

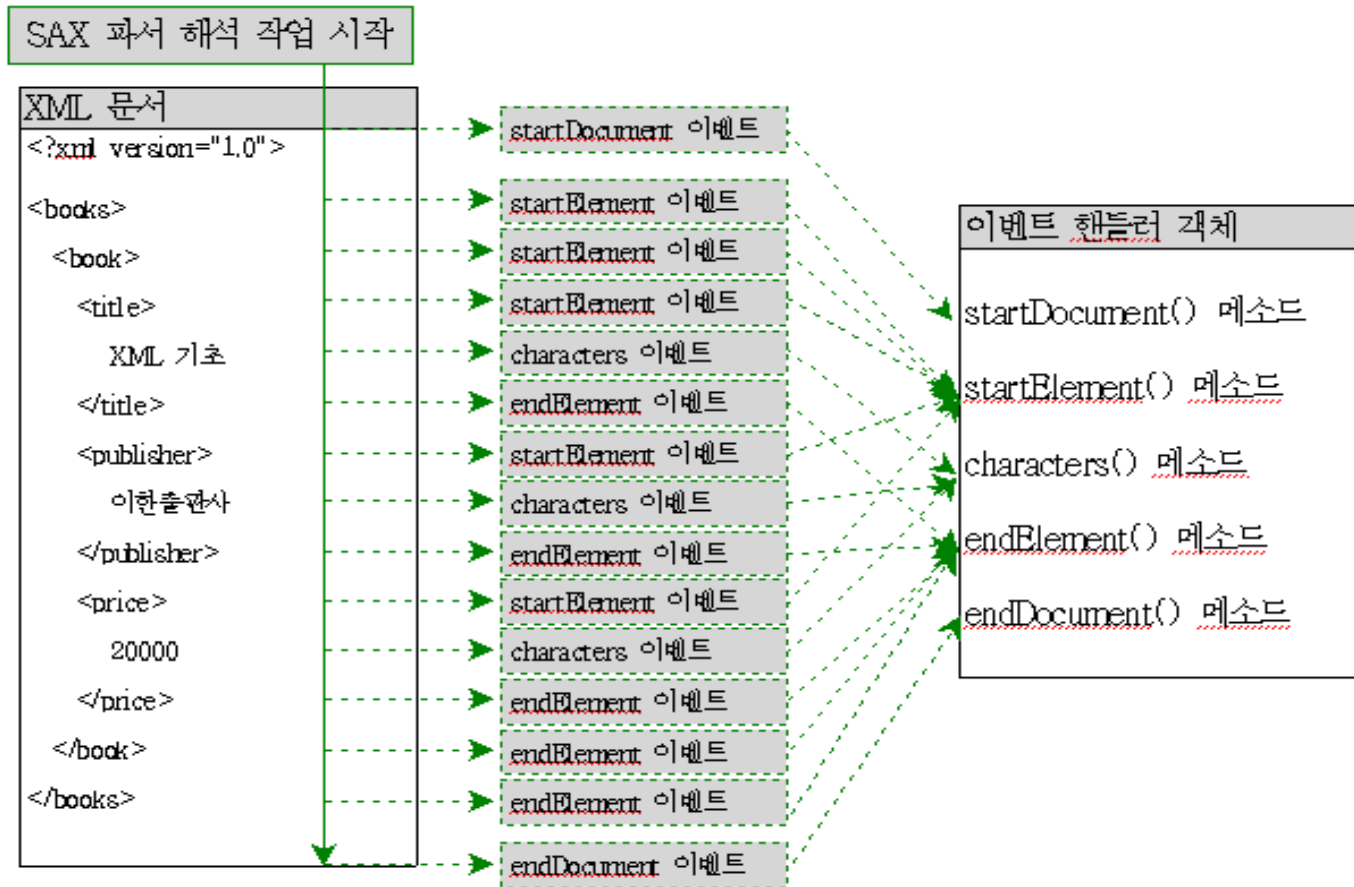
제 10 장 SAX 기반 문서 처리 프로그래밍

1. SAX 역사 및 개념
2. SAX 파서의 작동 메커니즘
3. SAX 스펙 관련 Java 패키지
4. SAX 파서 선택
5. SAX 파서의 생성
6. ContentHandler의 이벤트 처리
7. DTDHandler의 이벤트 처리
8. EntityResolver의 이벤트 처리
9. ErrorHandler의 이벤트 처리
10. DefalutHandler의 이벤트 처리
11. 다양한 형태의 XML 문서의 파싱

1. SAX 역사 및 개념

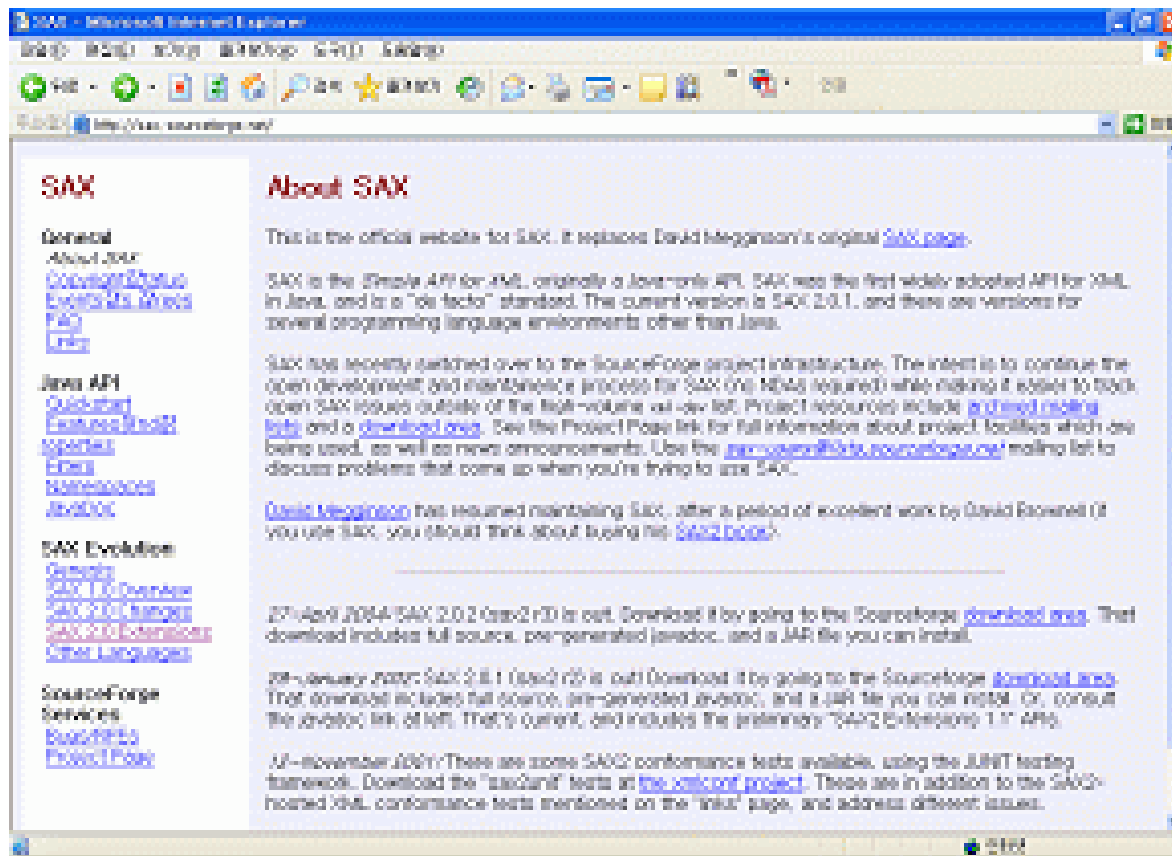
- SAX(Simple API for XML) 역사
 - XML-DEV라는 메일링 리스트에 속한 소프트웨어 기술자들의 의견 교환에 의해서 탄생
 - 초창기 XML 개발자인 Tim Bray, David Megginson, Jame Clark 가 참여
 - 1998년 5월 SAX 1.0 발표
 - 네임스페이스 지원하지 않음
 - 2000년 5월 SAX 2.0 발표
- SAX 개념
 - XML 문서 전체를 처음부터 끝까지 차례대로 읽어 가면서 **이벤트를 발생**시켜 처리하는 구조
 - Ex) XML 문서 해석 도중 시작 태그를 만나면 startElement 이벤트 발생, 문자 데이터를 만나면 characters 이벤트 발생 등 ...

1. SAX 역사 및 개념



1. SAX 역사 및 개념

– <http://sax.sourceforge.net>

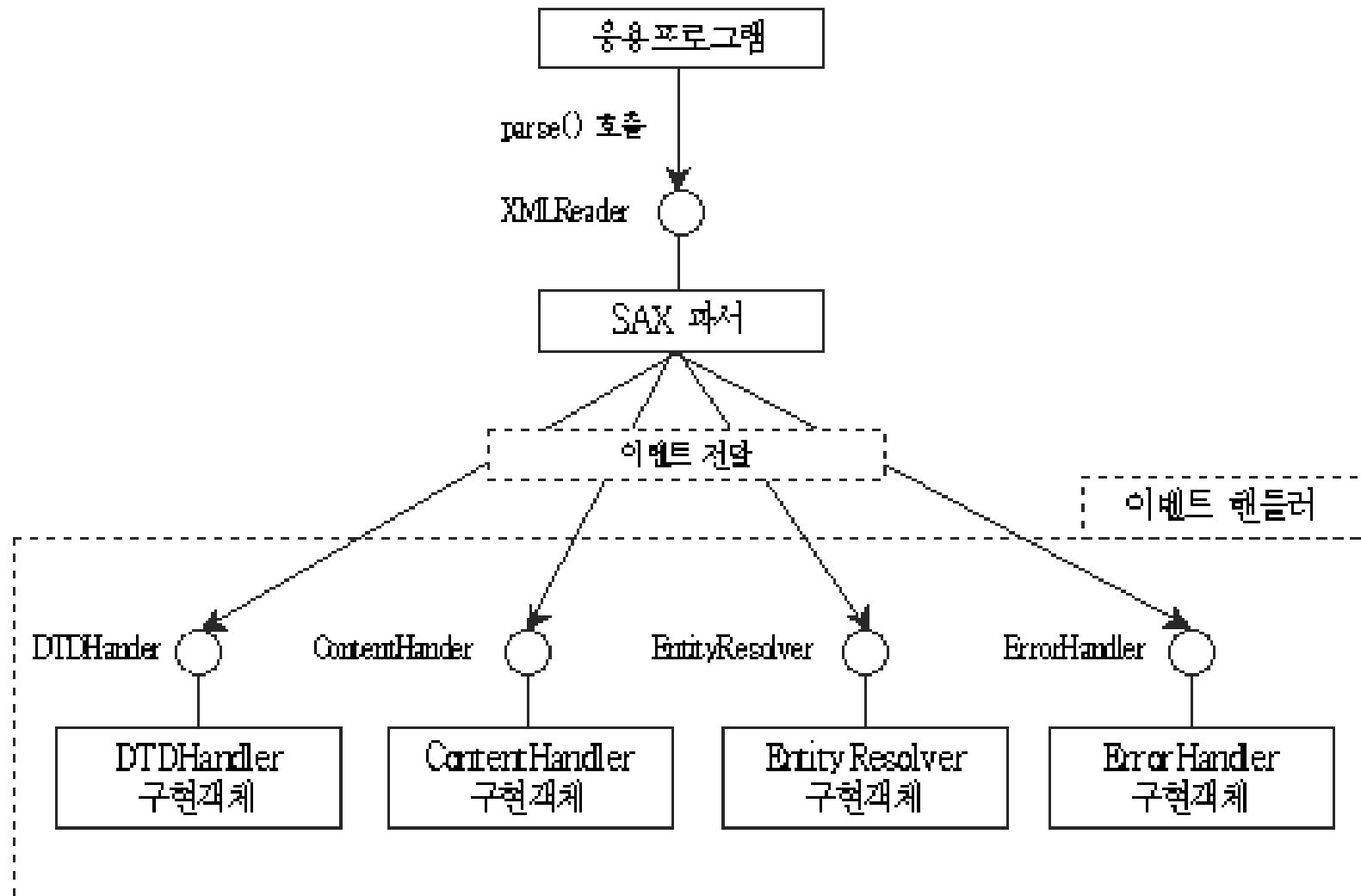


1. SAX 역사 및 개념

- SAX 파서와 DOM 파서의 비교

구분	DOM 파서	SAX 파서
파서 해석 결과	객체 트리	이벤트
데이터 접근 방식	네비게이션(스크롤) 접근	1회 순차적인 접근
메모리	XML 문서량에 비례하여 증가	XML 문서량과 상관없이 일정 메모리 사용
데이터 용량	적을 경우 유리	상관 없음
데이터 재사용	가능	불가능(재파싱해서 사용), 저장 메커니즘 필요
유리한 작업 형태	네비게이션 검색 작업	DB 입력, 순차적 검색 작업, XML 문서 유효성 검사 작업

2. SAX 파서의 작동 메커니즘



3. SAX 스펙 관련 Java 패키지

- J2SE 1.4부터 XML을 위한 파서 제공

패키지 명	설명
org.xml.sax	SAX의 Core 패키지라고 하며, SAX의 기본 인터페이스들과 클래스들을 포함하고 있다.
org.xml.sax.ext	SAX의 확장 인터페이스들을 포함하고 있다.
org.xml.sax.helpers	org.xml.sax 패키지의 인터페이스들을 구현하는 구현 클래스들을 포함하고 있다.

3. SAX 스펙 관련 Java 패키지

- SAX 1.0 및 SAX 2.0의 Core 인터페이스와 클래스

org.xml.sax 패키지의 구성		
타입	이름	설명
interface	Attributes	엘리먼트의 속성들을 표현하는 인터페이스이다.
interface	ContentHandler	XML 문서 파싱 도중에 마크업이나 문자 데이터를 만났을 때 호출되는 이벤트 핸들러의 인터페이스이다.
interface	DTDHandler	XML 문서 파싱 도중에 DTD에 포함된 노테이션 선언이나 언파스드 엔티티 선언을 만났을 때 호출되는 이벤트 핸들러의 인터페이스이다.
interface	EntityResolver	외부 DTD 서브셋, 외부 파라미터 엔티티 참조, 내부 파스드 파라미터 엔티티 참조 등을 만났을 때 호출되는 이벤트 핸들러의 인터페이스이다.
interface	ErrorHandler	XML 문서의 파싱 도중에 만난 파스 에러를 처리하는 이벤트 핸들러의 인터페이스이다.
interface	Locator	SAX 이벤트가 어느 엔티티의 어느 위치에서 발생했는지의 정보에 접근하는데 필요한 메소드들을 정의하고 있는 인터페이스이다.

3. SAX 스펙 관련 Java 패키지

org.xml.sax 패키지의 구성		
타입	이름	설명
interface	XMLFilter	XML Reader를 상속하는 인터페이스이다. 이 인터페이스는 XMLReader와 마찬가지로 SAX 파서를 표현하지만, 외부 소스로부터 XML 문서를 파싱하여 이벤트를 발생하는 것이 아니라 다른 XMLReader로부터 이벤트를 받는다.
interface	XMLReader	SAX 파서를 표현하는 인터페이스이다. SAX 파서의 작동 중에 호출될 이벤트 핸들러들을 등록하고, 파싱을 구동하는 데 필요한 메소드들을 정의하고 있다.
class	InputSource	XML 문서를 구성하는 외부 엔티티 하나를 표현하는 클래스이다.
class	SAXException	SAX 파서의 작동 중에 발생하는 예외 클래스이다.
class	SAXNotRecognizedException	SAXException을 상속하는 클래스로서, SAX 파서가 알 수 없는 피처나 프로퍼티에 대한 요청을 했을 때 발생한다.
class	SAXNotSupportedException	SAXException을 상속하는 클래스로서, SAX 파서가 지원하지 않는 피처나, 프로퍼티를 설정했을 때 발생한다.
class	SAXParseException	SAXException을 상속하는 클래스로서, XML 문서 파싱 도 중 파스에러나 워닝(Warning)이 일어났을 때 발생한다.

3. SAX 스펙 관련 Java 패키지

- SAX 확장(Extensions) 패키지

org.xml.sax.ext 패키지의 구성		
타입	이름	설명
interface	DecHandler	XML 문서 파싱 도중 DTD에 포함된 엔티티 선언, 엘리먼트 리스트 선언, 엔티티 선언을 만났을 때 호출되는 이벤트 핸들러의 인터페이스이다.
interface	LexicalHandler	XML 문서 파싱 도중에 만나는 마크업들 중 ContentHandler에 보고되지 않은 마크업들을 처리하기 위한 이벤트 핸들러의 인터페이스이다.

3. SAX 스펙 관련 Java 패키지

- 유틸 클래스 패키지 (SAX 2.0 Core 인터페이스)

org.xml.sax.helpers 패키지의 구성		
타입	이름	설명
class	AttributesImpl	Attributes 인터페이스의 구현 클래스이다.
class	DefaultHandler	ContentHandler, DTDHandler, EntityResolver, ErrorHandler 인터페이스를 모두 구현하는 클래스이다.
class	LocatorImpl	Locator 인터페이스의 구현 클래스이다.
class	NamespaceSupport	SAX 파서에 의해 사용되는 네임스페이스 로직을 캡슐화한 클래스이다.
class	ParserAdapter	SAX 1.0의 인터페이스인 Parser를 XMLReader로서 사용할 수 있도록 하는 클래스이다. XMLReader 인터페이스와 DocumentHandler 인터페이스를 구현한다.
class	XMLFilterImpl	XMLFilter 인터페이스와 EntityResolver, DTDHandler, ContentHandler, ErrorHandler 인터페이스를 모두 구현하는 클래스이다.
class	XMLReaderAdapter	XMLReader를 SAX 1.0의 인터페이스인 Parser로서 사용할 수 있도록 하는 클래스이다. Parser 인터페이스와 ContentHandler 인터페이스를 구현한다.
class	XMLReaderFactory	XMLReader의 인스턴스를 생성하는 팩토리 클래스이다.

5. SAX 파서의 생성

- SAX 파서 객체 생성

```
import javax.xml.parsers.*;
```

```
...
```

```
//SAX 파서 공장 생성
```

```
SAXParserFactory factory = SAXParserFactory.newInstance();
```

```
//SAX 파서 생성
```

```
SAXParser parser = factory.newSAXParser();
```

5. SAX 파서의 생성

- SAX 파서에 다양한 기능 추가하기

//SAX 파서 공장 생성

SAXParserFactory factory = SAXParserFactory.newInstance();

//① DTD Validation 검사를 하게함

factory.setValidating(true);

//② NameSpace를 해석하도록 함

factory.setNamespaceAware(true);

//③ XML Schema Validation 검사를 하게함(NameSpace 해석 기능 필수)

factory.setFeature("http://apache.org/xml/features/validation/schema", true);

//SAX 파서 생성

SAXParser parser = factory.newSAXParser();

6. ContentHandler의 이벤트 처리

- ContentHandler 인터페이스
 - SAX 파서가 XML 문서의 내용 중 **엘리먼트, 문자 데이터, 프로세싱 인스트럭션**에 관련된 이벤트를 처리

메소드명	설명
startDocument()	XML 문서의 시작을 만났을 때 발생하는 이벤트를 처리하는 메소드
endDocument()	XML 문서의 끝을 만났을 때 발생하는 이벤트를 처리하는 메소드
startElement(...)	엘리먼트의 시작 태그를 만났을 때 발생하는 이벤트를 처리하는 메소드
stopElement(...)	엘리먼트의 끝 태그를 만났을 때 발생하는 이벤트를 처리하는 메소드
characters(...)	문자 데이터를 만났을 때 발생하는 이벤트를 처리하는 메소드

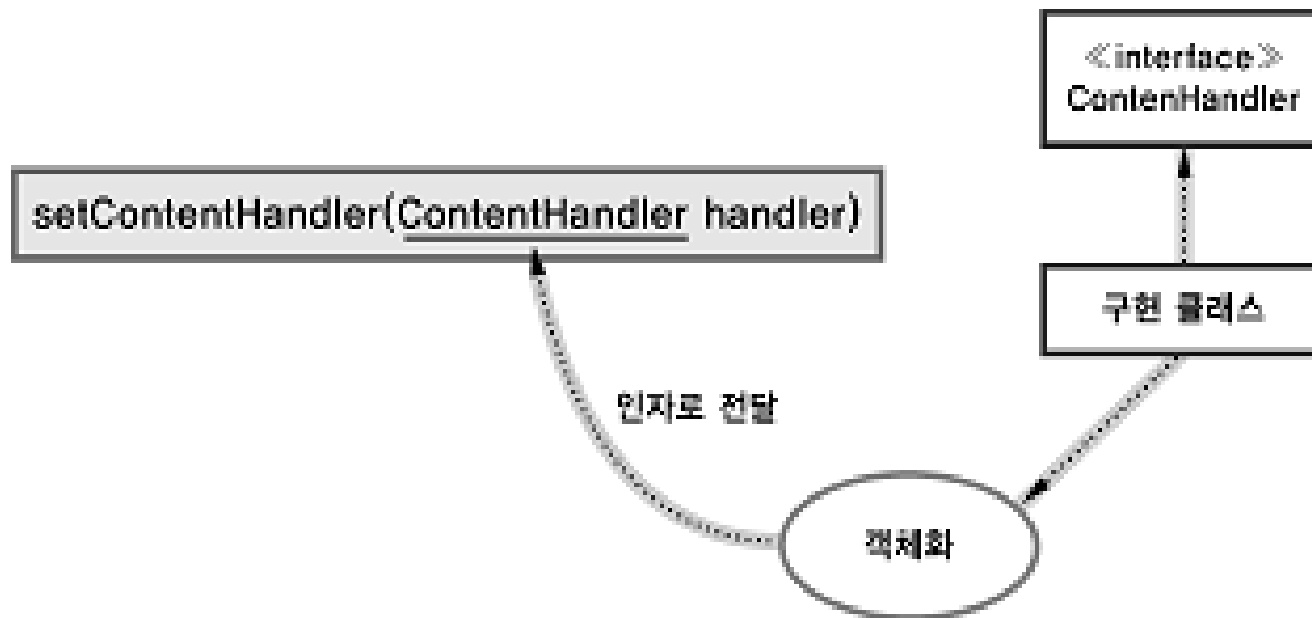
6. ContentHandler의 이벤트 처리

- ContentHandler 의 구현 클래스 작성
 - 사용자 정의 ContentHandler 객체는 반드시 org.xml.sax.ContentHandler 인터페이스를 구현

Ex) MyContentHandler.java

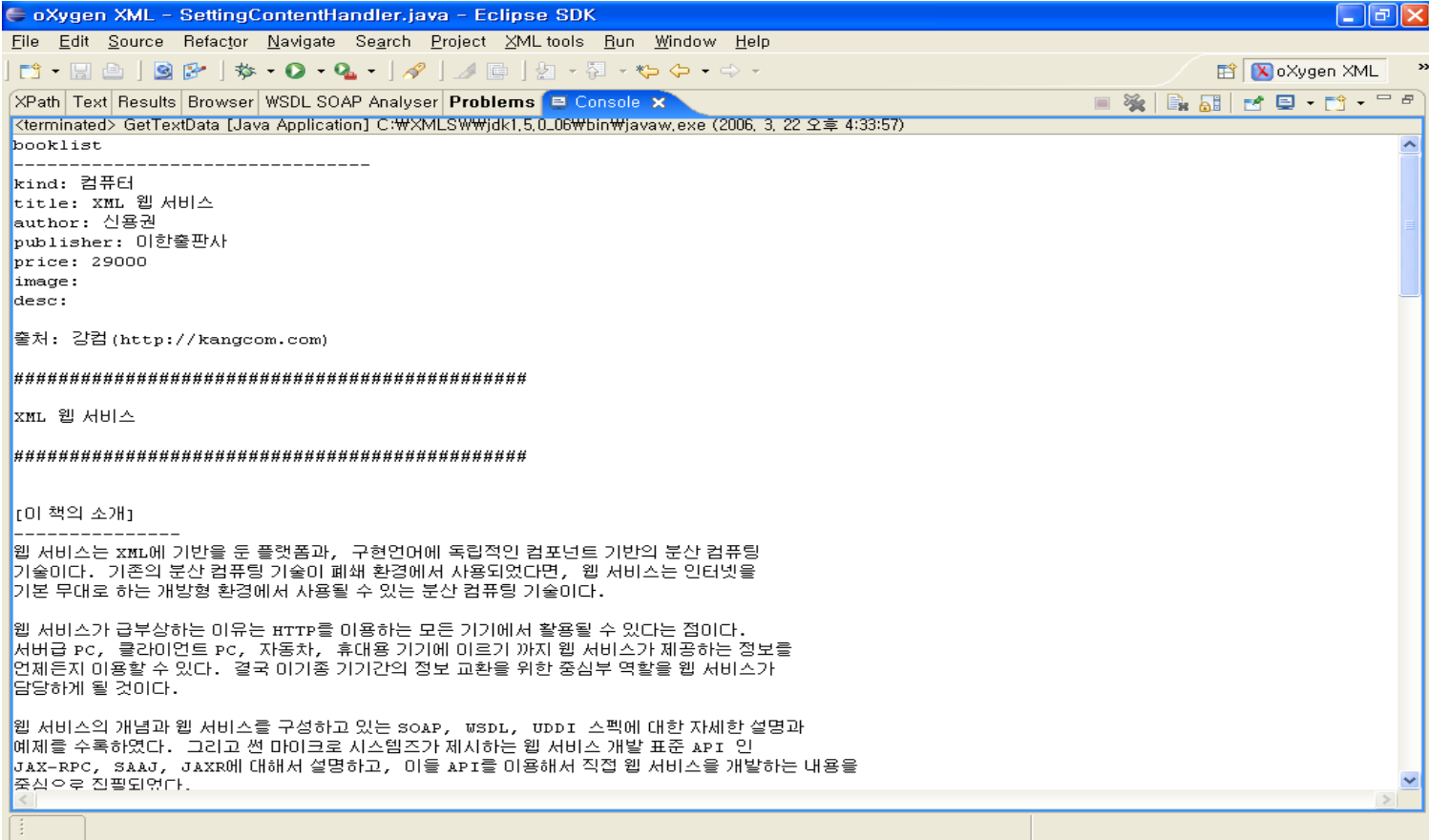
6. ContentHandler의 이벤트 처리

- ContentHandler 객체 등록
 - org.xml.sax.XMLReader를 구현한 객체를 얻음
 - Ex) XMLReader reader = parser.getXMLReader();



6. ContentHandler의 이벤트 처리

Ex) SettingContentHandler.java, bml.xml, bml.dtd



The screenshot shows the Eclipse IDE with the 'Console' window open. The title bar reads 'oXygen XML - SettingContentHandler.java - Eclipse SDK'. The console output shows a terminated Java application (javaw.exe) from 2006, 3, 22 at 4:33:57. The output is an XML document with the following structure:

```
<terminated> GetTextData [Java Application] C:\XMLS\jdk1.5.0_06\bin\javaw.exe (2006. 3. 22 오후 4:33:57)
booklist
-----
kind: 컴퓨터
title: XML 웹 서비스
author: 신용권
publisher: 미한출판사
price: 29000
image:
desc:

출처: 강컴 (http://kangcom.com)

#####

XML 웹 서비스

#####

[이 책의 소개]
-----
웹 서비스는 XML에 기반을 둔 플랫폼과, 구현언어에 독립적인 컴포넌트 기반의 분산 컴퓨팅
기술이다. 기존의 분산 컴퓨팅 기술이 폐쇄 환경에서 사용되었다면, 웹 서비스는 인터넷을
기본 무대로 하는 개방형 환경에서 사용될 수 있는 분산 컴퓨팅 기술이다.

웹 서비스가 급부상하는 이유는 HTTP를 이용하는 모든 기기에서 활용될 수 있다는 점이다.
서버급 PC, 클라이언트 PC, 자동차, 휴대용 기기에 이르기 까지 웹 서비스가 제공하는 정보를
언제든지 이용할 수 있다. 결국 이기간 기기간의 정보 교환을 위한 중심부 역할을 웹 서비스가
담당하게 될 것이다.

웹 서비스의 개념과 웹 서비스를 구성하고 있는 SOAP, WSDL, UDDI 스펙에 대한 자세한 설명과
예제를 수록하였다. 그리고 썬 마이크로 시스템즈가 제시하는 웹 서비스 개발 표준 API 인
JAX-RPC, SAAJ, JAXR에 대해서 설명하고, 이들 API를 이용해서 직접 웹 서비스를 개발하는 내용을
중심으로 진필되었다.
```

7. DTDHandler의 이벤트 처리

- DTDHandler 인터페이스
 - SAX 파서가 DTD 내용 중에서 노테이션 선언과 언파스드 엔티티 선언을 만났을 때 발생

메소드명	설명
notationDecl(...)	노테이션 선언을 만났을 때 발행하는 이벤트를 처리하는 메소드
unparsedEntityDecl(...)	언파스드 엔티티 선언을 만났을 때 발생하는 이벤트를 처리하는 메소드

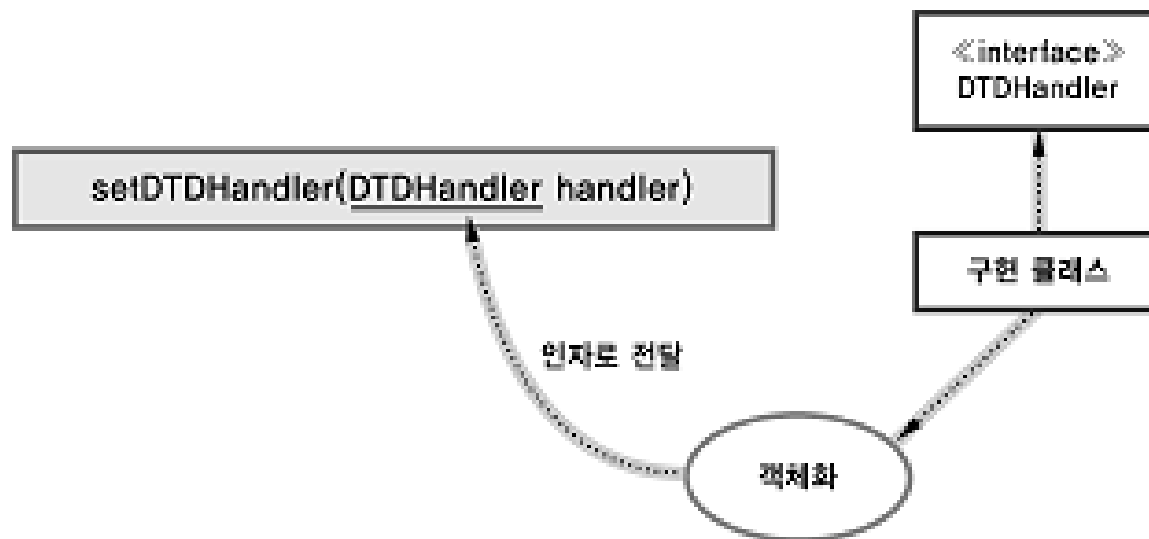
7. DTDHandler의 이벤트 처리

- DTDHandler 구현 클래스 작성
 - 사용자 정의 DTDHandler 객체는 반드시 `org.xml.sax.DTDHandler` 인터페이스를 구현

Ex) `MyDTDHandler.java`, `NotationDecl.java`, `UnparsedEntityDecl.java`

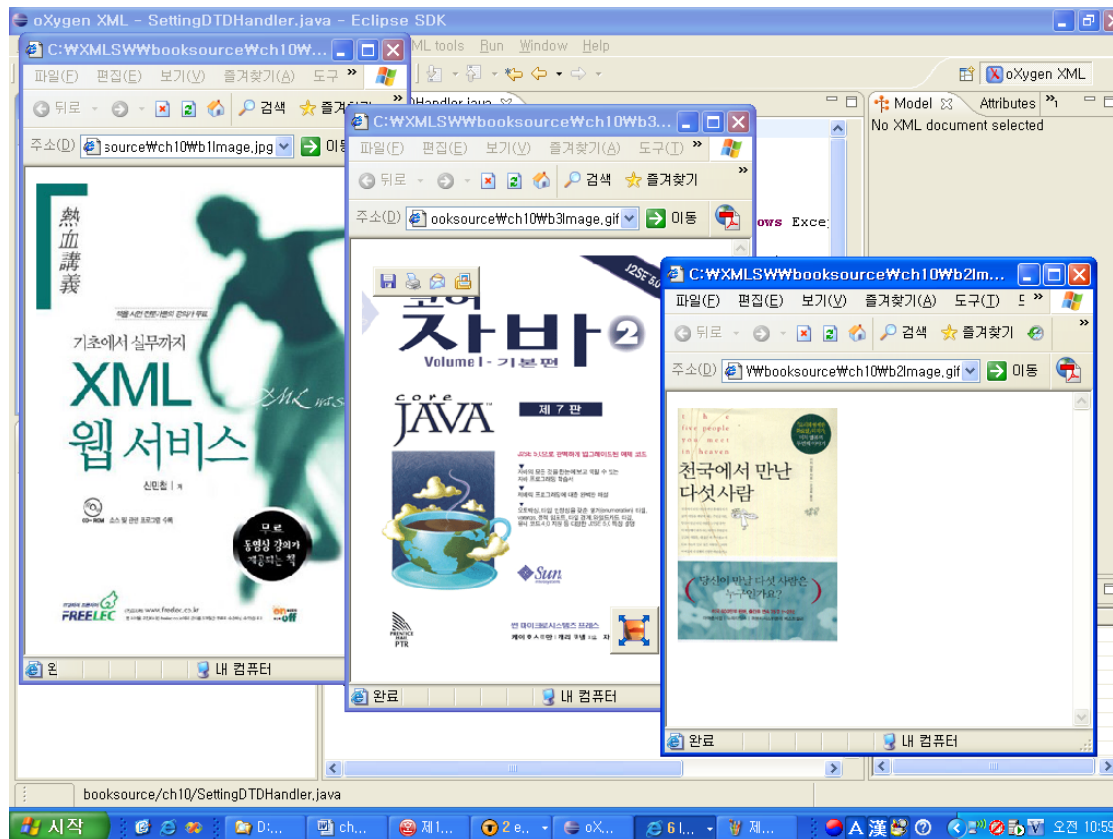
7. DTDHandler의 이벤트 처리

- DTDHandler 객체 등록
 - org.xml.sax.XMLReader 인터페이스를 구현한 객체를 SAXParser 객체로부터 얻음
- Ex) XMLReader reader = parser.getXMLReader();



7. DTDHandler의 이벤트 처리

ex) SettingDTDHanlder.java,



8. EntityResolver의 이벤트 처리

- EntityResolver 인터페이스
 - SAX 파서가 DTD나 XML 문서 내용 중에서 외부 파스드 엔티티에 대한 참조처리 하는 이벤트

메소드명	설명
resolveEntity(...)	외부 파스드 엔티티에 대한 참조를 만났을 때 발생하는 이벤트를 처리하는 메소드

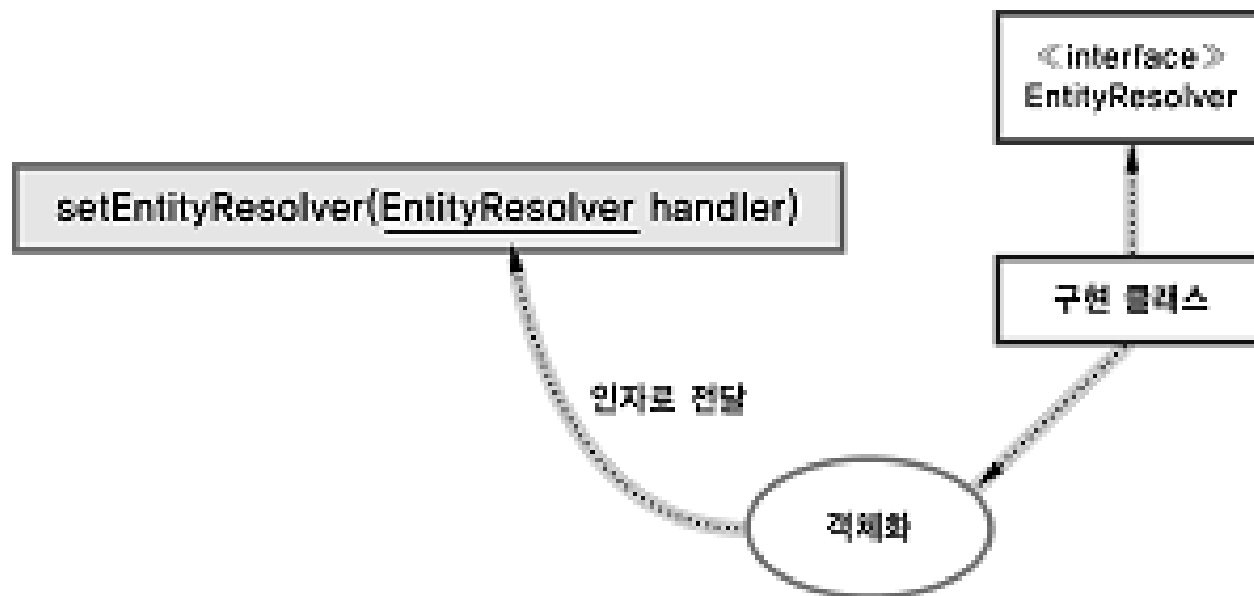
8. EntityResolver의 이벤트 처리

- EntityResolver 구현 클래스 작성
 - 사용자 정의 EntityResolver 객체는 반드시 org.xml.sax.EntityResolver 인터페이스를 구현한 객체

Ex) MyEntityResolver.java

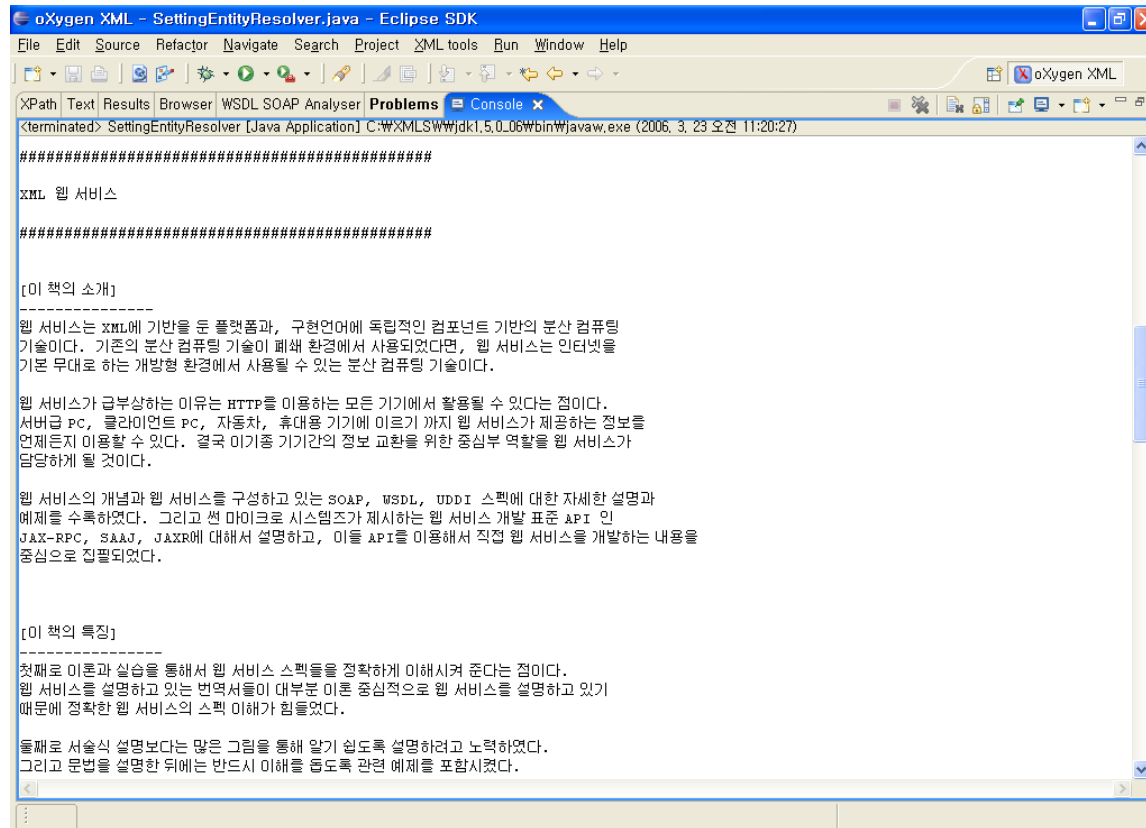
8. EntityResolver의 이벤트 처리

- EntityResolver 객체 등록
 - org.xml.sax.XMLReader 인터페이스를 구현한 객체를 SAXParser 객체로부터 얻음
- Ex) XMLReader reader = parser.getXMLReader();



8. EntityResolver의 이벤트 처리

Ex) SettingEntityResolver.java



```
oXygen XML - SettingEntityResolver.java - Eclipse SDK
File Edit Source Refactor Navigate Search Project XML tools Run Window Help
XPath Text Results Browser WSDL SOAP Analyser Problems Console
<terminated> SettingEntityResolver [Java Application] C:\XXMLSW\jdk1.5.0_06\bin\javaw.exe (2006. 3. 23 오전 11:20:27)

#####

XML 웹 서비스

#####

[이 책의 소개]
-----
웹 서비스는 XML에 기반을 둔 플랫폼과, 구현언어에 독립적인 컴포넌트 기반의 분산 컴퓨팅
기술이다. 기존의 분산 컴퓨팅 기술이 폐쇄 환경에서 사용되었다면, 웹 서비스는 인터넷을
기본 무대로 하는 개방형 환경에서 사용될 수 있는 분산 컴퓨팅 기술이다.

웹 서비스가 급부상하는 이유는 HTTP를 이용하는 모든 기기에서 활용될 수 있다는 점이다.
서버급 PC, 클라이언트 PC, 자동차, 휴대용 기기에 이르기 까지 웹 서비스가 제공하는 정보를
언제든지 이용할 수 있다. 결국 이기종 기기간의 정보 교환을 위한 중심부 역할을 웹 서비스가
담당하게 될 것이다.

웹 서비스의 개념과 웹 서비스를 구성하고 있는 SOAP, WSDL, UDDI 스펙에 대한 자세한 설명과
예제를 수록하였다. 그리고 썬 마이크로 시스템즈가 제시하는 웹 서비스 개발 표준 API 인
JAX-RPC, SAAJ, JAXR에 대해서 설명하고, 이들 API를 이용해서 직접 웹 서비스를 개발하는 내용을
중심으로 집필되었다.

[이 책의 특징]
-----
첫째로 이론과 실습을 통해서 웹 서비스 스펙들을 정확하게 이해시켜 준다는 점이다.
웹 서비스를 설명하고 있는 번역서들이 대부분 이론 중심으로 웹 서비스를 설명하고 있기
때문에 정확한 웹 서비스의 스펙 이해가 힘들었다.

둘째로 서술식 설명보다는 많은 그림을 통해 알기 쉽도록 설명하려고 노력하였다.
그리고 문법을 설명한 뒤에는 반드시 이해를 돕도록 관련 예제를 포함시켰다.
```

9. ErrorHandler의 이벤트 처리

- ErrorHandler 인터페이스
 - SAX 파서가 DTD나 XML 문서 파싱 중에 파싱 에러를 만났을 때 발생하는 이벤트를 처리

메소드명	설명
<code>fatalError()</code>	XML 문서가 Well-Formed하게 작성 되지 않았을 경우(즉 XML 권고안을 지키지 않았을 경우)에 발생하는 이벤트를 처리하는 메소드
<code>error()</code>	XML 문서가 유효한 문서가 아닐 경우(즉 DTD나 XML Schema 정의 구조에 맞게 작성 되지 않았을 경우)에 발생하는 이벤트를 처리하는 메소드
<code>warning()</code>	<code>fatalError</code> 및 <code>error</code> 이외의 이벤트를 처리하는 메소드

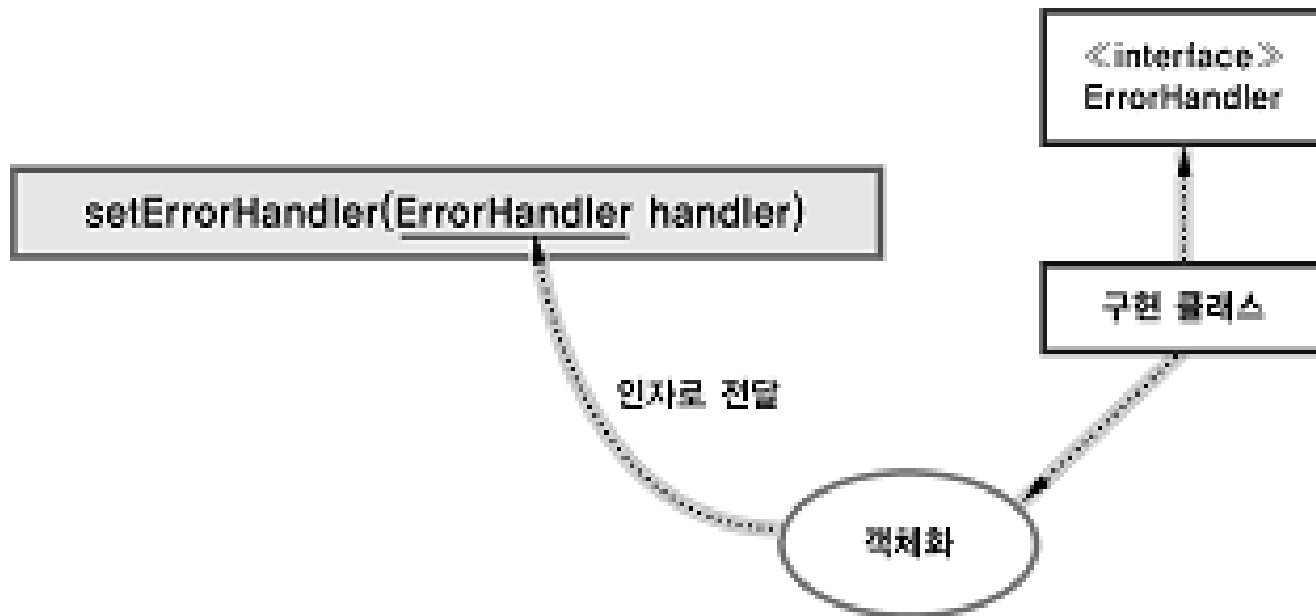
9. ErrorHandler의 이벤트 처리

- ErrorHandler 구현 클래스 작성
 - 사용자 정의 ErrorHandler 객체는 반드시 org.xml.sax.ErrorHandler 인터페이스를 구현한 객체

Ex) MyErrorHandler.java

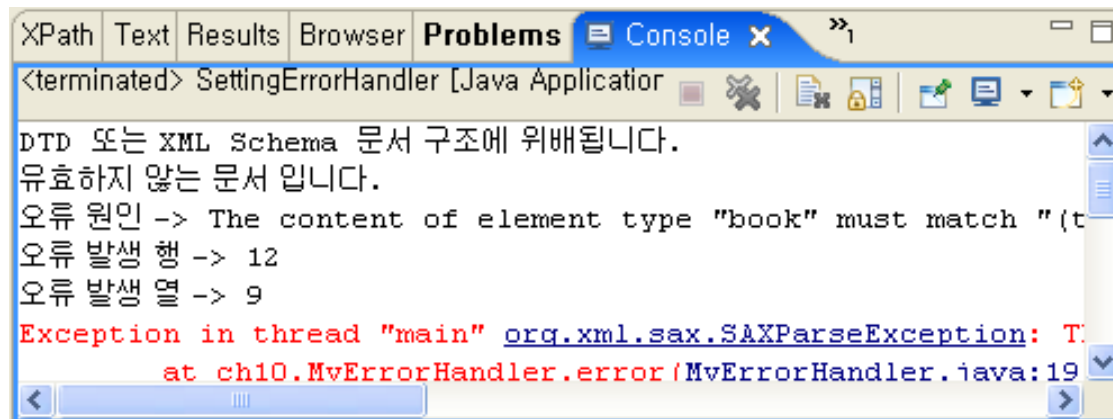
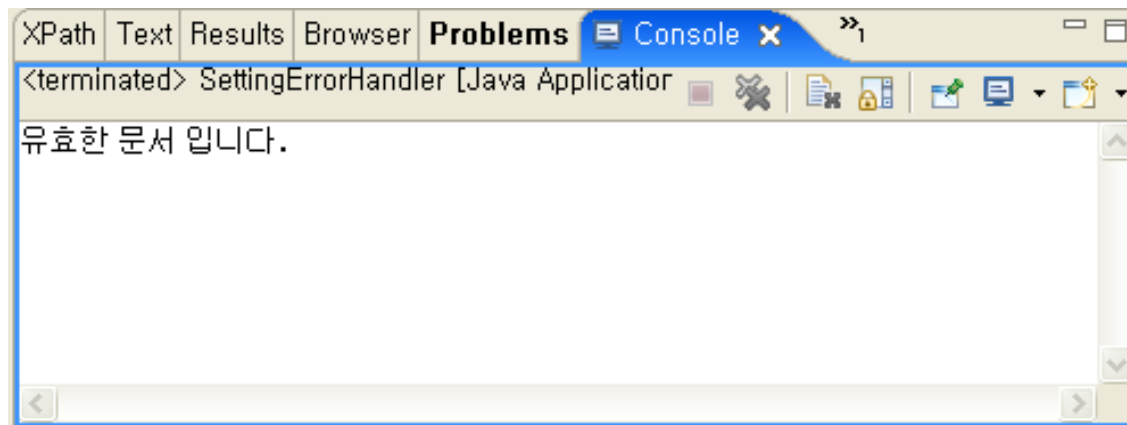
9. ErrorHandler의 이벤트 처리

- ErrorHandler 객체 등록
 - org.xml.sax.XMLReader 인터페이스를 구현한 객체를 SAXParser 객체로부터 얻음
- Ex) XMLReader reader = parser.getXMLReader();



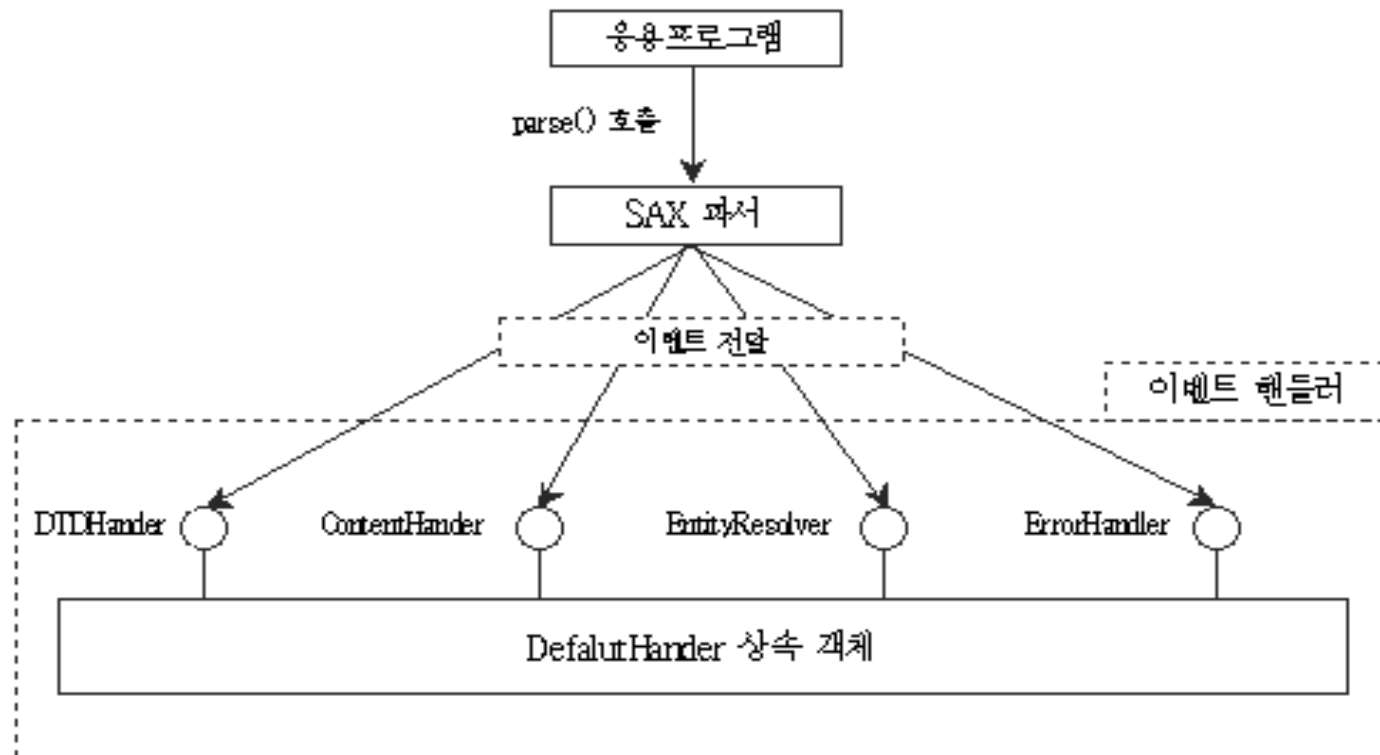
9. ErrorHandler의 이벤트 처리

Ex) SettingErrorHandler.java



10. DefaultHandler의 이벤트 처리

- DefaultHandler 클래스
 - SAX 파서에서 발생하는 몇몇 이벤트들에 대해서만 처리하고자 할 때 사용하는 어댑터 클래스
 - ContentHandler, DTDHandler, EntityResolver, ErrorHandler를 모두 구현한 클래스

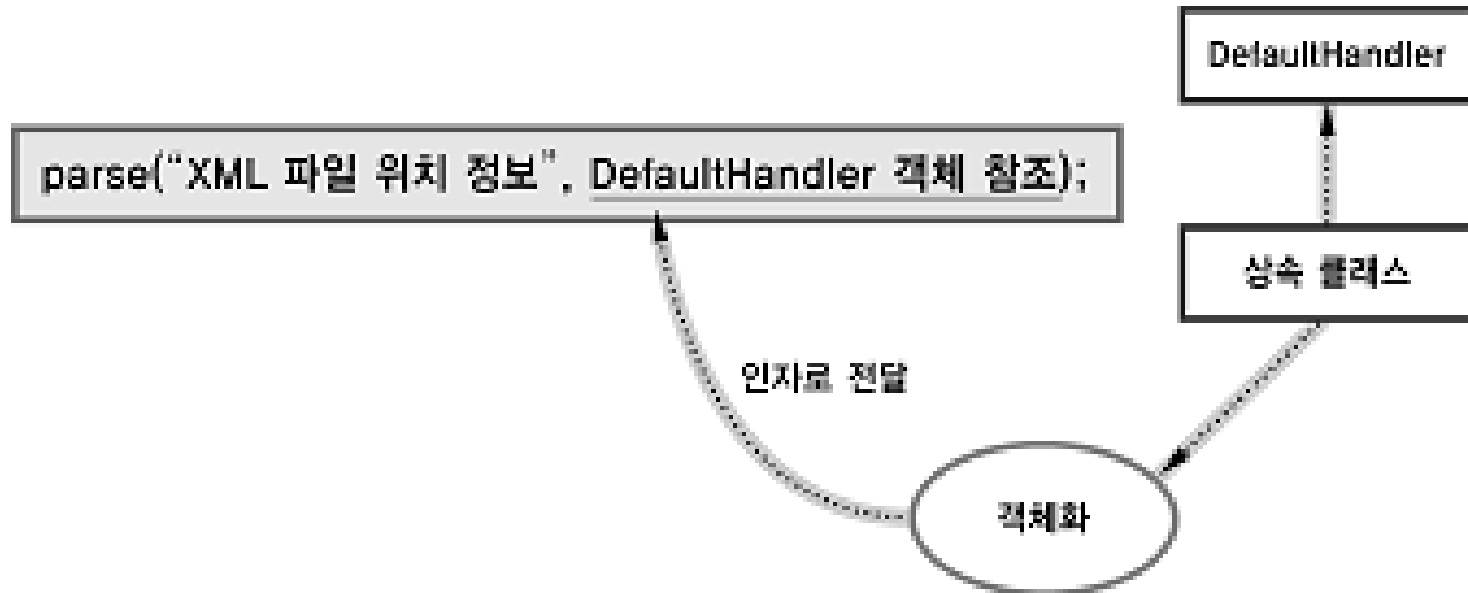


10. DefaultHandler의 이벤트 처리

- DefaultHandler 상속 클래스 작성
 - 사용자 정의 DefaultHandler 객체는 반드시 `org.xml.sax.helpers.DefaultHandler` 클래스를 상속한 객체
 - Ex) `FindBookEventHandler.java`

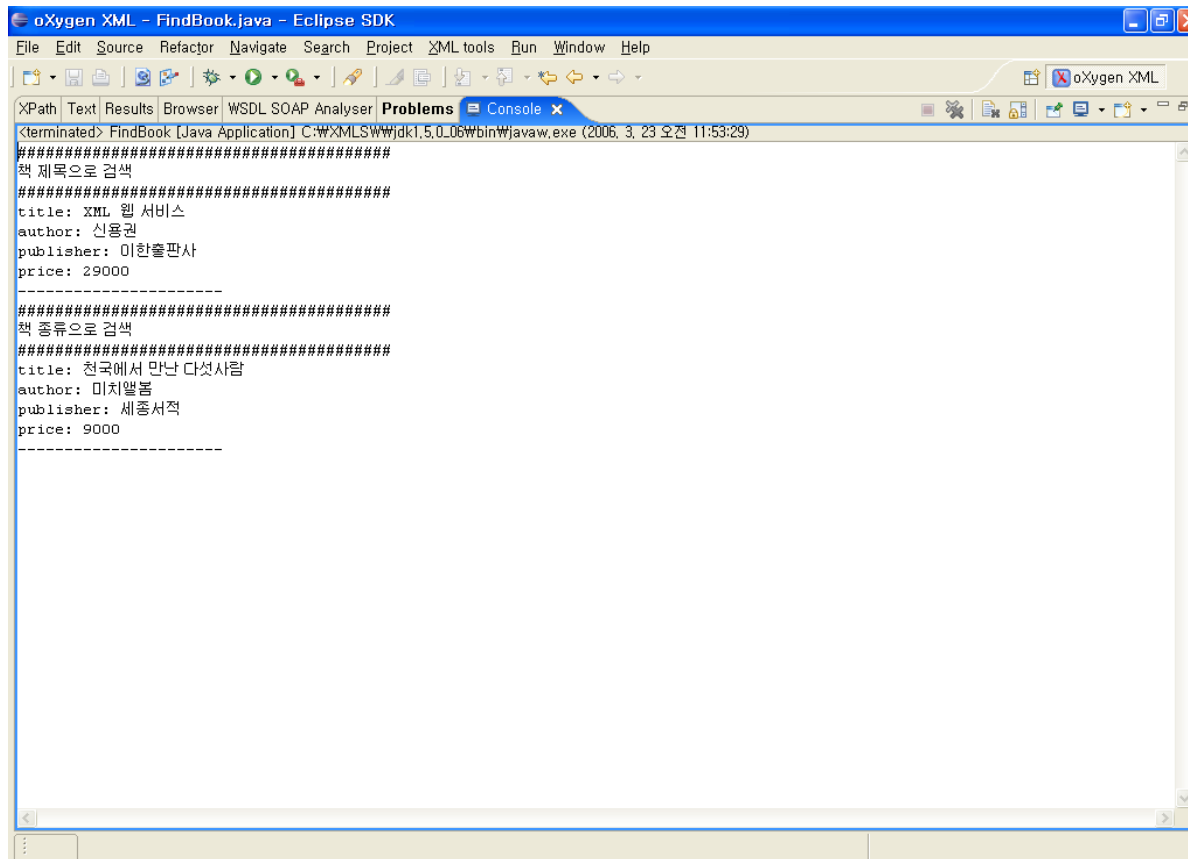
10. DefaultHandler의 이벤트 처리

- DefaultHandler 객체 등록
 - SAXParser 객체의 **parse()** 메소드를 사용



10. DefaultHandler의 이벤트 처리

Ex) FindBook.java



```
oXygen XML - FindBook.java - Eclipse SDK
File Edit Source Refactor Navigate Search Project XML tools Run Window Help

XPath Text Results Browser WSDL SOAP Analyser Problems Console x
<terminated> FindBook [Java Application] C:\XMLSDK\jdk1.5.0_06\bin\javaw.exe (2006. 3. 23 오전 11:53:29)
#####
책 제목으로 검색
#####
title: XML 웹 서비스
author: 신용권
publisher: 미한출판사
price: 29000
-----
#####
책 종류로 검색
#####
title: 천국에서 만난 다섯사람
author: 미치앨론
publisher: 세종서적
price: 9000
-----
```

11. 다양한 형태의 XML 문서 파싱

- XML 문서 파싱

- javax.xml.parsers.SAXParser의 parse()

```
//로컬 하드 디스크에 있는 XML 문서 파싱
parse(String uri, DefaultHandler dh)
parse(File f, DefaultHandler dh)
// 네트워크를 통해 전달되는 XML 문서를 파싱
parse(InputSource is, DefaultHandler dh)
parse(InputStream is, DefaultHandler dh)
parse(InputStream is, DefaultHandler dh, String systemId)
```

- javax.xml.parsers.SAXParser의 parse()

```
//로컬 하드 디스크에 있는 XML 문서 파싱
parse(String systemId)
// 네트워크를 통해 전달되는 XML 문서를 파싱
parse(InputSource input)
```