблемные рубрики и подрубрики, включая периферические рубрики с соподчинением по схеме “слабой” иерархии, — наружные периферические ветви “дерева”, завершающие иерархические “связи” рубрикаторов.

Д. Фактографические таблицы или проблемные фасетки (однородные элементы), надстраиваемые наружные периферические ветви “дерева” и представляемые в совокупности при поисковой графо-навигации по данной GRID-инфраструктуре, с учетом ее локальной проблематики.

Естественно, что проблемная модернизация древоподобного рубрикатора в процессе мониторинга содержания статей из журналов с извлечением новой или модулированной проблематики осуществляется преимущественно в С- и D-ярусах с установлением дополнительных связей по схеме “слабой” иерархии с боковыми внутренними стволами и главными ветвями “деревьев”, что элементарно отражается на языке адекватной иерархической буквенной (латиница) иерархии.

4. ПРИМЕНЕНИЕ ТЕХНОЛОГИИ “ПРОЕЦИРОВАНИЯ” ДЛЯ РАЗЛИЧНЫХ НАУЧНО-ИНФОРМАЦИОННЫХ ИНТЕРНЕТ-РЕСУРСОВ В РАМКАХ КОНЦЕПЦИИ GRID-ИНТЕГРАЦИИ С СОЗДАНИЕМ РАСПРЕДЕЛЕННОЙ ССЫЛОЧНОЙ СТРУКТУРЫ



Принципиальная схема технологии “проецирования” научно-информационных Интернет-ресурсов

4.1. Программы проецирования публикаций Американского института физики

Отличительной особенностью процесса проецирования публикаций Американского института физики является необходимость двупроходного процесса проецирования. На первом проходе собирается оглавление выпуска журнала, на втором — аннотации опубликованных статей. Листинг программ проецирования приводится в Приложении 7 к Итоговому отчету по теме Президиума РАН за 2005 г. [2].

4.2. Программы проецирования публикаций издательства Elsevier

Структура сайта издательства Elsevier позволяет за один запрос получить библиографическую информацию и аннотации к статьям журнального выпуска, что существенно упрощает процесс. Однако структура сайта требует использования более сложного внешнего планировщика загрузки так как имеет древовидный, а не линейный вид. Листинг программ проецирования приводится в Приложении 7 к Итоговому отчету по теме Президиума РАН 2005 г. [2].

4.3. Контроль целостности полученной информации

Так как структура и содержание сайтов издательств постоянно меняются, то проблема контроля целостности полученной информации выходит на первый план. Кроме этого, возможны проблемы с доступностью и работой сайтов, из-за чего возможно неполное получение запрошенных документов. К сожалению, универсального решения данной проблемы не существует, поэтому используются сразу несколько методов. Во-первых, контроль размера. Размер полученного документа сравнивается с размером, указанным в HTTP-ответе сервера (если таковой указан), а также с наперед заданным минимальным значением, выбираемым индивидуально для каждого типа информационного ресурса. Во-вторых, производится проверка наличия стандартных частей документов (например, стандартного окончания каждого документа) и проверка отсутствия сообщений об ошибках сервера. В случае невыполнения какого-либо из условий предпринимаются повторные попытки получения документов через различные промежутки времени и в случае их неуспешного завершения процесс останавливается с сообщением об ошибке. Данный подход дает вполне удовлетворительные результаты на практике, без вмешательства оператора в процесс проецирования.

4.4. Распределение процесса проецирования по различным узлам сети

Процесс проецирования прекрасно распределяется по различным узлам сети с помощью использования внешнего планировщика и удаленного доступа к узлам. Распределение процесса позволяет использовать различные каналы для соединения с сайтами издательств и более полно проводить обработку полученных документов.

4.5. Стандартизация формата получаемых данных

Процесс стандартизации полученной информации сводится к разбору текста страниц сайтов и вычленению стандартных блоков: библиографической информации и аннотации статьи.

Структура получаемой информации:

1. Название журнала
2. Номер журнала
3. Выпуск журнала
4. Дата публикации
5. Название статьи
6. Автор(ы)
7. Страницы
8. Текст аннотации
9. Ссылка на полный текст статьи.

5. УСТАНОВЛЕНИЕ СВЯЗИ ОСНОВНОГО ЭЛЕМЕНТА НАУЧНОЙ РАСПРЕДЕЛЕННОЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ, АДЕКВАТНО ДАННОМУ МОМЕНТУ ВРЕМЕНИ, С ФАКТИЧЕСКИМ ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ GRID ДЛЯ ИНТЕРНЕТ-РЕСУРСОВ

Как хорошо известно, способы общения между учеными, представителями различных наук, уже к настоящему моменту претерпели достаточно радикальные изменения.

5.1. Мониторинг научных исследований в “журнальную эпоху” (бумажный период)

Наиболее эффективное общение между специалистами родственных наук осуществлялось примерно с 17-го века в форме первичной бумажной публикации в одном из периодических научных журналов. Вплоть до конца 20-го века первичная публикация являлась для ученых главным источником новой научной информации, а именно, описанием конечного результата или продукта научного исследования, эксперимента и даже открытия. Естественно, что основной формой научно-информационной деятельности для каждого ученого являлся при этом тематический поиск необходимых публикаций разными способами, включая опосредованную детальную проблемную систематизацию их содержания.

5.2. “Интернет-эпоха” для научных журналов и приоритетная проблемная GRID виртуализация статей (“проблемные модули”) для научно-информационной среды

“Интернет-эпоха” для научных журналов как бы завершает лингвистический период их информационного анализа, поскольку на первый план выступает приоритетная задача их точечно-адресной проблемной обработки с определением наборов тематически релевантных статей. Адекватным механизмом для этой цели служит выявление экспертами-специалистами “проблемных модулей” при GRID-виртуализации статей и последую-

щая их систематизация по древовидному рубриктору. Из сравнения таблиц выпускаемых журналов за 2000 г. и за 2004–2005 гг. можно легко увидеть темпы перехода от “журнальной эпохи” к “интернет-эпохе”, где дополнительно к издательствам журналов начинают подключаться научно-электронные библиотеки (типа НЭБ РФФИ).

5.3. Научные интернет-ресурсы в нежурнальной форме

Автономность отписок отдельных статей в научных журналах “интернет-эпохи” позволит ставить вопрос о дальнейшей “автономизации” тематических или даже фактографических интернет-ресурсов, поступающих от конкретных авторов. Все это должно реализоваться в новых GRID-инфраструктурах на периферии древоподобного рубриктора с образованием “фасеточных блоков”, т. е. может возникнуть задача замены “цепочки” журнал—оттиск—“проблемный модуль” другой, более адекватной, информационной конструкцией.

6. ЗАКЛЮЧЕНИЕ

В заключение необходимо отметить, что основы представленной концепции GRID-интеграции и

некоторые из технологий были успешно применены в разработанном в 2004 г. демонстрационном сайте ЭИС по информатике, который может быть немедленно преобразован в действующий вариант ЭИС по информатике с момента начала финансирования работ по регулярной систематизации массивов статей из научно-информационных интернет-ресурсов по информатике на основе созданного виртуализированного рубриктора. Последующая реализация ЭИС по другим разделам физико-математических наук позволит расширить диапазон концептуальных исследований GRID-технологий.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Асрибеков В. Е., Авраменко Н. С., Вермишева Л. Ю., Дисперова Т. А., Попов А. Г., Сливкина Ю. Е. ВИНТИ в эпоху интернет-обеспечения первоисточниками: роль Электронной информационной службы по физике (ЭИС-Ф) // НТИ. Сер. 1. — 2006. — № 12. — С. 30–32.
2. Итоговый сводный отчет за 2005 г. по программе фундаментальных исследований Президиума РАН: “Разработка фундаментальных основ создания научной распределенной информационно-вычислительной среды на основе технологии GRID”.

Материал поступил в редакцию 18.01.07.