

# **XML**

Разбор  
Приложения

# XML - спецификации

- Extensible Markup Language (XML) 1.0. W3C Recommendation 10-Feb-1998
- ...
- Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26-Nov-2008
- Extensible Markup Language (XML) 1.1. W3C Recommendation 05-Nov-2003
- Extensible Markup Language (XML) 1.1 (Second Edition). W3C Recommendation 16-Aug-2006 / 29-Sep-2006.

# Общая структура XML-документа

<?xml version="..." encoding="..."?>

<?имя строка?> (“processing instructions”)

<!DOCTYPE имя-элемента PUBLIC  
“идентификатор” “URI” [DTD]>

<имя-элемента>

...

</имя-элемента>

# Разбор

- Два уровня корректности XML-документов:
  - Well-formed – отсутствуют ошибки микросинтаксиса, документ может быть успешно разобран
  - Valid – документ соответствует заданному DTD.
- Два типа программ разбора:
  - Валидирующие
  - Не валидирующие

# Разбор

## Классы алгоритмов разбора

- Событийные
  - Алгоритм генерирует вызовы функций-callback на каждый значимый элемент синтаксиса
- DOM
  - Алгоритм оставляет в ОЗУ структуру данных (дерево), отражающее структуру документа, и доступную через интерфейсы DOM

!Сравнение подходов

# Событийный разбор

Синтаксические элементы, на которые генерируются вызовы функций

- Открывающий тэг
- Закрывающий тэг
- Блок текста без разметки
- Processing instruction
- Entity
- Некоторые другие

# Событийный разбор

## Программное обеспечение

- expat.c (C, free, доступна в исходном коде)

## Интерфейс SAX (Simple API for XML)

- Объектно-ориентированный интерфейс для событийного разбора
- Первоначально ПО разбора с интерфейсом SAX разработано на языке Java (и для программ на Java)
- Позднее ПО разбора с интерфейсом SAX появилось и для других языков
  - C++
  - Perl
  - Python
  - PHP (PHP5)
  - и проч.

# DOM - разбор

- DOM API стандартизовано W3C
- DOM API позволяет:
  - Анализировать документ
  - Модифицировать документ
  - Сериализировать документ (преобразовать обратно в XML)
- DOM – разбор реализован во многих библиотеках для XML для почти всех языков



# ПО разбора

- expat.c («библиотека»)
- MSXML (SAX, DOM, валид./невалид.; COM)
- Apache Xerces (SAX, DOM, Java, C++, Perl)
- Libxml2 (SAX, DOM-like, C, C++, Python, Pascal, Ruby, tcl, PHP5, Java)
  - + gdom2: реализация DOM

# Приложения XML

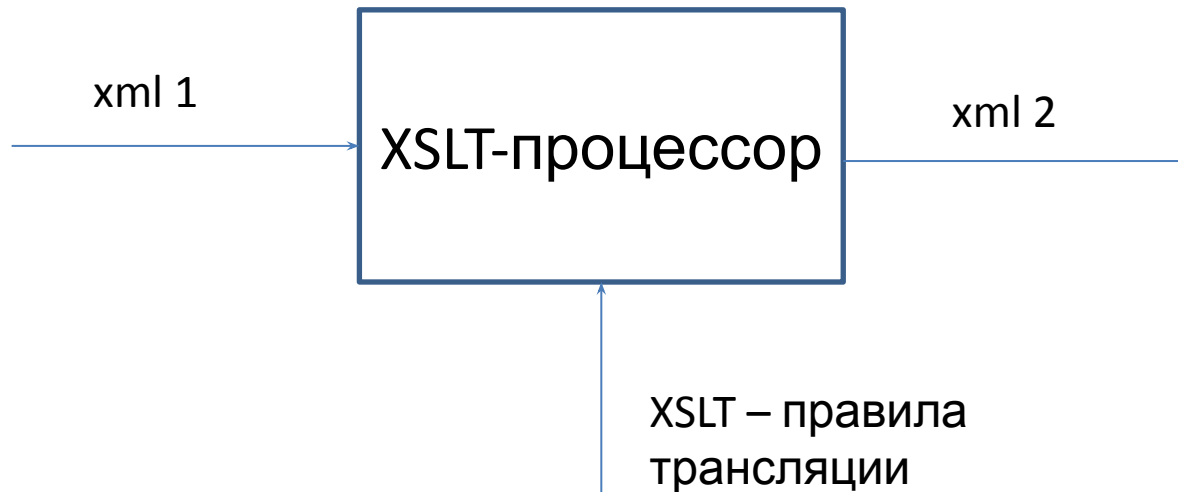
- XHTML
- XML Schema
- SVG
- MathML
- XForms
- XQuery
- SOAP
- XLink/XPointer/XPath
- XSL (“XML Stylesheet language”)

# XSL

XSL - XML Stylesheet language

- В отличие от элементов HTML/ХHTML, у элементов (других приложений) XML нет способа отображения по умолчанию.
- Тем более, у элементов языка разметки, только что созданного под некоторые нужды, такого отображения нет.

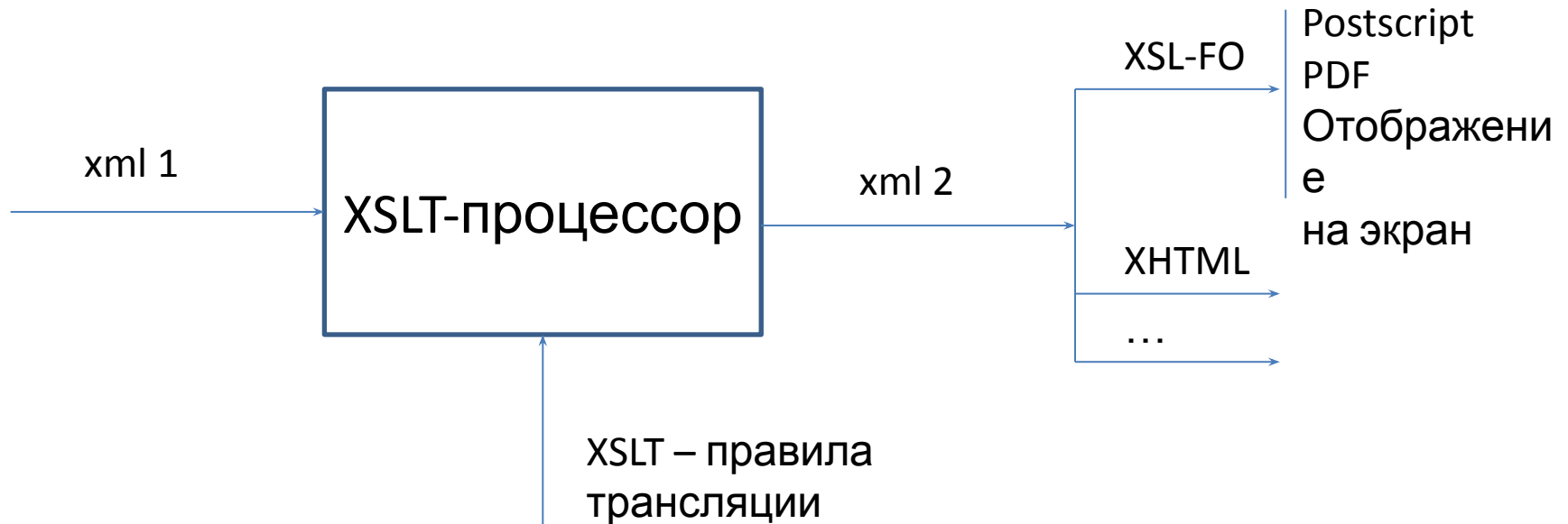
# XSL



## XSLT – XSL Transformation

язык описания преобразования (перевода)  
документа из одного приложения XML в другое

# XSL



## XSL-FO (XSL Formatting objects)

Язык описания представления данных,  
выводимых на печать.

# Приложения XML

- RDF
- RDFS
- OWL

(«Semantic Web»)