

ПРОГРАММНЫЕ СРЕДСТВА ДЛЯ СОЗДАНИЯ И ПОДДЕРЖКИ XML-ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ HIPNS

А.В. Дугинов, И.Е. Куликов, Е.И. Литвиненко, Т.Л. Пикельнер,
Р.Н. Семенов
Объединенный институт ядерных исследований,
Россия, 141980, г. Дубна Московской области, ул. Жолио-Кюри, 6
litvin@nf.jinr.ru

SOFTWARE TO CREATE AND SUPPORT XML VERSION OF THE INFORMATION SYSTEM HIPNS

A.V. Duginov, I.E. Kulikov, E.I. Litvinenko, T.L. Pikelner, R.N. Semenov
Joint Institute for Nuclear Research, Joliot-Curie 6, Dubna, Moscow region,
Russia, 141980
litvin@nf.jinr.ru

The HIPNS project (the information system on neutron sources and scientific research based on these sources) supported by the Russian Foundation of Basic Research (*grant 01-07-90347*) is under development at the Frank Laboratory of Neutron Physics of JINR. The main distinction of the system is an orientation toward the potential users of neutron instruments, while others are directed to the neutron source specialists. The core of the information system is a relational database, which contains the main parameters of the setups and includes a lot of links to the electronic documents, providing an access to the remote resources and electronic collections. The report presents the current results of XML format application for HIPNS. The work was aimed to the following tasks: XML code generation from the database content, modern design of the resulting pages, tools for HTML code generation from XML/XSL codes and some others. This practical experience could be helpful for other developers because of the fact, that the use of XML is still quite complicated due to the incompleteness of the corresponding software.

Введение

В настоящее время в Лаборатории нейтронной физики им. И.М.Франка ОИЯИ при финансовой поддержке Российского Фонда Фундаментальных Исследований (*грант 01-07-90347*) ведутся работы над проектом по развитию информационной системы по нейтронным источникам и проводимым на их базе научным исследованиям [1-4]. Особенностью разрабатываемой системы является ориентация не на специалистов по источникам нейтронов, а на потенциальных пользователей, работающих на

них установок. Ядром информационной системы является реляционная база данных, которая содержит основные параметры установок и включает в себя значительное число ссылок на электронные документы, обеспечивая доступ к удаленным ресурсам и электронным коллекциям.

В целях наиболее эффективного разделения функциональности между программными модулями, обеспечивающими получение запрошенной информации из базы данных, и модулями, обеспечивающими внешнее представление полученной информации, было принято решение о внедрении в разрабатываемую систему современных XML-технологий.

В настоящей работе представлены текущие результаты работ по применению формата XML [5] для интерфейсных страниц информационной системы HIPNS. Работы проводились при использовании СУБД Microsoft SQL Server 7, не имеющей встроенной поддержки XML. Практический опыт участников данного проекта может представлять интерес для разработчиков других систем, поскольку внедрение формата XML затруднено недостатком соответствующего программного обеспечения.

Средства формирования статических XML-страниц

Авторами работы просматривались возможности генерации XML-страниц из таблиц реляционной базы данных некоторыми коммерческими продуктами, из них наиболее эффективным оказался пакет XML Junction 7.0. При использовании этого пакета возможно сформировать некоторое XML-представление таблицы, содержащейся в базе данных, однако для него отсутствует возможность задать определенную структуру тегов в результирующем файле.

Пример XML-кода, полученного с помощью XML Junction см. в **Пример 1**. Как видно из примера, при преобразовании в XML целые числа (атрибуты "Type" и "Type2") получаются вещественными, в остальном форма представления в основном приемлемая, только для включения стилевых файлов необходимо изменить первые строки.

Помимо просмотра готовых приложений, проводилось изучение нескольких базовых программных пакетов в качестве библиотек процедур по генерации XML-страниц из таблиц реляционной базы данных. Рассматривались Borland Delphi, Borland Kylix, Microsoft Visual Basic.

В частности, с этой целью было написана программа на базе Delphi, позволяющая подсоединиться к удаленной базе данных и провести редактирование таблицы, сохраняя результаты в базе и в XML-файле (Рисунок 1).

В **Примере 2** представлен пример XML-файла, полученного с помощью этого приложения. Вид результирующего XML кода в данном случае совершенно неприемлемый, поскольку вся содержательная информация перенесена в атрибуты тегов. Приложения на базе Delphi и Kylix позволяют в интерактивном или пакетном режиме получить XML-страницы,

однако, в форме, крайне неудобной для дальнейшего использования совместно со стилевыми файлами.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<recordset>
<R1>
  <ID>1002.0000000000053</ID>
  <TYPE>2.000000000000013</TYPE>
  <SID>10.00000000000053</SID>
  <E_MAIL>vondreele@lanl.gov</E_MAIL>
  <COMMENTS>High Intensity Powder Diffractometer: atomic structures
of materials that are available only in polycrystalline or non-
crystalline form.</COMMENTS>
  <WEB><A href=http://www.lansce.lanl.gov/lujan/instruments/
hipd.htm>http://www.lansce.lanl.gov/lujan/instruments/hipd.htm
</A></WEB>
  <FAX>505-665-2676</FAX>
  <MAIN_PERSON>Von Dreele</MAIN_PERSON>
  <INSTRUMENT>HIPD</INSTRUMENT>
  <TYPE2>1.000000000000013</TYPE2>
  <MAINPER_NAME>Robert</MAINPER_NAME>
  <GLINK />
</R1> ...
```

Пример 1

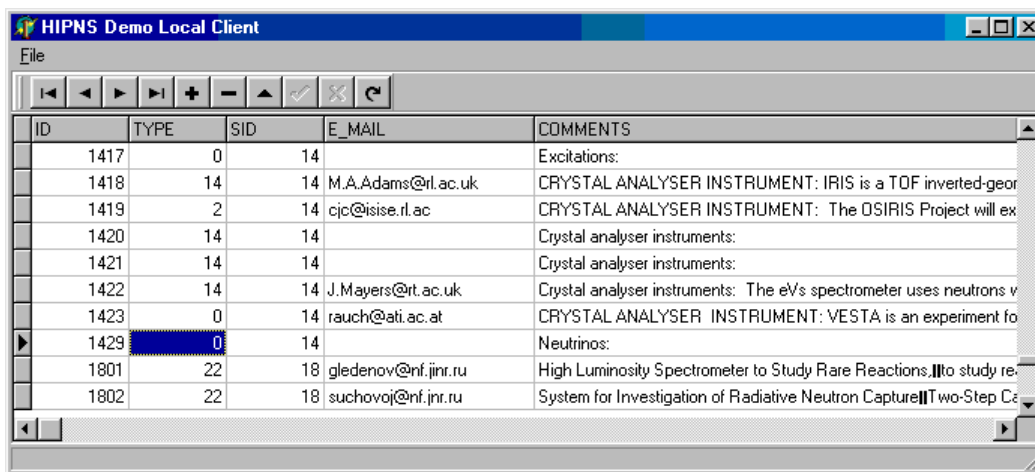


Рисунок 1. Приложение на базе Borland Delphi

Наиболее успешные результаты по автоматической генерации XML-страниц были достигнуты при создании программных модулей на базе динамических библиотек Msxml*.dll, поставляемых вместе с пакетом Microsoft Office XP.

Программные модули в этом случае были реализованы как приложения пакета Microsoft Access и написаны на языке Visual Basic for Applications. Основной библиотечной процедурой, необходимой для генерации

XML-файлов по результатам SQL запроса типа SELECT, является в нашем случае метод ExportXML класса Application.

```
<?xml version="1.0" standalone="yes"?>
<DATAPACKET Version="2.0">
<METADATA>
  <FIELDS>
    <FIELD attrname="ID" fieldtype="i4"/>
    <FIELD attrname="TYPE" fieldtype="i4"/>
    <FIELD attrname="SID" fieldtype="i4"/>
    <FIELD attrname="E_MAIL" fieldtype="string" WIDTH="60"/>
    <FIELD attrname="COMMENTS" fieldtype="string"
WIDTH="2000"/>
    <FIELD attrname="WEB" fieldtype="string" WIDTH="200"/>
    ...
  </FIELDS>
  <PARAMS CHANGE_LOG="1 0.....1465 4 8"/>
</METADATA>
<ROWDATA>
  <ROW RowState="4" ID="902" TYPE="1" SID="9" E_MAIL=""
  COMMENTS="" WEB="&lt;A href=&quot;http://nfdfn.jinr.ru/
  flnph/fks/hrfd.html&quot;&gt;http://nfdfn.jinr.ru/flnph/
  fks/hrfd.html&lt;/A&gt;" FAX=""
  MAIN_PERSON="Pomjakushin" INSTRUMENT="HRFD" TYPE2="1"
  MAINPER_NAME="Vladimir" GLINK=""/>
  <ROW RowState="4" ID="906" TYPE="3" ...
```

Пример 2

Разработанные программы позволяют подсоединиться к удаленной базе данных под управлением Microsoft SQL Server и автоматически сгенерировать набор XML-страниц, отражающих текущее состояние таблиц в базе данных. На следующей странице представлен пример получаемого XML-файла (Пример 3).

Результирующий XML-код соответствует желаемому виду. Для формирования первых трех строк XML-представления использовались специальные приемы потому, что текущая реализация метода ExportXML, как и пакет XMLJunction, не отрабатывает описанное в документации конфигурирование вида представления. Понятно, что при создании собственных программных приложений этот недостаток можно преодолеть.

Для внешнего представления полученных XML страниц были написаны стилевые файлы в формате XSL. Внешний вид полученных страниц при просмотре их браузером Microsoft Internet Explorer 6 для двух вариантов дизайна показан на Рисунке 2.

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml:stylesheet href='Main.xsl' type='Text/xsl'?>
<dataroot xmlns:od="urn:schemas-microsoft-com:officedata">
<TITLE>BERII</TITLE>
<_x0035_>
  <NEU_INSTR.ID>504</NEU_INSTR.ID>
  <NEU_INSTR.TYPE>1</NEU_INSTR.TYPE>
  <SID>5</SID>
  <E_MAIL>stuesser@hmi.de</E_MAIL>
  <WEB>http://www.hmi.de/bensc/instrumentation/instrumente/
    e4.html</WEB>
  <MAIN_PERSON>Stuesser</MAIN_PERSON>
  <INSTRUMENT>E4</INSTRUMENT>
  <TYPE2>3</TYPE2>
  <MAINPER_NAME>Norbet</MAINPER_NAME>
  <NEU_INSTR.GLINK><![CDATA[<A HREF="http://159.93.21.210/
    cgi-bin/hipnsimg.cgi?iid=504">View</A>]]></NEU_INSTR.GLINK>
  <NEU_SOURCES.ID>5</NEU_SOURCES.ID>
  <TOWN>Berlin</TOWN>
  <SOURCE>BERII</SOURCE>
  <INSTITUTE>HMI</INSTITUTE>
  <YEAR_START>1993</YEAR_START>
  <WEB_SOURCE>http://www.hmi.de/bensc/instrumentation/
    berII.html</WEB_SOURCE>
  <WEB_INSTITUTE>http://www.hmi.de/</WEB_INSTITUTE>
  <WEB_INSTRUMENTS>http://www.hmi.de/bensc/instrumentation/
    instrumentation_en.html</WEB_INSTRUMENTS>
  <NEU_SOURCES.TYPE>reactor</NEU_SOURCES.TYPE>
  <POWER_KWT>10000</POWER_KWT>
  <COUNTRY>Germany</COUNTRY>
  <STATUS>http://www.hmi.de/grossgeraete/bensc/
    reactor-cycles_97.html</STATUS>
  <NEU_INST_TYPE.ID>1</NEU_INST_TYPE.ID>
  <TYPE_NAME>Monochromatic Neutron/X-ray Powder Diffractome-
    ter</TYPE_NAME>
</_x0035_>

```

Пример 3

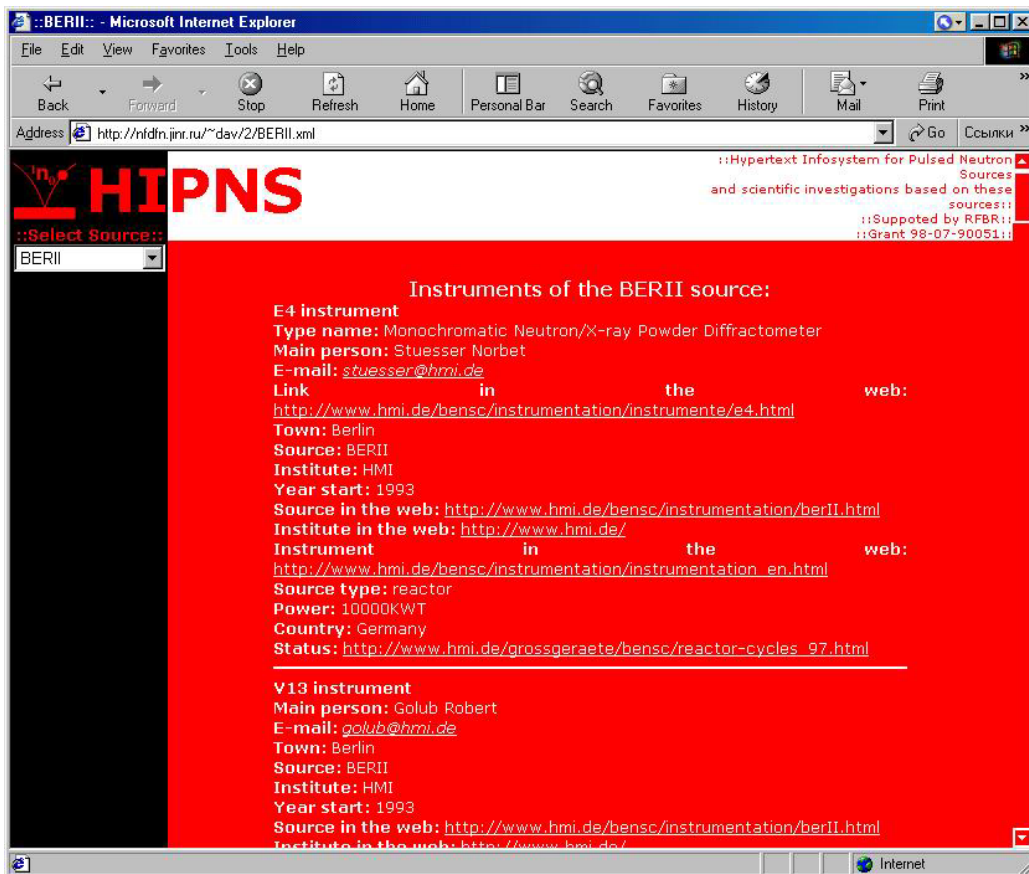
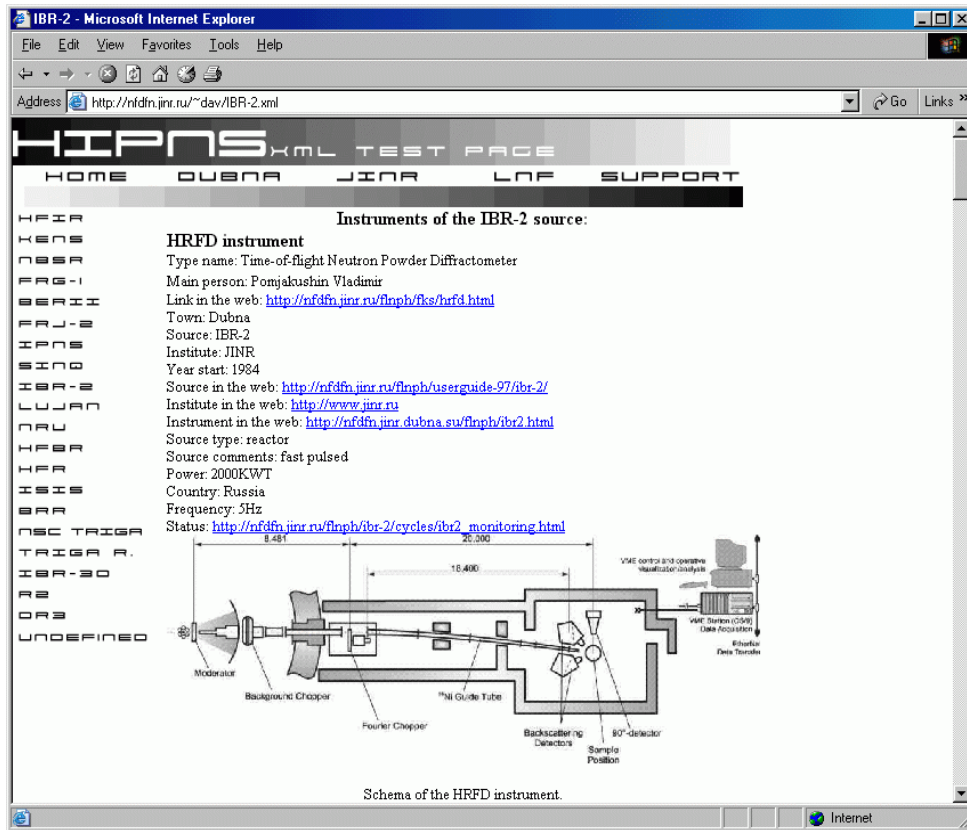


Рисунок 2. Два варианта дизайна интерфейсных страниц системы

Преобразование XML в HTML с помощью Apache Cocoon

На следующем шаге необходимо было обеспечить возможность просмотра полученных страниц браузерами, которые не в полной мере поддерживают формат XML (это Netscape, Microsoft IE ниже версии 5.5 и др.). Для этой цели исследовались варианты серверных приложений, обеспечивающих в первую очередь преобразование XML в HTML.

Известно, что серверные приложения на базе Java-сервлетов (JSP) не вполне подходят для генерации динамических XML-страниц, потому что JSP API не поддерживает организацию конвейеров для сервлетов [7-8]. Наиболее продуктивным подходом в этом направлении представляется подход, реализованный в проекте Apache Cocoon [6], основной идеей которого является разделение этапов производства, обработки и представления информации.

В Cocoon производство исходных данных осуществляется с помощью сервлета, называемого генератором (см. Рисунок 3). Генератор может просто считывать XML-код из файла (File Generator), или формировать XML-листинг директории (Directory Generator), или возвращать XHTML-код, полученный преобразованием HTML-файла (Html Generator), и т.д.

Выходные XML-данные генератора подаются на процессор (Transformer). Процессор проводит преобразование (или цепочку преобразований) этих данных в XML-код, точнее, преобразует входной поток SAX [9] событий в выходной поток SAX событий. Предлагается использовать XSLT Transformer, который при преобразовании обращается к стилевому описанию в формате XSL, SQL Transformer, который переправляет запрос к базе данных и возвращает XML-код результата, и др.

На последнем этапе реализуется представление или форматирование окончательных результатов. Cocoon предлагает такие типы представления, как HTML, XML, текст, PDF, Postscript, формат MS Excel и др. Модули форматирования (Serializers) преобразуют поток SAX событий в поток бинарных или текстовых данных, в ряде случаев просто вызывая для выполнения форматирования компоненты web-сервера Apache.

При использовании Cocoon статические и динамические данные должны размещаться в различных каталогах. В каталогах для динамических данных должны присутствовать конфигурационные файлы, в которых указываются способы обработки и представления данных. Для функционирования компонент пакета Cocoon необходим совместимый с Servlet 2.2 процессор сервлетов. В нашем случае тестирование пакета Cocoon проводилось под Linux, в качестве web-сервера и процессора сервлетов использовался Apache Tomcat 4.0.3 [10].

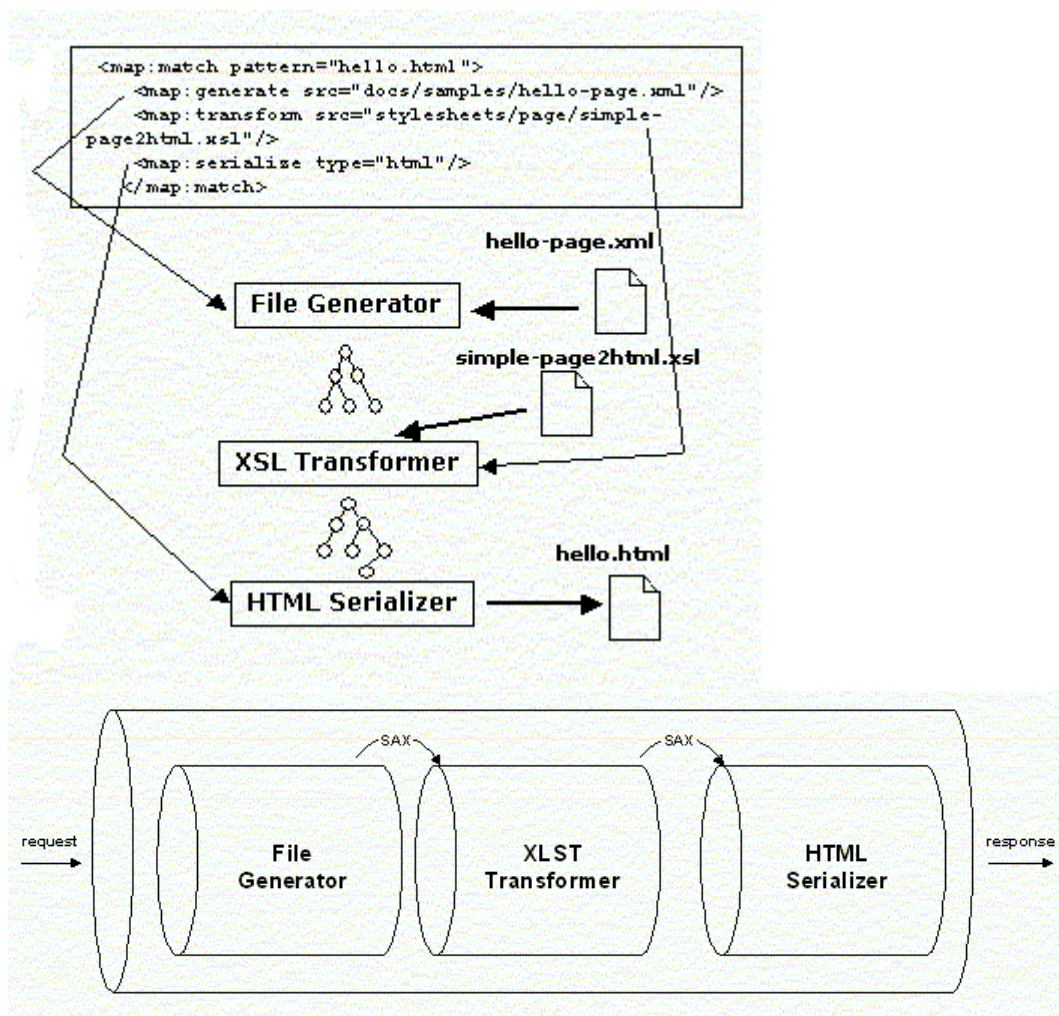


Рисунок 3. Схема получения HTML-представления XML-страницы средствами Apache Cocoon

Наше тестирование показало, что пакет Cocoon достаточно хорошо производит преобразование XML+XSL данных в HTML (см. Рисунок 4), однако пока не позволяет получать в приемлемом виде PDF файлы, несмотря на наличие соответствующего процессора.

Получение HTML кодов при использовании двух шаговой схемы представления данных

На следующем шаге можно поставить задачу еще более детального разделения задач по получению информации и по ее представлению. При этом желательно иметь возможность использования одного и того же дизайна страниц для исходной информации разной природы.

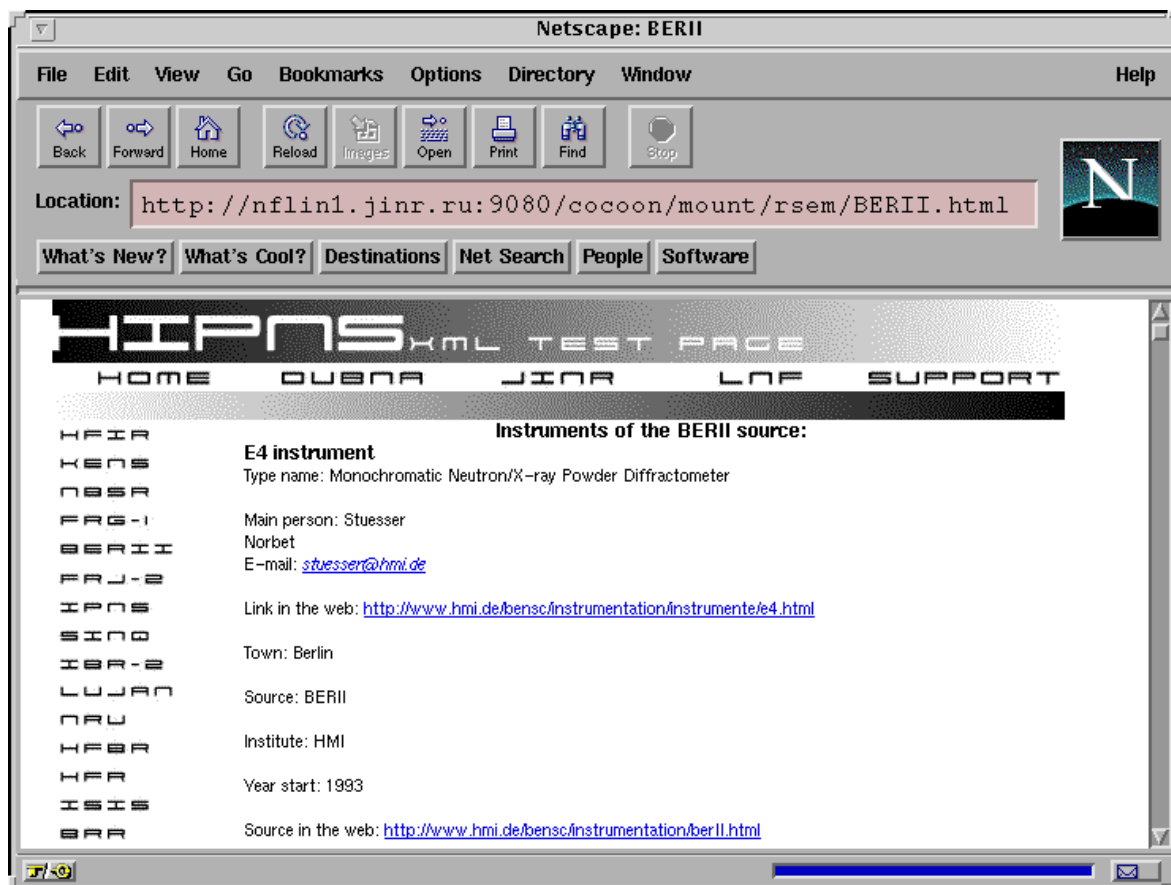


Рисунок 4. Результат преобразования XML/XSL-кодов в HTML с помощью Apache Cocoon

В рамках пакета Apache Cocoon эта идея реализуема при использовании двух-шаговой схемы обработки данных [11]. Достаточно иметь один стилевой файл (XSL) для проблемно-ориентированных XML кодов (и перевода их в экранно-ориентированные коды) и другой XSL для экранно-ориентированных XML кодов.

Другим путем решения этой задачи могут быть специальные дизайн ориентированные приложения. Для нашей системы такое приложение было разработано на языке PHP 4. Пример работы с этим приложением для наших данных представлен на рисунке (см. Рисунок 5)

Заключение

В настоящей работе изложен практический опыт по внедрению XML-технологий. Учитывая острый недостаток русскоязычной литературы по данному вопросу, авторы предполагают, что их опыт окажется полезным разработчикам многих других информационных систем.

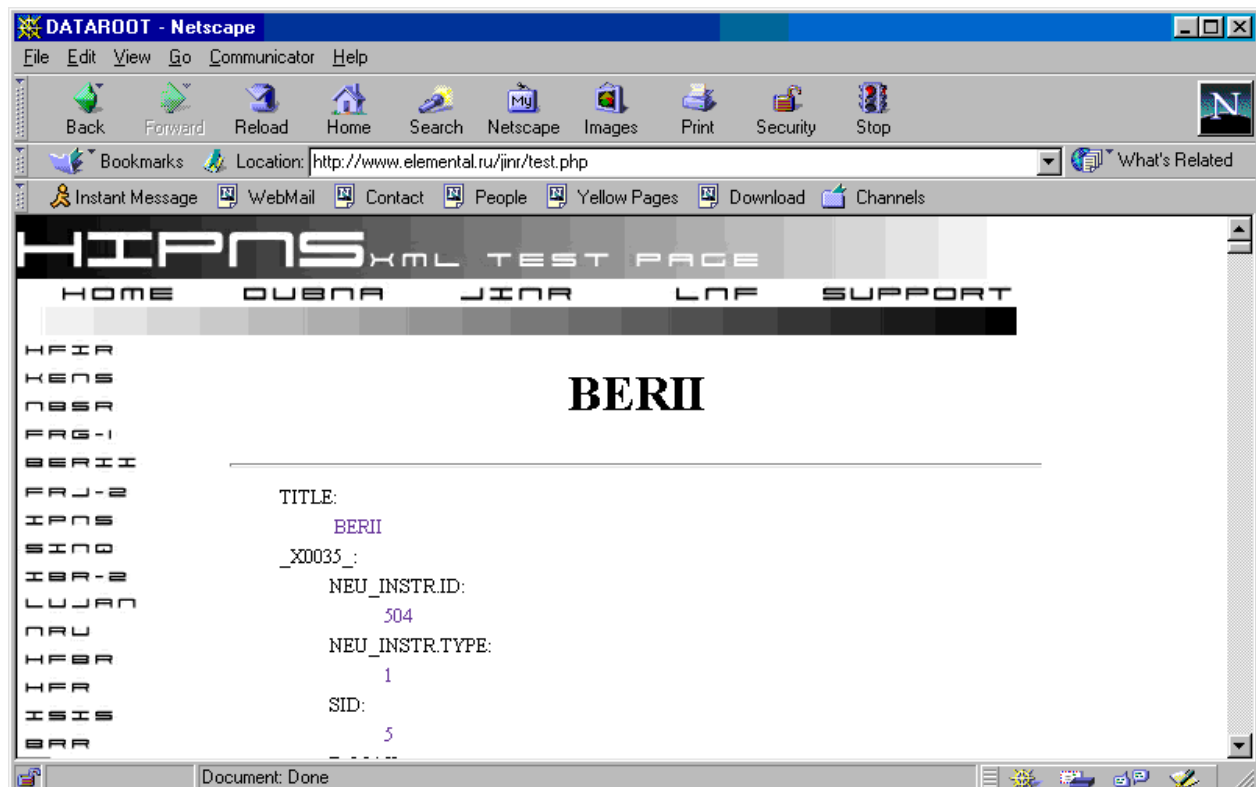
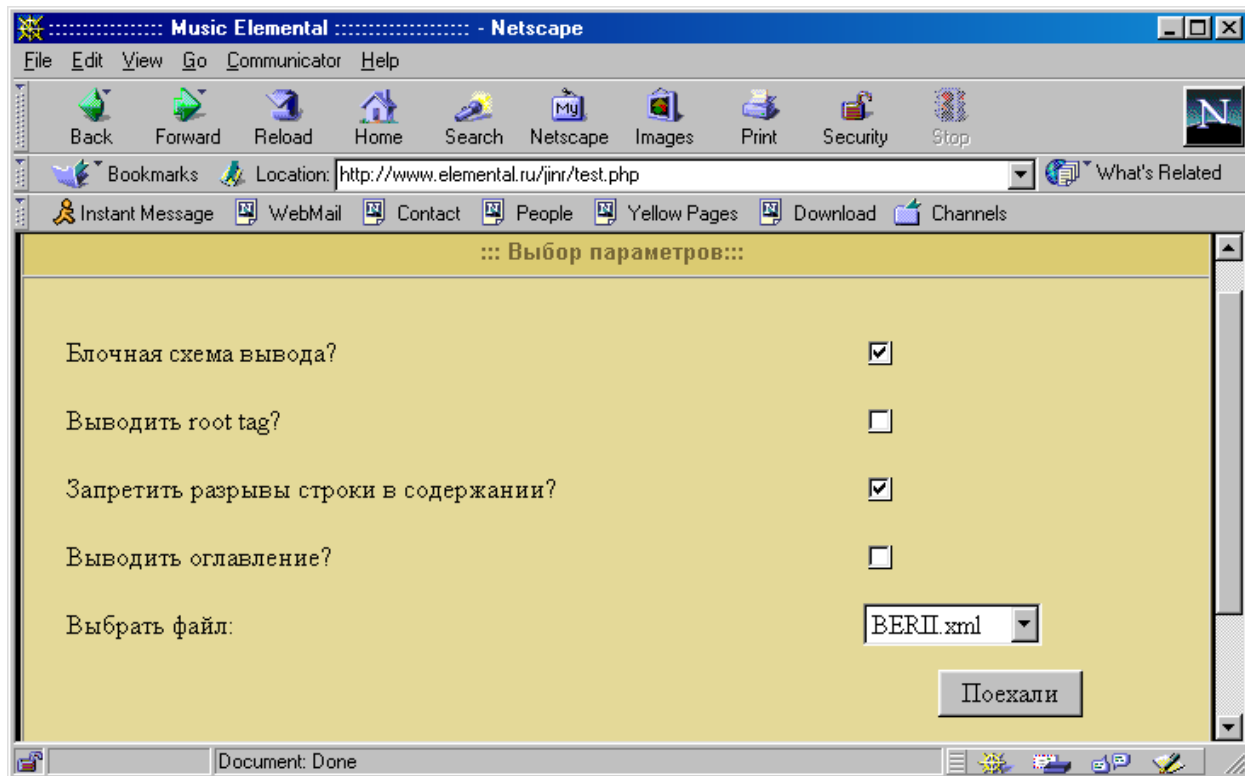


Рисунок 5. Получение и просмотр HTML данных, полученных в результате преобразования XML/XSL кодов с помощью PHP-4-приложения

Литература

1. Litvinenko E.I., Hypertext infosystem for pulsed neutron sources and scientific investigations based on these sources, Proceedings of the Intern.Workshop on Advances in Databases and Information Systems (ADBIS'96), Moscow, MEPhI, 1996, pp. 42-46,
<http://nfdfn.jinr.ru/~litvin/ADBIS96/ab1/ab1.html>
2. E.I. Litvinenko, E.P. Akishina, Y.A. Astakhov, R.N.Semenov, I.S. Smolkov, The hypertext information system on pulsed neutron sources and scientific investigations based on thesesources, Proceedings of the Intern.Workshop NOBUGS'97
http://www.aps.anl.gov/xfd/bcda/nobugs/proceed/litvin_hipns.html
3. E.I. Litvinenko, T.L. Pikelner T.L., R.N. Semenov, The status of HIPNS, The poster presentation on the Intern. Workshop NOBUGS-2000
http://http://nfdfn.jinr.ru/~litvin/nobugs2000/nobugs2000_litvin_hipns_proceeding.htm
4. E.I. Litvinenko, T.L. Pikelner, A.V. Prikhodko, R.N. Semenov, Hypertext information system on neutron sources and neutron instruments, Proceedings of the 3rd International Workshop on Computer Science and Information Technologies CSIT'2001, USATU Publishers, Ufa, Russia, 2001, Vol.2, pp.77-82
5. <http://www.xml.org>
6. <http://xml.apache.org/cocoon/>
7. http://www.ccc.ru/magazine/depot/01_08/read.html?web2.htm
8. http://www.onjava.com/pub/a/onjava/2001/02/22/jsp_servlets.html
9. <http://www.saxproject.org/>
10. <http://jakarta.apache.org/tomcat/>
11. <http://www.martinfowler.com/isa/htmlRenderer.html>