

# *Практическое использование XML*

---

Ростислав ТИТОВ

*Группа e-бизнеса отдела ИТ*

ЦЕРН – Женева, Швейцария

## eXtensible Markup Language

### «Расширяемый язык разметки»

- **SGML (стандарт ISO, 1986)**  
В основном для технической документации
- **XML (стандарт W3C, 1998)**  
Упрощение и развитие SGML, широкая область применения

# Зачем нужна разметка данных?

```
<book lang="Hungarian">
  <chapter>
    Bevezetés
    <section> Szöveg </section>
    <section> Táblázatok </section>
  </chapter>
  <chapter>
    Dokumentumjelölés-jelek
    <section> Barokk stílus </section>
    <section> A csodálatos <del>szöveg</del> </section>
  </chapter>
</book>
```

Разметка позволяет  
добавить информацию о  
структуре документа

# XML: Правила построения

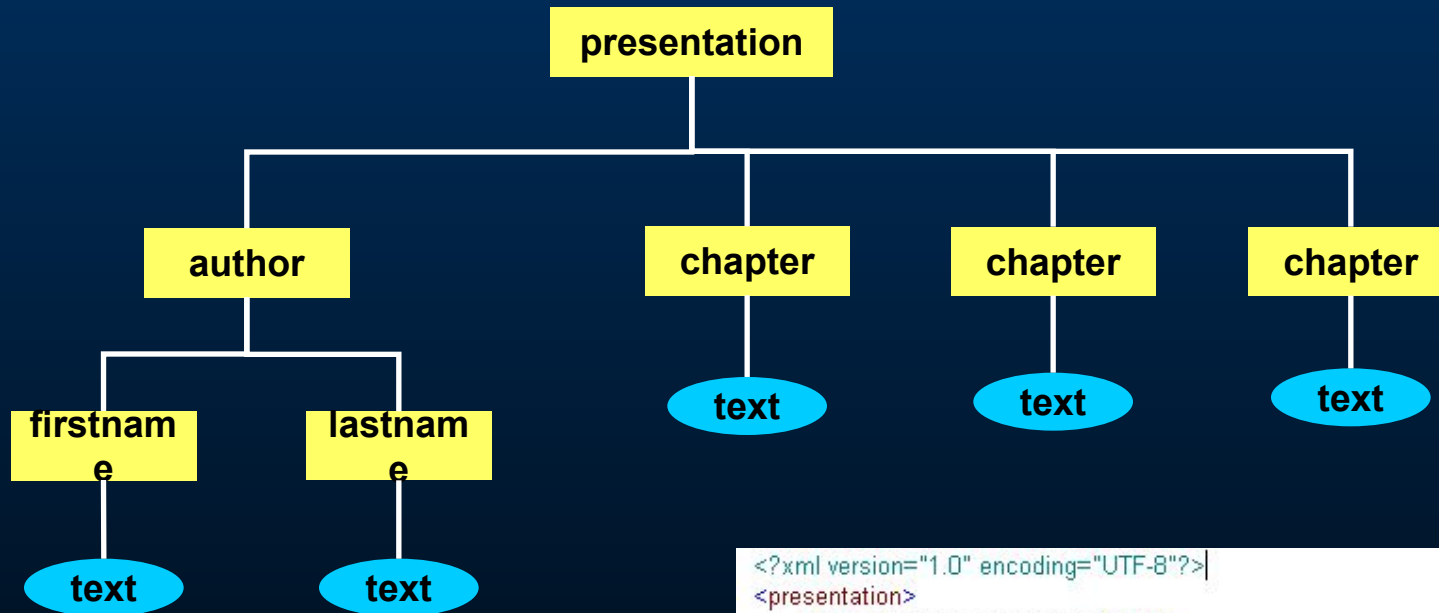
- Заголовок
- Один корневой тэг
- Иерархия тэгов
- Атрибуты
- Текстовые элементы
- Пустые элементы

## Некоторые правила

- Имена элементов чувствительны к регистру букв
- Каждый элемент должен закрываться
- Элементы не могут пересекаться (**<a><b></a></b>**)
- Значения атрибутов - в кавычках или апострофах

```
<?xml version="1.0" encoding="UTF-8"?>
<presentation>
  <author>
    <firstname>Rostislav</firstname>
    <lastname>Titov</lastname>
  </author>
  <chapter number="1" title="What is
XML">
  XML (Extensible Markup Language)
  is ...
  </chapter>
  <conclusion/>
</presentation>
```

# XML: Дерево



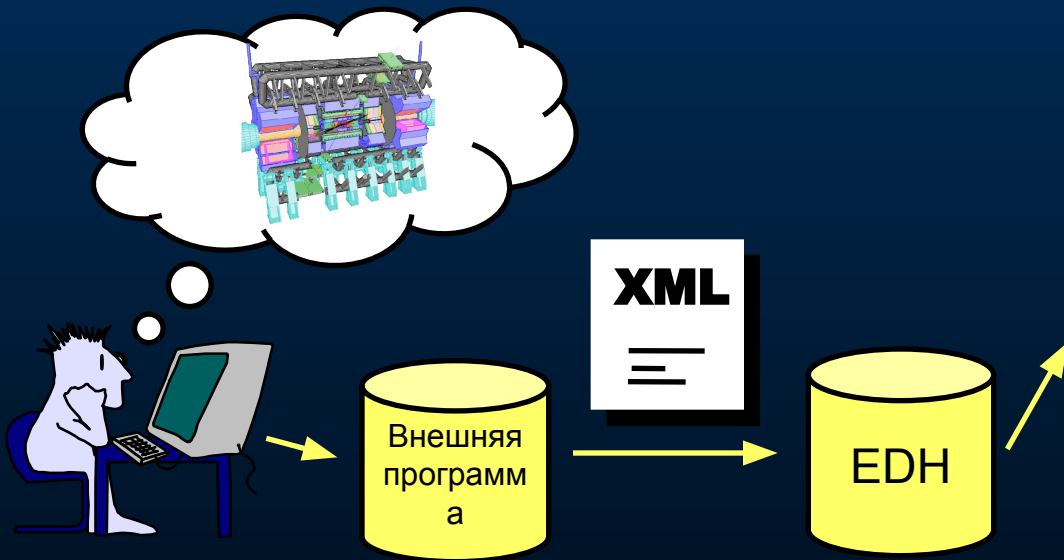
```
<?xml version="1.0" encoding="UTF-8"?>
<presentation>
  <title>Practical Use of XML</title>
  <author>
    <firstname>Rostislav </firstname>
    <lastname>Titov</lastname>
  </author>
  <chapter number="1" title="What is XML">
    XML (Extensible Markup Language) is a standard, proposed by the W3C
    consortium in 1996.
  </chapter>
  <chapter number="2" title="XML Structure">
    XML is a normal text file that could be edited in any text editor, such as
    NotePad.
  </chapter>
</presentation>
```

# XML: Передача данных

- **Независимость от платформы и языка**
- **Простота создания, простота обработки**
- **Понятность для человека и компьютера**
- **Открытый стандарт**
  - Большое количество библиотек обработки
  - Большое количество литературы
  - Специализированные XML-редакторы
- **Возможность проверки структуры**

# XML: Передача данных

## Пример: CERN Electronic Document Handling (EDH)



The screenshot shows a web browser window titled 'Internal Transport / Handling Request - Microsoft Internet Explorer'. The address bar shows 'https://edh.cern.ch/Document/TransportRequest'. The page content includes a title 'Internal Transport / Handling Request' and a note: 'Fields with asterisks (\*) are obligatory and must be filled in.' The form fields are:

- Nature:
- General Description \*:
- Contact \*:
- INB Traceability:

Below these fields is a section titled 'Transport' with the following fields:

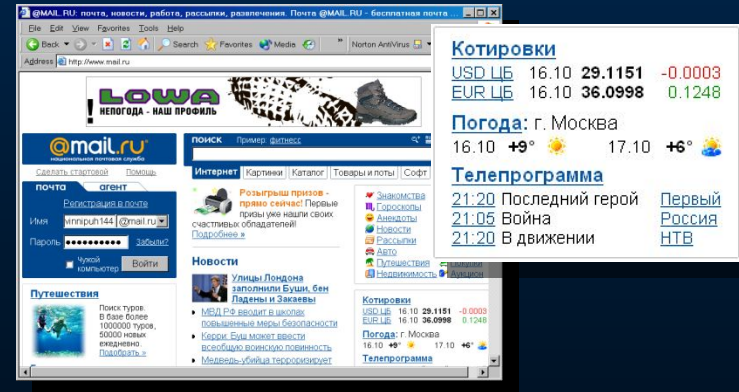
- Recipient:
- From \*:
- Desired Date \*:
- Latest Date:
- To \*:
- Desired Time:

At the bottom, there is a table with columns 'Mass' and 'Quantity'.

- Автоматическая генерация форм из внешних программ
- XML в качестве формата передачи данных
- Анализ XML-схемы - гарантия правильности данных

# Web Services

- Обмен данными между программами через Интернет
- Стандарт
- Независимость от платформы и языка (Java, .Net, ...)



**WSDL – Web Service Definition Language**  
**SOAP – Simple Object Access Protocol**



# XML: Хранение данных

- **Хранение структуры данных вместе с данными**
- **Объектное «дополнение» реляционных СУБД**
- **Проверка структуры**
- **Поддержка на уровне баз данных**
  - **Microsoft SQL Server 2000 +, Oracle 9i +,**
  - **Специальный тип данных для хранения XML**
  - **Специализированные XML-индексы**
  - **Запросы к XML (XQuery и пр.)**
  - **Выдача данных в формате XML**

# XML: Хранение данных

## Пример: Поисковая система EDH

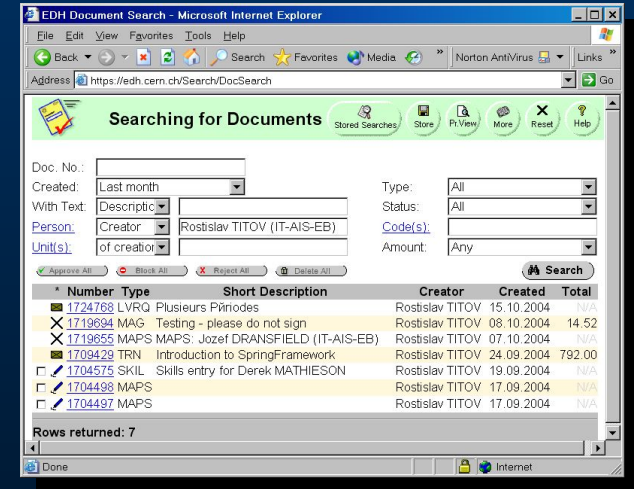
Проблема: Эффективный поиск по произвольному числу критериев – нетривиальная задача

Решение:

- Все документы хранятся в XML
- Контекстный поиск по XML (Oracle InterMedia)

Пример: «Найти документы, которые создал Титов»:

```
Select DOC_ID from DOC_XML where  
Contains(XML, "Titov within creator") > 0;
```



# XML: Преобразование данных

- XML может быть преобразован в HTML, текст, PDF, ...
  - Не требуется специальных программных средств
  - Коммерческие визуальные редакторы
  - Платформонезависимость

# XML: Стандарты на базе XML

- Возможность формального описания структуры
- Независимость от платформы и языка
- Понятность для человека и компьютера
- Возможность использования XML-средств (преобразования XSLT, запросы XQuery)...
  - XHTML (HTML, удовлетворяющий стандарту XML)
  - WSDL (Web Services Definition Language)
  - SOAP (Simple Object Access Protocol)
  - SVG (Scalable Vector Graphics)
  - ebXML (XML for e-Business)
  - ...

# Формализация структуры XML

- **Существуют способы формального определения структуры XML-документов**

*Устарело!  
Не для новых разработок*

- ~~DTD (Document Type Definition)~~
- XML-Схема (XML Schema)



This file is not valid:  
Mandatory element 'firstname' expected in place of 'middlename'

# XML-схема: когда это нужно?

- **Формальное описание структуры для будущего использования**
- **Программисты могут не беспокоиться о правильности входных данных**
- **Создатели XML-документов могут заблаговременно проверить их правильность**

# XML-схема: когда это НЕ нужно?

- Когда заведомо известно, что XML имеет правильную структуру
- Когда правильность структуры не играет роли
- Когда нужна максимальная скорость обработки
- Небольшие «одноразовые» проекты

# XML-схема: ВОЗМОЖНОСТИ

- Набор и порядок следования элементов
- Последовательный порядок элементов (sequence) или выбор (choice)
- Количество повторений элементов и групп элементов
- Набор и наличие/отсутствие атрибутов
- Тип элементов и атрибутов
- Ограничения на значения элементов и атрибутов
- Значения атрибутов по умолчанию
- Уникальность значений
- Поддержка пространств имен (namespaces)



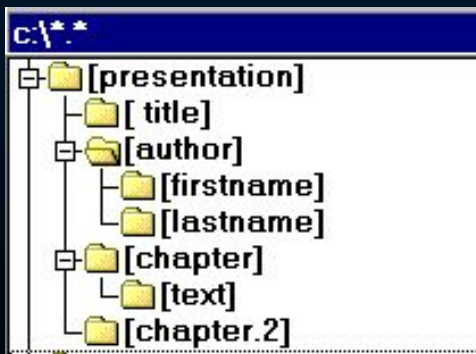
# XML-схема: демонстрация

```
<?xml version="1.0" encoding="UTF-8"?>|
<presentation>
  <title>Practical Use of XML</title>
  <author>
    <firstname>Rostislav </firstname>
    <lastname>Titov</lastname>
  </author>
  <chapter number="1" title="What is XML">
    XML (Extensible Markup Language) is a standard, proposed by the W3C
    consortium in 1996.
  </chapter>
  <chapter number="2" title="XML Structure">
    XML is a normal text file that could be edited in any text editor, such as
    NotePad.
  </chapter>
</presentation>
```

# XPath: Навигация по XML

- Обеспечивает доступ к отдельным элементам XML
- Результатом XPath-выражения может быть:
  - Элемент XML
  - Набор элементов
  - Логическое выражение
  - Строка
  - Число
  - Пустое множество

*C:\presentation\author\firstname*



*/presentation/author/firstname*

```
<?xml version="1.0" encoding="UTF-8"?>
<presentation>
  <title>Practical Use of XML</title>
  <author>
    <firstname>Rostislav </firstname>
    <lastname>Titov</lastname>
  </author>
  <chapter number="1" title="What is XML">
    XML (Extensible Markup Language) is a standard, proposed by the W3C
    consortium in 1996.
  </chapter>
  <chapter number="2" title="XML Structure">
    XML is a normal text file that could be edited in any text editor, such as
    NotePad.
  </chapter>
</presentation>
```

# XPath: Примеры

- Найти имя ректора  
`/institute/rector/person/text()`
- Найти названия факультетов  
`/institute/faculty/@name`
- Найти всех сотрудников  
`//person`
- Найти имя декана факультета «Б»  
`/institute/faculty[@shortname="Б"]/dean/person/text()`
- Найти имя второго по счету заместителя Малюка А. А.  
`//dean/person[starts-with(., "Малюк")]  
/../../deputies/person[position() = 2]`

```
<?xml version="1.0" encoding="UTF-8"?>
<institute name="МИФИ">
  <direction>
    <rector><person>Оныкий Б.Н.</person></rector>
  </direction>
  <faculty name="Факультет автоматики и электроники" shortname="А">
    <dean><person>Рыбин В.М.</person></dean>
    <deputies>
      <person>Шуренков В.В.</person>
      <person>Никитин А.М.</person>
    </deputies>
  </faculty>
  <faculty name="Факультет кибернетики" shortname="К">
    <dean><person>Панферов В.В.</person></dean>
    <deputies>
      <person>Березкин Е.Ф.</person>
    </deputies>
  </faculty>
  <faculty name="Факультет информационной безопасности" shortname="Б">
    <dean><person>Малюк А.А.</person></dean>
    <deputies>
      <person>Кондратьева Т.А.</person>
      <person>Горбатов В.С.</person>
      <person>Толстой А.И.</person>
    </deputies>
  </faculty>
</institute>
```

# XPath: Примеры

## Пример: Система обработки событий



«Хочу уведомления о всех документах на сумму более 600 CHF»

/ document [amount > 600]

# XPath: Использование в программах

## XPath

```
System.out.println(((XMLDocument)xml).selectSingleNode(
"/config/report[@name='Slava']/title/text()").getNodeValue());
```

## DOM Model

```
Element root = xml.getDocumentElement();
Node child;
for (child = root.getFirstChild(); child != null; child = child.getNextSibling())
    if (child.getNodeName().equals("report") && ( (Element)child ).getAttribute("name").equals("Slava"))
        break;
for (child = ((Element)child).getFirstChild(); child != null; child = child.getNextSibling())
{
    if (child.getNodeName().equals("title") )
    {
        for (Node child2 = child.getFirstChild(); child2 != null; child2 = child2.getNextSibling())
            if ( child2 instanceof Text )
                System.out.println(( (Text)child2 ).getData().trim());
    }
}
}
```

```
<config>
  <report name="Vasya">
    <author>X</author>
    <title>Vasya's report</title>
  </report>
  <report name="Slava">
    <author>Y</author>
    <title>Slava's report</title>
  </report>
</config>
```

# Зачем нужен XPath

---

**«XPath является критической составляющей XML-преобразований (XSLT) и запросов XQuery.»**

# XQuery – Язык XML запросов

- **XQuery – это SQL для XML**
  - Независимость от конкретной СУБД
  - Простота использования
- **Поддержка популярными СУБД  
(Microsoft SQL Server 2003, Oracle 9i и 10g)**
- **Базируется на XPath, но более понятен и может работать на множестве документов**

# XSLT: XML Transformations

- Transforms XML to HTML, text or other XML
- **XSLT 1.0 (Current)**, XSLT 2.0 (Draft)
- XSLT is a “Human Interface” to XML
- Supported by Web Browsers

```
<?xml version="1.0" encoding="UTF-8"?>
<presentation>
  <title>Practical Use of XML</title>
  <author>
    <firstname>Rostislav </firstname>
    <lastname>Titov</lastname>
  </author>
  <chapter number="1" title="What is XML">
    XML (Extensible Markup Language) is a standard, proposed by the W3C
    consortium in 1996.
  </chapter>
  <chapter number="2" title="XML Structure">
    XML is a normal text file that could be edited in any text editor, such as
    NotePad.
  </chapter>
</presentation>
```

**XSLT**  
→

## Practical Use of XML

*Author: Rostislav Titov*

### Table of Contents

1. What is XML
2. XML Structure

### Chapter 1. What is XML

*XML (Extensible Markup Language) is a standard, proposed by the W3C consortium in 1996.*

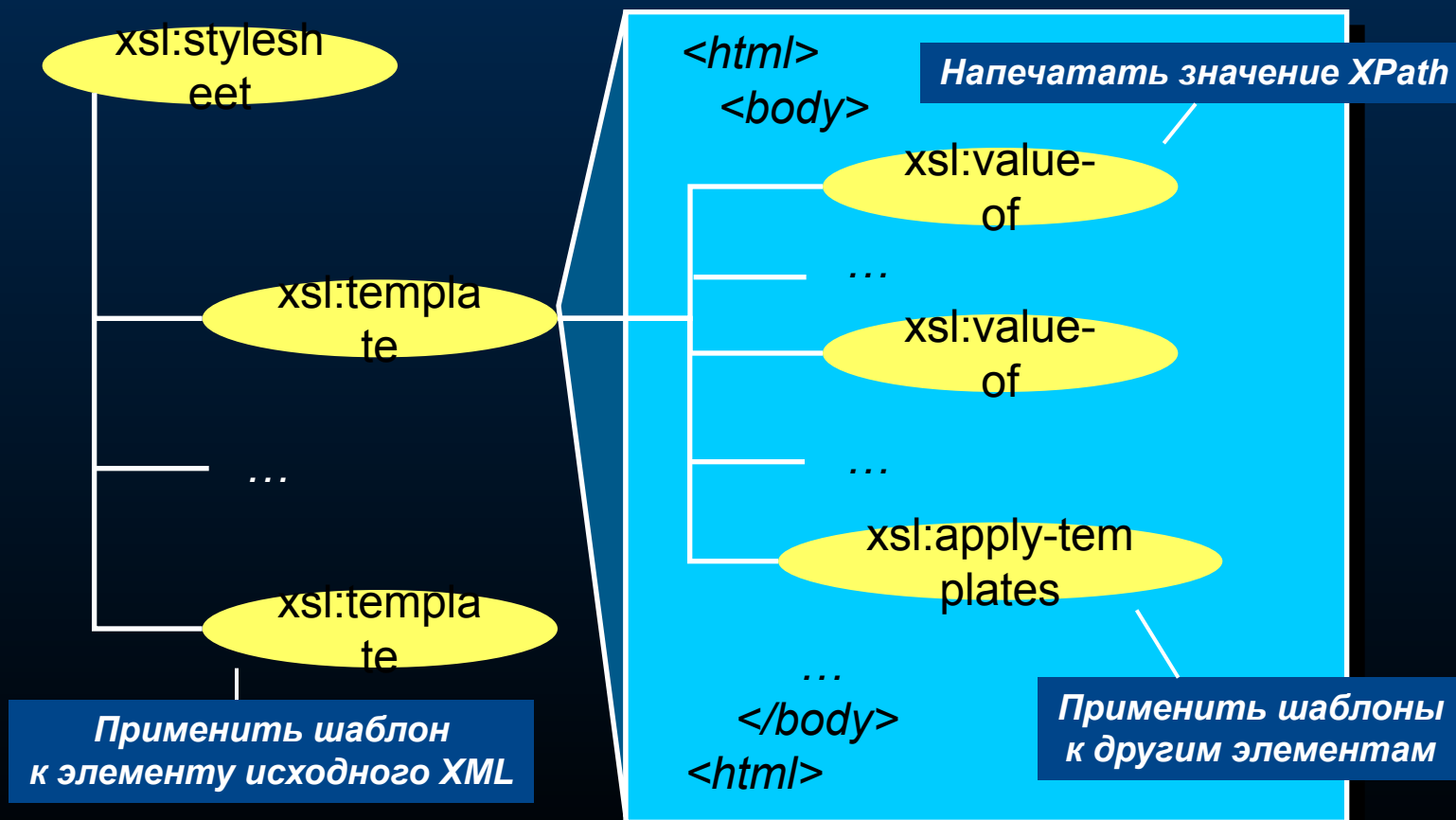
### Chapter 2. XML Structure

*XML is a normal text file that could be edited in any text editor, such as NotePad.*



# XSLT: Упрощенная структура

- XSLT – это файл в формате XML
- Активное использование XPath



# XSLT: Пример

```
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">  
  <xsl:output method="html" version="1.0" encoding="UTF-8" indent="yes"/>
```

```
  <xsl:template match="presentation">
```

```
    <html>
```

```
      <body bgcolor="#FFCCFF">
```

```
        <h1><font color="darkblue"><xsl:value-of select="title"/></font></h1>
```

```
        <h4><font color="green"><i>Author: <xsl:value-of
```

```
select="author"/></i></font></h4>
```

```
        <b>Table of Contents</b><br/><br/>
```

```
        <xsl:apply-templates select="chapter" mode="contents"/>
```

```
        <br/><br/>
```

```
        <xsl:apply-templates select="chapter" mode="normal">
```

```
          </body>
```

```
        </html>
```

```
  </xsl:template>
```

```
  <xsl:template match="chapter" mode="normal">
```

```
    <b>Chapter <xsl:value-of select="@number"/>. <xsl:value-of select="@title"/></b><br/><br/>
```

```
    <i><xsl:value-of select="text()"/></i><br/><br/>
```

```
  </xsl:template>
```

```
  <xsl:template match="chapter" mode="contents">
```

```
    <xsl:value-of select="@number"/>. <xsl:value-of select="@title"/><br/>
```

```
  </xsl:template>
```

```
</xsl:stylesheet>
```

## Practical Use of XML

Author: Rostislav Titov

1. What is XML
2. XML Structure

### Chapter 1. What is XML

*XML (Extensible Markup Language) is a standard, proposed by the W3C consortium in 1996.*

### Chapter 2. XML Structure

*XML is a normal text file that could be edited in any text editor, such as NotePad.*

# XSLT: Другие возможности

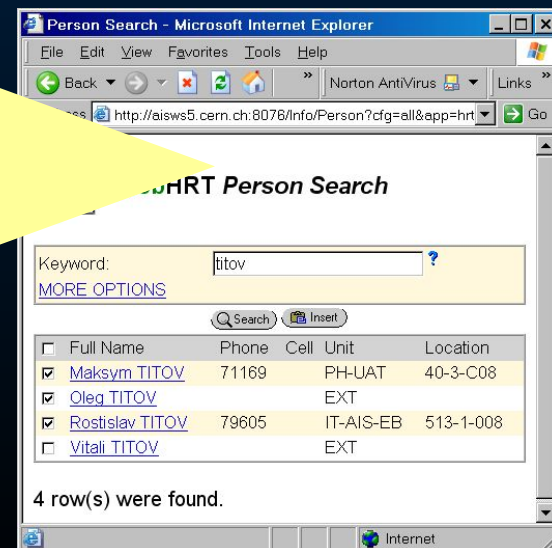
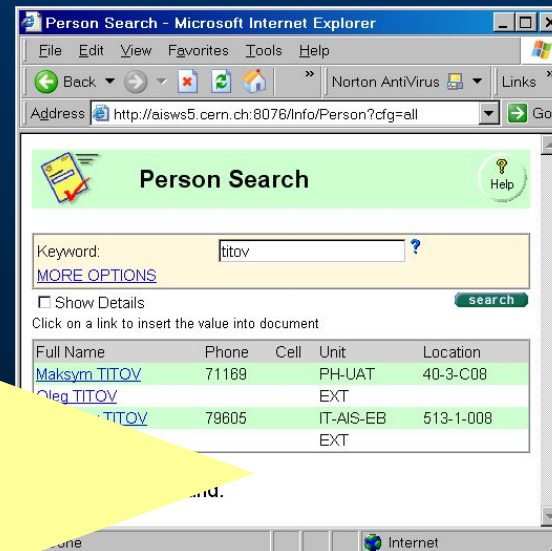
- Условия (<xsl:if>)
- Циклы (<xsl:for-each>)
- Переменные (<xsl:variable>)
- Сортировка (<xsl:sort>)
- Нумерация [1., 1.1., 1.1.a, 2.,] (<xsl:number>)
- Форматирование чисел (format-number())
- Многошаговая обработка (mode)
- Работа со строками (через XPath)

## XSLT 2.0 (Draft)

- XPath 2.0
- Создание собственных функций
- Анализ строк при помощи регулярных выражений
- Форматирование даты и времени

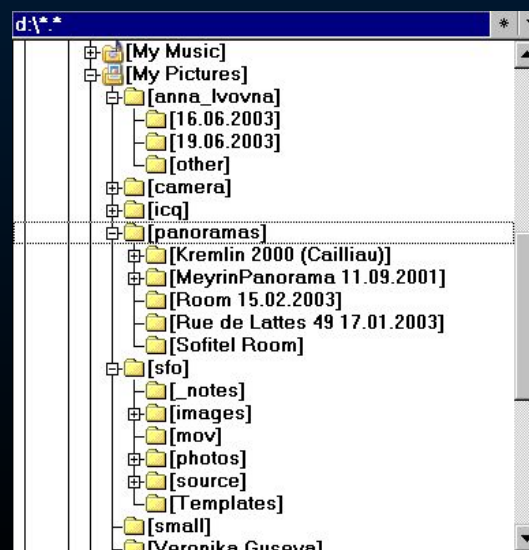
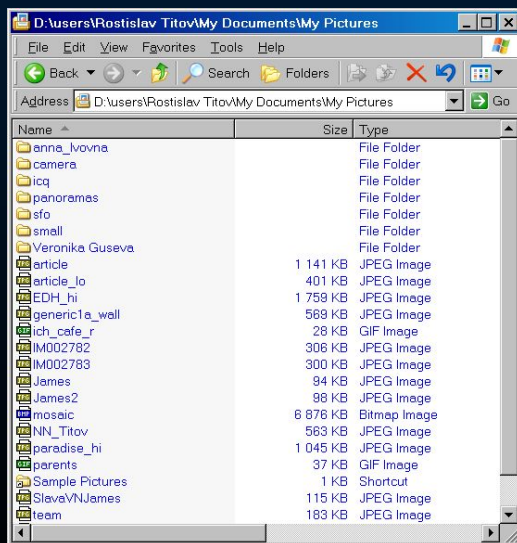
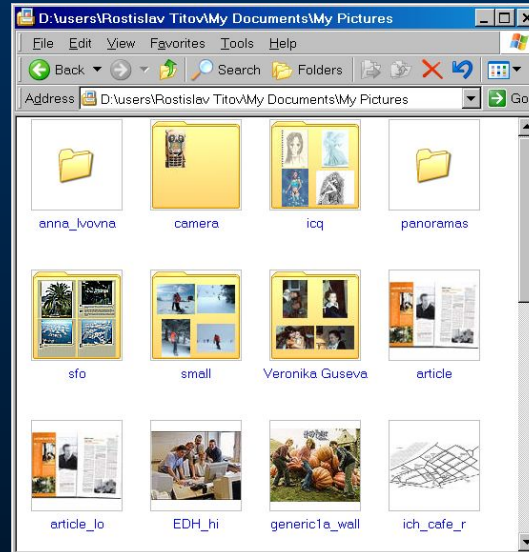
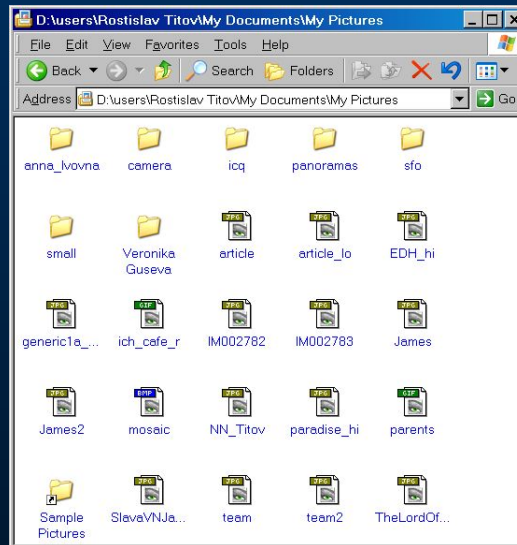
# XSLT: Web “Skins”

```
<aissearchscreen>
  <head><title>Person Search</title></head>
  <body>
    <input type="hidden" name="isAdvanced" value="false"/>
    <input show="always" type="text" label="Keyword"
value="titov"/>
    <input type="checkbox" label="Fuzzy search" value="No"/>
    <result>
      <header>
        <tablecell>Full Name</tablecell>
        ...
      </header>
      <row>
        <tablecell>Maksym TITOV</tablecell>
        <tablecell>71169</tablecell>
        <tablecell>40-3-C08</tablecell>
        ...
      </row>
      <row>
        <tablecell>Oleg TITOV</tablecell>
        <tablecell>EXT</tablecell>
        ...
      </row>
      ...
    <rowcount>4</rowcount>
  </result>
</body>
</aissearchscreen>
```



# XSLT: Web “Skins” - 2

XSLT



# XSLT: Интерфейс пользователя

## CERN Stores Catalog

- Загрузка данных через XML
- Все данные хранятся в XML
- Чистый XML-XSLT
- 15000 наименований
- +10000 пользователей
- Используется ежесекундно
- ~15-20К XML на каждую страницу
- Страницы разного формата (переопределение XSLT)

The screenshot shows the CERN Stores Catalog website in Microsoft Internet Explorer. The search results for 'batteries' are displayed, listing various battery types and their SCEM codes. The top result is '01.24.08 CADMIUM-NICKEL BATTERIES'. Below the list, there is a detailed view of the selected product, including technical information and a table of specifications.

Buy	SCEM Code	Unit	Unit Price	U rated V	DIMENSIONS max. mm	SANYO TYPE
to	01.24.08.100.9	PC	15.50	3.6	Ш15,5 x 58	07009733-N-100SB3

Technical information: [product manager](#)

STANDARD: DIN 49620  
REINFORCED CONSTRUCTION  
VIBRATION-RESISTANT  
Internally frosted glass

WANDER-LAMP: [03.50.10.C](#)  
FLASHING WARNING BOX: [50.64.35](#)

01 BATTERIES AND ACCESSORIES

- 01.24 DRY BATTERIES
  - 01.24.08 CADMIUM-NICKEL BATTERIES
  - 01.24.09 LITHIUM BATTERIES
  - 01.24.20 ALKALI - MANGANESE BATTERIES
  - 01.24.22 NIMH RECHARGEABLE BATTERIES (Nickel-Metal Hydride)
- 01.28 BATTERIES - ACCESSORIES

# XSLT: XML to Text

## Пример:

- Автоматическая генерация кода

### XML-описание

```
<document>
  <input type="person" name="A"/>
  <input type="number" name="B"/>
  ...
</document>
```

General Description *:	Oracle8i	?
Technical Contact *:	Derek MATHIESON (AS-IDS)	?
Supplier:	ORACLE CORPORATION, 20, DAVIS DRIVE, CA.94002 BELMONT (ORAC37, M	?
Country of Distribution *:	US	?
Currency *:	USD Dollar US (1.7)	?
Total Value	\$4.95 (SFr. 7.00)	?

```
String m_GeneralDescription;
Person m_TechnicalContact;
Supplier m_Supplier;
Country m_DistribCountry;
Currency m_Currency;
```

Purchase Order CBO

```
TextInput m_GeneralDescription;
PersonInput m_TechnicalContact;
SupplierInput m_Supplier;
CountryInput m_DistribCountry;
CurrencyInput m_Currency;
```

Purchase Order ServletExecutor

### Программа

Интерфе  
йс

Бизнес-  
логика

SQL

# XSLT: XML to XML

---

- Обновление конфигурационных файлов
- XSL:FO



# XSL-FO: Formatting Objects

- FO: XML-описание макета документа
- XSL-FO: XSLT преобразование документа XML в документ FO
- FOP Processor: программа, преобразующая документ FO в формат для печати (PDF, PS, ...)

Документ XML

```
<?xml version="1.0"?>  
<presentation>  
  <title>  
    XXX  
  </title>  
</presentation>
```

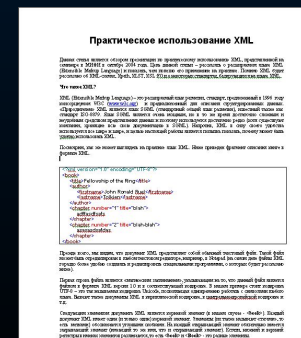
Преобразование XSL:FO

Документ FO

```
<fo:root>  
<fo:page-sequenc  
e>  
  <fo:flow>  
  
  ...  
</fo:flow>  
</fo:page-sequen  
ce>  
</fo:root>
```

FOP Processor

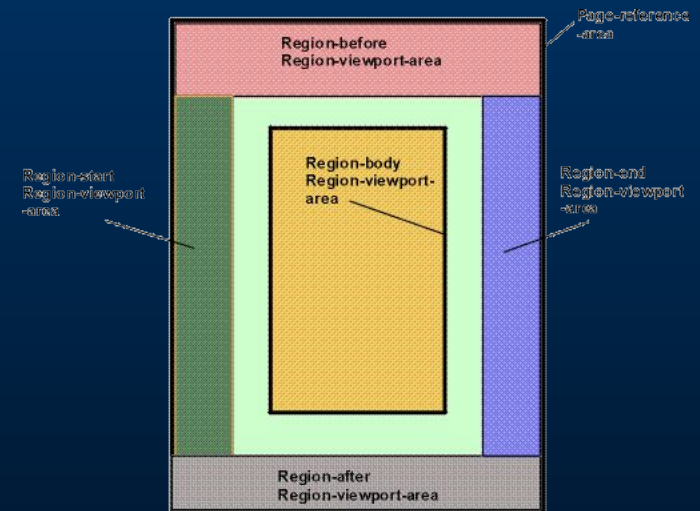
Документ PDF



# XSL-FO: Formatting Objects

*FO обладает всеми возможностями современных текстовых редакторов:*

- Шрифты
- Разбивка на страницы
- Колонтитулы
- Нумерация страниц
- Четные/нечетные страницы
- Отступы и интервалы
- Неразрывные абзацы
- «Висячие» строки
- Таблицы
- Графика
- ...



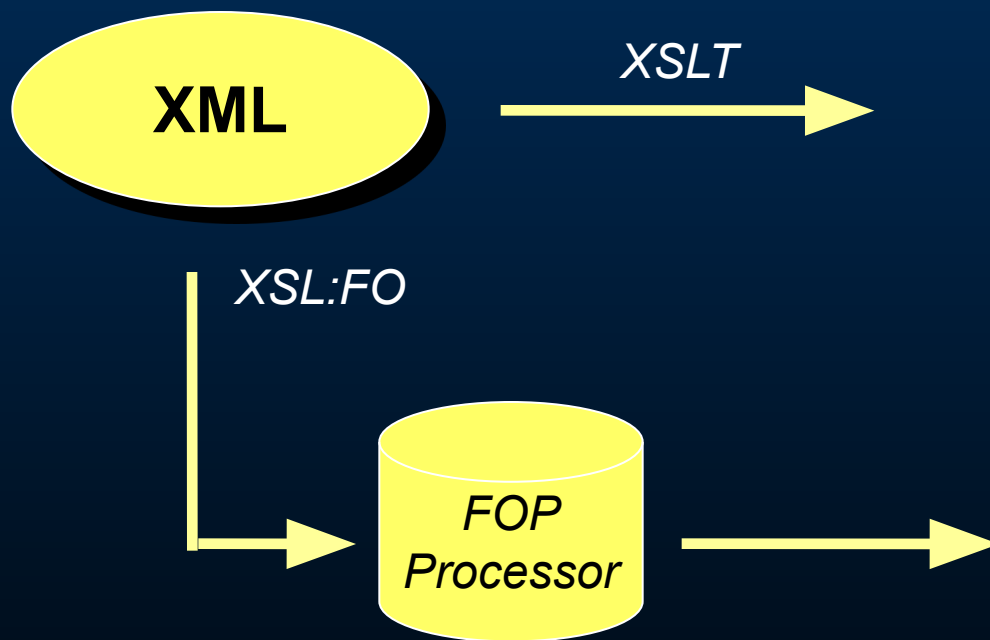
## FOP Processor:

*Apache FOP Processor*



# XSL-FO: Пример

## «Банковская система»



Поддержка PDF не требует написания дополнительного программного кода!

## Web Interface

demo.CQV1 - Asset Allocation

At: Sunday, September 26, 2004 7:12:55 PM

Policy: demo.CQV1 - Asset Allocation (Pos-Instrument + Pos-Currency (Risk) + Pos-Instrument Detail + Sec-Security)

Clients: 1 Assets: 28 Portfolios: 4 Orders: 0 Amount: 1,677,946.21 USD

	Total	USD	EUR	CHF	GBP	CAD
Total	100.00	71.84	32.94	-10.26	4.07	1.40
Bonds	60.00	48.13	7.80		4.07	
Liquities	21.40	21.01	12.96	-13.98		1.40
Equities	12.00	1.21	9.92	1.67		
Miscellaneous	5.72	1.49	2.18	2.05		

Portfolio

Amount	%	Local %
1,677,946.21	100.00	100.00

Print Report... (PDF)

## Client Statement (PDF)

NUMBER : 8066526  
NAME : Advanced Network SA

VALUATION 09/18/04 in USD

MANAGEMENT :

	Valuation	Accrued Interests	Valuation Consolidate	% Total
Funds	2 187 815		2 187 815	39.24%
Stocks	2 182 768		2 182 768	39.15%
Current Acc.	667 495	10 708	678 203	12.54%
Fwd Fx	376 059		376 059	6.75%
Conv. Bond	124 689	854	125 543	2.25%
Str. Bond	24 300	495	24 804	0.44%
total	5 563 133	12 657	5 575 190	100.00%

ASSET CLASS DISTRIBUTION

- Current Amount: 678 203 (12.16%)
- Forward Forex: 376 059 (6.75%)
- Convertible Bond: 125 543 (2.25%)
- Strap Bond: 24 804 (0.44%)
- Stocks: 2 182 768 (39.15%)
- Funds: 2 187 815 (39.24%)

page 1 / 1

# XML: Программная обработка

- **DOM (Document Object Model)**
  - Построение дерева
- **SAX (Simple API for XML)**
  - Обработка событий
  - `startElement()`
  - `endElement()`

**SAX** - быстрее,  
**DOM** -  
универсальнее

## Java, C++:

- Apache Xalan
- Oracle XML Parser
- ...

## PERL, .Net:

- Встроенные библиотеки

# IT-корпорации и XML

- **Чтобы лучше понять значение XML, посмотрим как относятся к нему ведущие IT-корпорации**
  - **Microsoft**
  - **Adobe**
  - **Sun**
  - **Oracle**

# XML и Microsoft

- **Internet Explorer: просмотр XML, поддержка XSLT и XML-схем**
- **Разработчики стандарта XML-схем**
- **Microsoft XML Parser**
- **Поддержка внутри Microsoft Office 2003 (XML, схемы)**
- **Поддержка в .Net**
- **Поддержка в SQL Server 2005: FOR XML (SQL Server 2000), XML Data Type, XQuery-запросы, поддержка схем, индексирование XML, ...**

# XML и Microsoft

- **InfoPath 2003**

- Корпоративная система обработки электронных форм
- Полностью основана на XML
- Описание бизнес-правил в виде XML-схемы
- Проверка правильности ввода данных при помощи XML-схемы

# XML и Adobe

- **Adobe Intelligent Document Platform**





# XML и Oracle

- Oracle XML Parser
- Основной формат описания данных в JDeveloper, Oracle IAS, ...
- Oracle 9i: XML Data Type, XQuery-запросы, поддержка схем, индексирование XML, ...
- Oracle 10g: еще больше XML

# XML и Sun

---

- XML API – стандартная библиотека Java 2
- Веб-приложения - описание при помощи XML
- Сотрудничество с W3C и Apache XML Group

# Заключение

---

**«XML является одним из важнейших достижений ИТ-технологий последних лет. Сегодня в мире насчитывается огромное количество XML-приложений, и с каждым годом это количество будет расти»**

**Вывод: XML нужно знать и уметь его применять!**