

## Разработка SAX-парсера учебных планов для формирования семестровых рабочих учебных планов

А. Н. Вильданов

*Башкирский государственный университет, Нефтекамский филиал*

*Россия, Республика Башкортостан, 452681 г. Нефтекамск, улица Трактовая, 1.*

*Email: alvild@mail.ru*

Статья посвящена разработке веб-приложения, позволяющего автоматически сгенерировать семестровые учебные планы по файлу учебного плана, подготовленного в программе Gosinsp или Планы Лаборатории ММИС. Приложение позволит избежать ошибок при ручном рутинном составлении документации и ускорить его. Разработанное приложение не требует установки, для его работы достаточно современного браузера. Изложенная методология и алгоритмы будут полезны разработчикам при создании приложений, работающих с XML-документами.

**Ключевые слова:** JavaScript, SAX-парсер, парсинг, рабочий учебный план, xml, автоматизированное формирование учебно-методической документации.

Планирование и организация учебного процесса в высших учебных заведениях осуществляются прежде всего на основе действующего рабочего учебного плана. В настоящее время большинство вузов используют для проектирования и проверки учебных планов программные продукты Лаборатории ММИС – программу «Макет УП ВПО (универсальный)» или бесплатную версию программы с ограниченными возможностями – Gosinsp. Первая из них располагает достаточно широким функционалом:

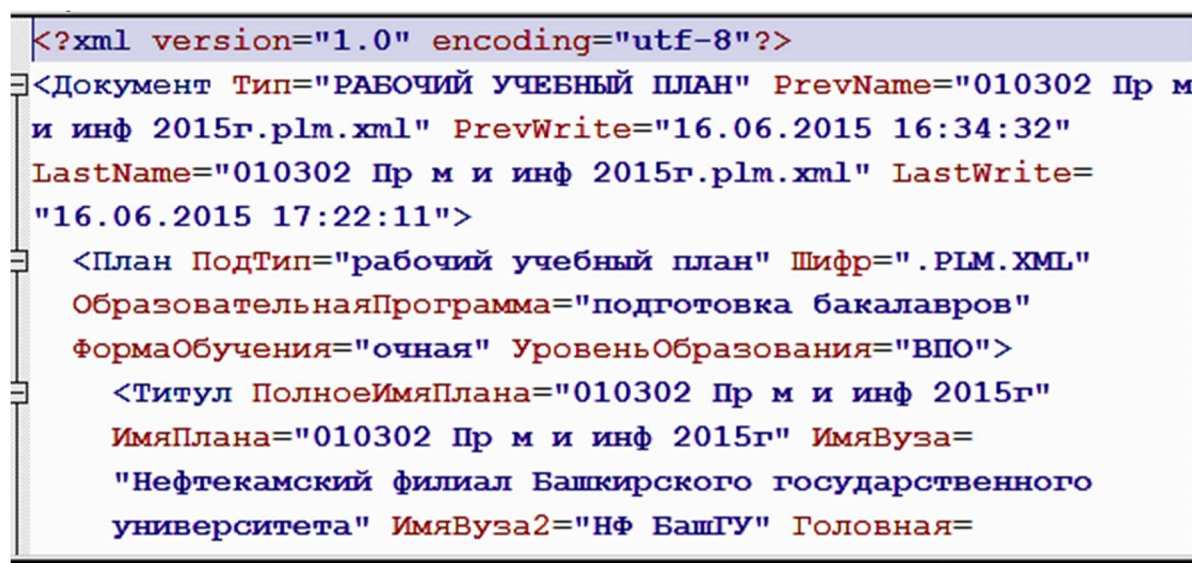
- анализ учебного плана на предмет соответствия требованиям ФГОСЗ и ФГОСЗ+;
- возможность генерации плана для заочной формы обучения на основе очной;
- возможность распределения компетенций по дисциплинам, практикам и ГИА;
- возможность просмотра планируемых дисциплин занятий по отдельным курсам, и т.д.

Несмотря на столь внушительные возможности, их не всегда хватает: иногда в силу каких-либо особенностей ведения номенклатуры конкретного вуза, иногда требуемый функционал попросту не оплачен и отсутствует. В таком случае можно попробовать самому изучить структуру файлов учебного плана и разработать собственные интерфейсы для получения разнообразных отчетов в удобной и приемлемой для своего вуза форме. Например, в работе [1] разработан процесс автоматизированного формирования рабочих программ на основе XML-файлов учебных планов.

Учебный план Лаборатории ММИС представляет собой файл с расширением .xml – т.е. документ в формате XML. XML-документ представляет собой обычный текстовый

файл, в котором при помощи HTML-образных тегов описываются элементы данных, последовательность и вложенность которых определяет структуру документа и его содержание. В настоящее время формат XML находит широкое применение и является, например, де-факто стандартом для описания ресурсов приложения.

Открыть для просмотра или редактирования файл с расширением .xml можно с помощью обычного текстового редактора. Если открыть учебный план в Notepad++, то можно увидеть его структуру (рис. 1):



```
<?xml version="1.0" encoding="utf-8"?>
<Документ Тип="РАБОЧИЙ УЧЕБНЫЙ ПЛАН" PrevName="010302 Пр м
и инф 2015г.plm.xml" PrevWrite="16.06.2015 16:34:32"
LastName="010302 Пр м и инф 2015г.plm.xml" LastWrite=
"16.06.2015 17:22:11">
  <План ПодТип="рабочий учебный план" Шифр=".PLM.XML"
ОбразовательнаяПрограмма="подготовка бакалавров"
ФормаОбучения="очная" УровеньОбразования="ВПО">
  <Титул ПолноеИмяПлана="010302 Пр м и инф 2015г"
ИмяПлана="010302 Пр м и инф 2015г" ИмяВуза=
"Нефтекамский филиал Башкирского государственного
университета" ИмяВуза2="НФ БашГУ" Головная=
```

Рис. 1. Пример содержания XML-файла учебных планов.

Данная статья посвящена решению следующей задачи – разработать веб-интерфейс, который позволял бы получить семестровый учебный план для конкретного курса. Семестровый учебный план (СУП) – это документ, который описывает какие объемы учебной нагрузки необходимо прочитать на выбранном учебном курсе в двух семестрах. Семестровый учебный план по сути просто аккумулирует определенную информацию из учебного плана направления.

Чтобы получить информацию из учебного плана, нужно его спарсить. Сегодня парсинг текстовых документов является эффективным решением для автоматизации процесса извлечения из них информации. Под парсингом документа обычно понимают получение его текстового содержимого с последующим разбором с целью извлечения данных [2]. Разбор текста, в свою очередь, может быть осуществлен с помощью регулярных выражений или программных библиотек, предназначенных для восстановления его древовидной структуры. Но регулярные выражения для сложных текстов теряют наглядность, их становится трудно воспринимать и отлаживать. Построение струк-

турного дерева текста требует больших затрат памяти и является излишним для многих практических задач [3].

Поскольку мы не собираемся менять исходный документ, а хотим получить из него «выжимку», нам отлично подойдет SAX-парсер. SAX-парсер не строит дерево данных, тем самым не расходует много памяти, а последовательно пробегает по XML-файлу, генерируя время от времени события, в которых сообщает вашему приложению о встреченных элементах XML-разметки.

SAX-парсер уже давно стал родным для языка Java, и jdk-библиотеки содержат соответствующий класс SAXParser. Реализации этого парсера для языка JavaScript разрабатываются различными энтузиастами. Мы будем использовать библиотеку jsXmlSaxParser [4]. Она обеспечивает SAX-подобные API для работы с данными. SAX-парсер содержит следующие основные методы [5]:

- startDocument уведомляет о начале документа;
- endDocument уведомляет о конце документа;
- StartElement сообщает об открывающем теге и содержит информацию о его атрибутах и их значениях;
- endElement сообщает об закрывающем теге;
- characters содержит информацию о тексте между текущими тегами и его длине.

Для получения семестрового учебного плана нам необходимо посмотреть, как выглядит структура описания изучаемых дисциплин внутри файла (рис. 2):

```
<Строка Дис="Алгебра и геометрия" НовЦикл="Б1.Б"
НовИдДисциплины="Б1.Б.9" Цикл="Б1" ИдентификаторВидаПлана
="2" ИдентификаторДисциплины="Б1.Б.9" ГОС="216" СР="43"
КомпетенцииКоды="7&amp;10&amp;11&14" Компетенции=
"ОК-7, ОПК-1, ОПК-2, ПК-1" Кафедра="3" ПодлежитИзучению=
"216" КредитовНаДисциплину="6" ЧасовВЗЕТ="36" СемЭкз=
"12">
  <Сем Ном="1" Лек="18" Пр="26" КСР="4" СРС="24" ЧасЭкз=
  "36" ЗЕТ="3" Экз="1" КонтрРаб="1" Контр="1">
  </Сем>
  <Сем Ном="2" Лек="18" Пр="22" КСР="4" СРС="19" ЧасЭкз=
  "45" ЗЕТ="3" Экз="1" КонтрРаб="1" Контр="1">
  </Сем>
</Строка>
```

Рис. 2. Пример описания изучаемых дисциплин в XML-файле учебного плана.

Из рисунка видно, что вся подробная информация об изучаемой дисциплине содержится внутри тегов <Строка> и </Строка>. Например, атрибут «Дис» отражает название дисциплины, атрибут «Компетенции» содержит перечень кодов компетенций, на усвоение которых направлена данная дисциплина, и т.д. Внутри тегов <Строка> находятся один и более тегов <Сем>, которые показывают, в каком семестре (семестрах) происходит изучение данной дисциплины, и количество отведенных часов на различные формы проведения занятий и промежуточные аттестации.

Изучив структура плана, можно предложить следующий алгоритм для извлечения СУП:

- последовательно продвигаемся по всему документу;
- при достижении открывающего тега *Строка* извлекаем название дисциплины с помощью атрибута *Дис*;
- при достижении открывающего тега *Сем* извлекаем номер семестра, количество лекционных, практических, лабораторных занятий, количество часов, отводимых на самостоятельную работу и на ее контроль, вид промежуточной аттестации, наличие контрольной работы и т.д.;
- при достижении закрывающего тега *Строка* готовим место для следующей дисциплины.

Результатом разработки является приложение, которое позволяет, загрузив макет учебного плана, вывести для данного курса список всех дисциплин с часами (рис 3):

Выберите файл

oop.plm.xml

третий курс ▼

☐ Показывать код дисциплин

start

Название дисциплины	Семестр/Сессия	Лекц.	Практ.	Лаб.	Зачет	Зачет с оценкой	Экз.	Курс. раб. / проект	Контр. р.	РГР
Социология	5	14	22		1					
Функциональный анализ	5	16	24				1			
Теория вероятностей и математическая статистика	5	24	40				1			
Теория вероятностей и математическая статистика	6	24	40				1		1	
Базы данных	5	28	44				1			

Рис. 3. Вывод семестрового учебного плана.

Приложение позволяет вывести список дисциплин как для отдельного курса, так и сразу для всех годов обучения. Оно позволит избежать ошибок при ручном, довольно рутинном и утомительном способе составления документации, и значительно ускорить этот процесс. Приложение не требует установки, ведь оно разработано на языке JavaScript, поэтому для его работы достаточно современного браузера с выходом в сеть Интернет.

Предложенная технология и алгоритмы будут полезны разработчикам при создании подобных приложений, работающих как с учебными планами, так и, вообще, с произвольными XML-документами. Ознакомиться с работой приложения можно по адресу <https://gevaraweb.github.io/uchplan/uchplan.html>.

### Литература

1. Широбокова С. Н., Щербакова Е. А., Кацупеев А. А., Евсин В. А. Математическая модель и программная реализация инструментария для автоматизированного формирования учебно-методической документации // Известия высших учебных заведений. Северо-Кавказский регион. Серия: Технические науки. 2016. №3 (191). С. 18–23.
2. Вильданов А. Н. Парсинг HTML-документа на языке Java (на примере расписания НФ БашГУ) // Достижения и приложения современной информатики, математики и физики: материалы V Всероссийской научно-практической заочной конференции. – г.Нефтекамск, 2016. – С. 17–25. – ISBN 978–5–7477–4280–2.
3. Орещенков Игорь. Извлечение информации из HTML-страниц средствами PHP / Игорь Орещенков // Системный администратор. №1–2 (182–183). Издательский дом «Положевец и партнеры», Москва, 2018. – С. 60–64.
4. Nicolas Debeissat. SAX parser in javascript [Электронный ресурс]. URL: <http://debeissat.nicolas.free.fr/sax.php> (дата обращения: 01.11.2017).
5. Yong Mook Kim. How to read XML file in Java – (SAX Parser) [Электронный ресурс]. URL: <https://www.mkyong.com/java/how-to-read-xml-file-in-java-sax-parser> (дата обращения: 01.12.2017).

Статья рекомендована к печати экономико-математическим факультетом Нефтекамского филиала Башкирского Государственного университета (канд. физ.-мат. наук, доц. Г. Ф. Сафина)

## Development of SAX parser working curriculum for the formation of semester workplans

A. N. Vildanov

*Bashkir State University, Neftekamsk Branch*

*1 Traktovaya Street, 452681 Neftekamsk, Republic of Bashkortostan, Russia.*

*Email: alvild@mail.ru*

The article is devoted to the development of a web application that allows automatically generating semester curricula on the curriculum file prepared in the Gosinsp program or the MMIS Laboratory Plans. The application will avoid mistakes during manual routine compilation of documentation and accelerate it. The application does not require installation, for its operation is quite modern browser. The stated methodology and algorithms will be useful for developers when creating applications that work with XML documents.

**Keywords:** JavaScript, SAX parser, parsing, working curriculum, xml, automated formation of educational-methodical documentation.