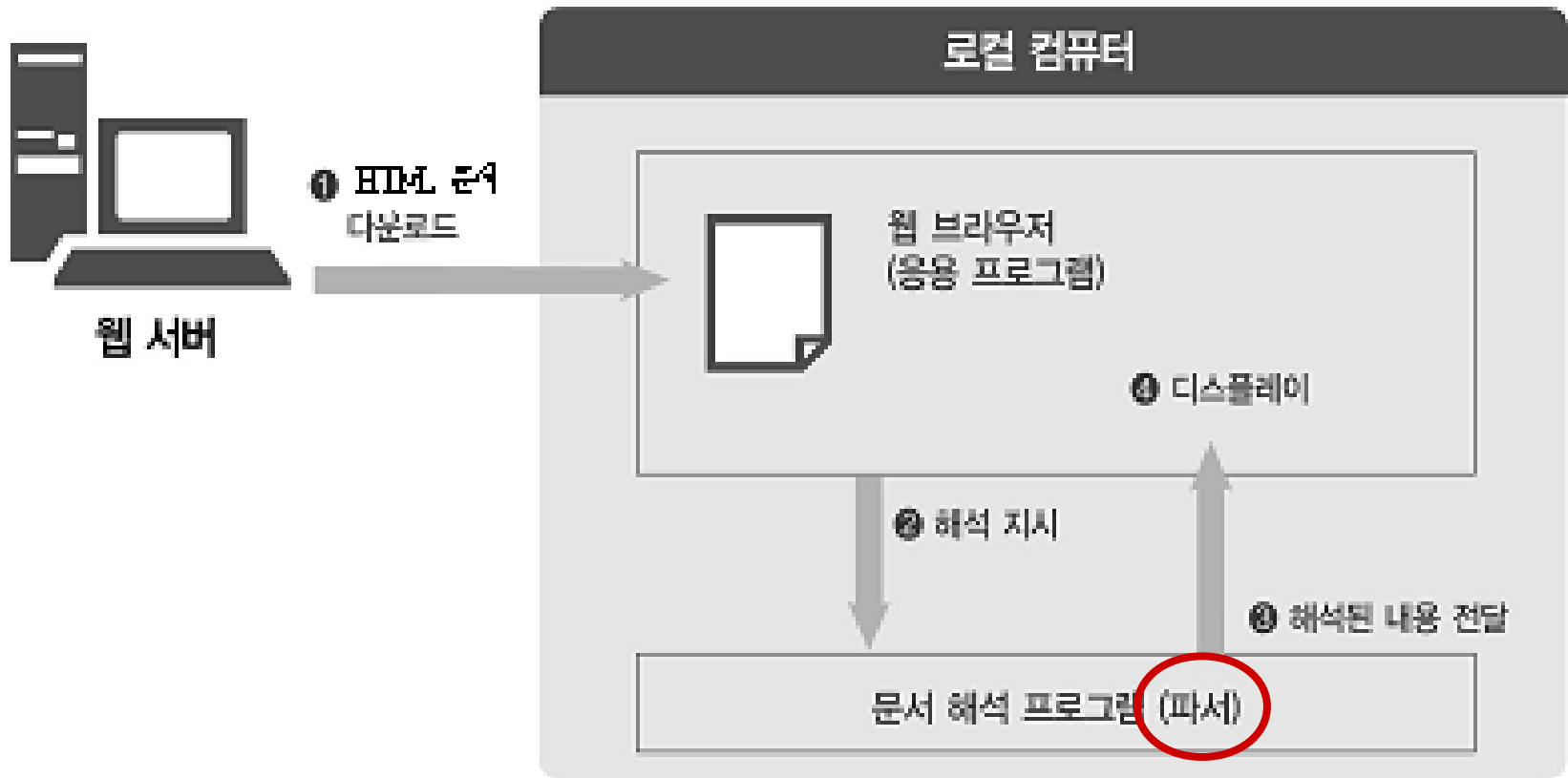


제 9 장 DOM 기반 문서 처리 프로그래밍

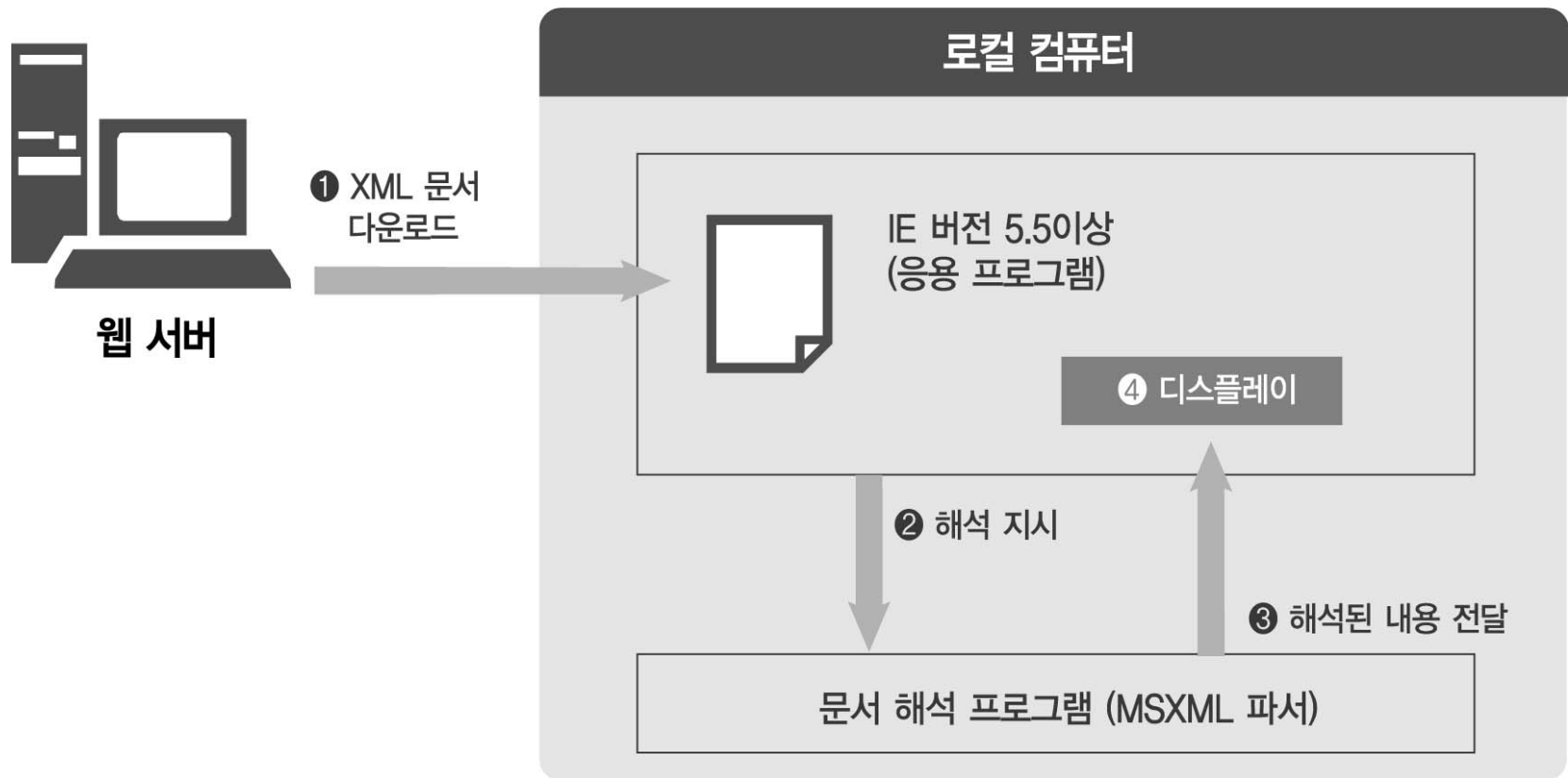
- 9-1. XML 문서 처리 이해
- 9-2. XML 파서의 역할
- 9-3. XML 파서에 종류
- 9-4. DOM 역사 및 개념
- 9-5. DOM 스펙
- 9-6. DOM Core 파트의 구조
- 9-7. DOM 파서 생성
- 9-8. XML 문서 파싱과 Document 객체 생성
- 9-9. XML 문서 검색
- 9-10. DOM 객체 수정
- 9-11. DOM 객체 추가
- 9-12. DOM 객체 제거
- 9-13. DOM 객체를 파일로 저장
- 9-14. XML 문서 뷰어 개발

9-1. XML 문서 처리 이해



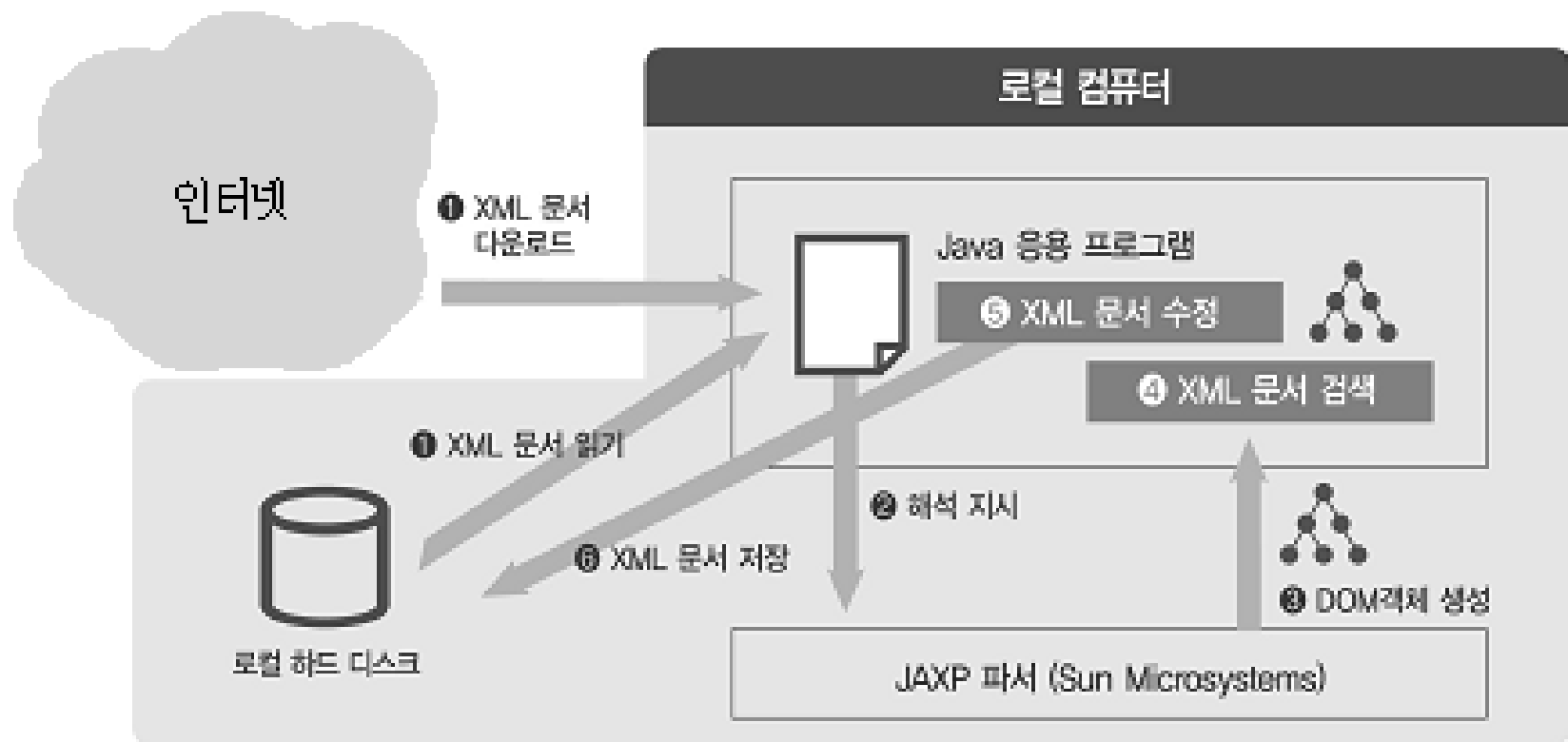
9-1. XML 문서 처리 이해

- 인터넷 익스플로러 : 5.5부터 MSXML 파서 설치



9-1. XML 문서 처리 이해

- JAXP(Java API for XML Processing) 파서 처리 과정



9-2. XML 파서의 역할

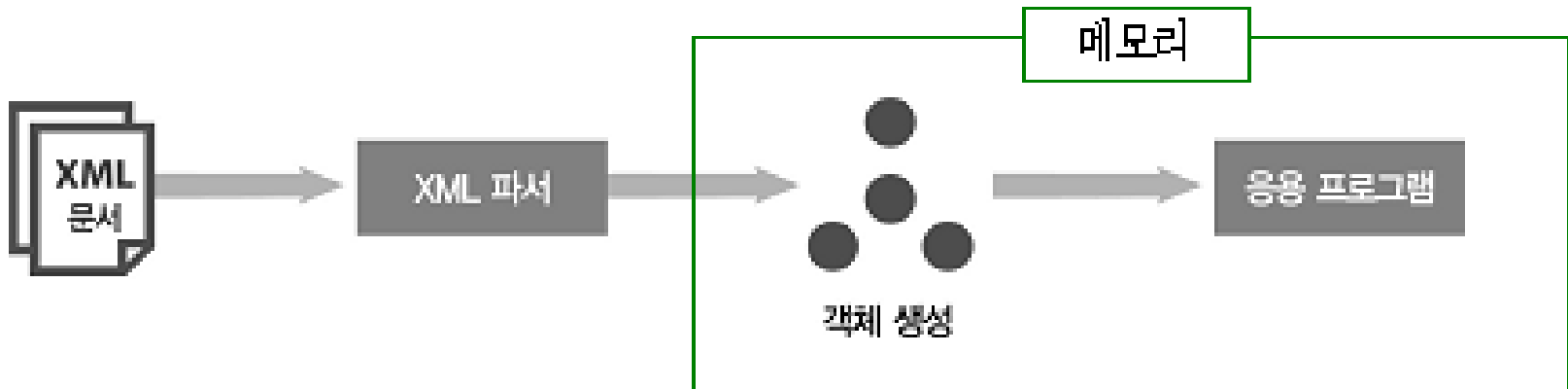
- 파서(Parser) : 파싱(Parsing)을 수행하는 프로그램
 - 파싱 : 주어진 문장을 분석하거나 문법적 관계를 해석하는 행위
- XML 파서 : XML 문서를 읽고, 해석하여 태그명, 속성명, 속성값 및 엘리먼트 내용을 분리시켜 주는 프로그램
- 자바에서 사용하는 XML 파서
 - Apache Group의 Xerces 파서
 - IBM의 XML4J 파서
 - Sun, MS, Oracle ...

9-3. XML 파서의 종류

- DOM(Document Object Model) 파서
 - XML 문서를 해석한 후, 해석 결과를 메모리에 DOM(Document Object Model) 이라는 객체 트리 구조로 생성시키는 파서
 - 데이터를 검색, 수정, 삭제 가능
- SAX(Simple API for XML) 파서
 - XML 문서를 해석과 동시에, 특정 이벤트를 발생시키는 파서
 - 이벤트는 이벤트 처리 핸들러에서 해석되고, 데이터를 응용프로그램에게 전달
- J2SE 1.4 부터는 Xerces 파서를 내장

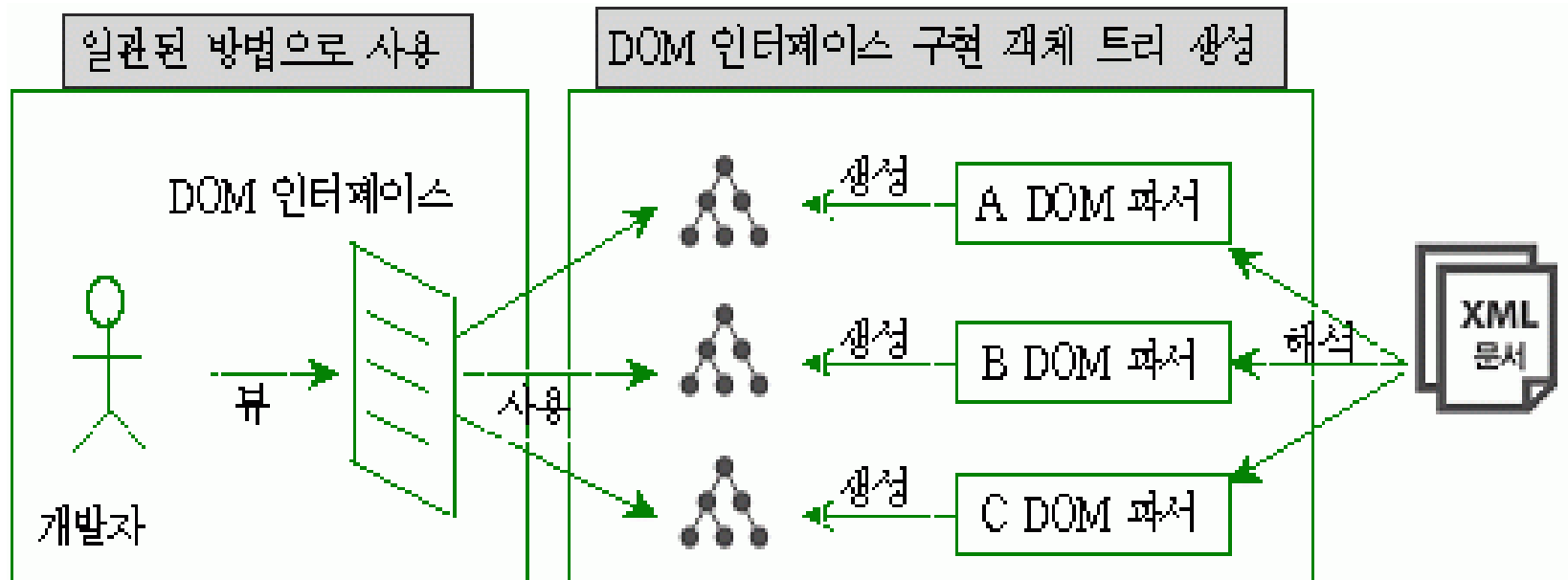
9-4. DOM 역사 및 개념

- DOM의 역사
 - XML 및 HTML 문서를 응용프로그램에서 사용하기 위한 API 규격.
 - DOM이란 문서의 각 부분들을 객체로 표현한 API
 - 1997년 W3C에서 워킹그룹을 결성
- DOM의 개념



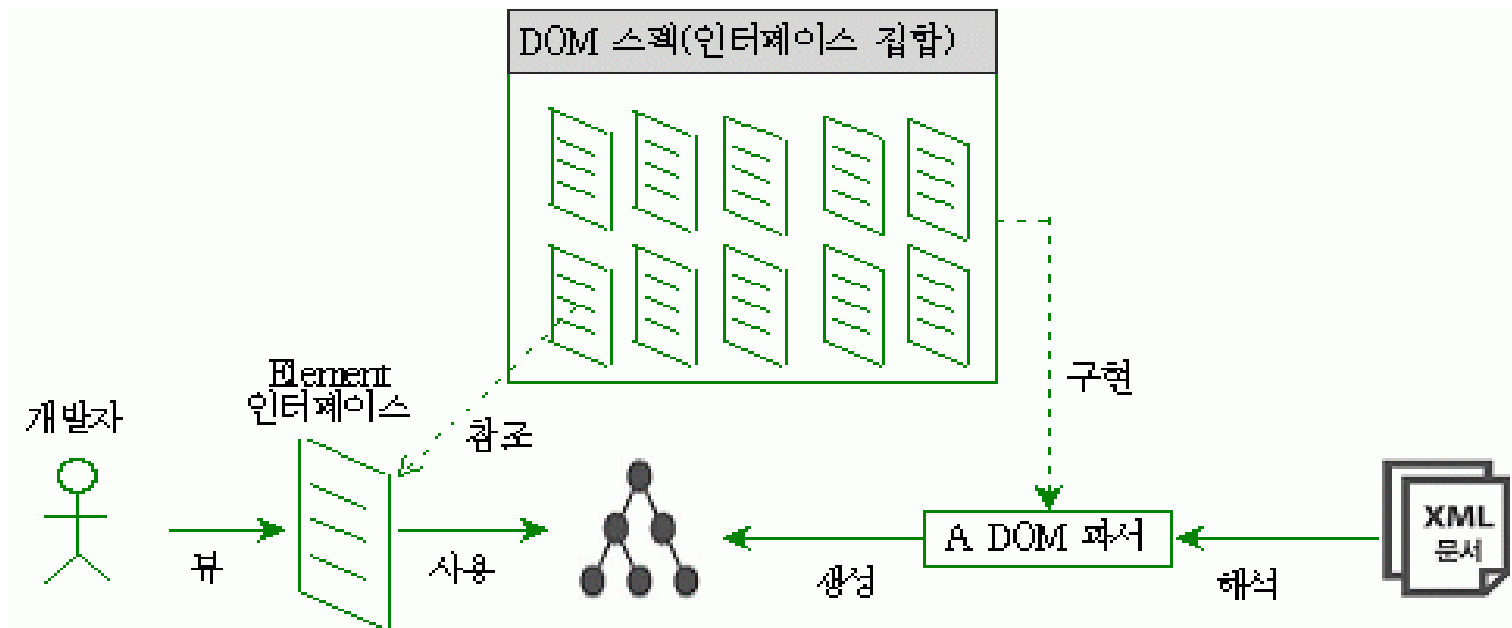
9-4. DOM 역사 및 개념

- W3C에서 XML 파서가 기본적으로 만들어야 할 객체 유형을 인터페이스 형태의 스펙으로 지정 → DOM



9-5. DOM 스펙

- DOM 스펙
 - 특정 플랫폼과 특정 프로그래밍 언어에 종속적이지 않는 인터페이스 언어 → IDL(Interface Definition Language)
 - DOM Level 1, DOM Level 2, DOM Level 3



9-5. DOM 스펙

- DOM Level 1
 - 1998년 10월에 발표

파트 이름	설명
Core 파트	구조화된 문서를 표현하는 기초 인터페이스(fundamental interface)들의 규격. XML 문서만을 위한 확장 인터페이스(extended interface)들도 포함
HTML 파트	HTML 문서를 더 쉽게 다룰 수 있는 하이레벨 인터페이스들의 규격

9-5. DOM 스펙

- DOM Level 2
 - 1999년 1월에 발표 된 네임스페이스, 뷰, 스타일시트, 이벤트, 문서 범위 지정, 노드 횡단을 지원

스펙 이름	포함하는 파트
Core	▶ Core 파트
HTML	▶ HTML 파트
Style	▶ Style Sheets 파트
	▶ CSS 파트
View	▶ Views 파트
Events	▶ Events 파트
Traversal and Range	▶ Traversal 파트
	▶ Range 파트

9-5. DOM 스펙

- DOM Level 3
 - DOM Level 2 내용 보안, 추상 XML Schema, Load and Save

스펙 이름	포함하는 파트
Core	▶ Core 파트
Load and Save	▶ Load and Save 파트
Validation	▶ Validation 파트
XPath	▶ XPath 파트
Events	▶ Events 파트
Views and Formatting	▶ Views and Formatting 파트
Abstract Schemas	▶ Abstract Schemas 파트

9-6. DOM Core 파트의 구조

- 기본 인터페이스
 - 마크업 언어로 작성된 문서를 처리하기 위한 기초적인 인터페이스

구분	인터페이스 종류	사용 용도
기본 인터페이스	Node	모든 객체 유형 인터페이스의 부모 인터페이스로, 다른 객체 유형 인터페이스가 공통적으로 가지는 메소드들이 정의되어 있다.
	NodeList	특정 노드들을 인덱싱 처리하여 반복 처리에 편리하도록 설계되어진 객체 유형 인터페이스 이다.
	DOMImplementation	Document 객체를 생성하는 메소드가 정의되어 있는 인터페이스이다.
	DocumentFragment	객체 트리에서 서브 트리를 조각으로 분리할 때 사용되는 인터페이스이다.
	Document	DOM 트리 구조의 최상위 노드로서, XML 문서 자체에 해당되는 객체 유형 인터페이스이다. 주요 메소드로 동적으로 다른 객체 유형을 생성키는 메소드가 있다.
	Element	XML 문서에서 엘리먼트에 해당되는 객체 유형 인터페이스이다.
	Attr	XML 문서에서 속성에 해당되는 객체 유형 인터페이스이다.
	CharacterData	XML 문서에서 데이터에 해당되는 객체 유형 인터페이스이다.
	Text	XML 문서에서 문자 데이터(엘리먼트의 내용)에 해당되는 객체 유형 인터페이스이다.

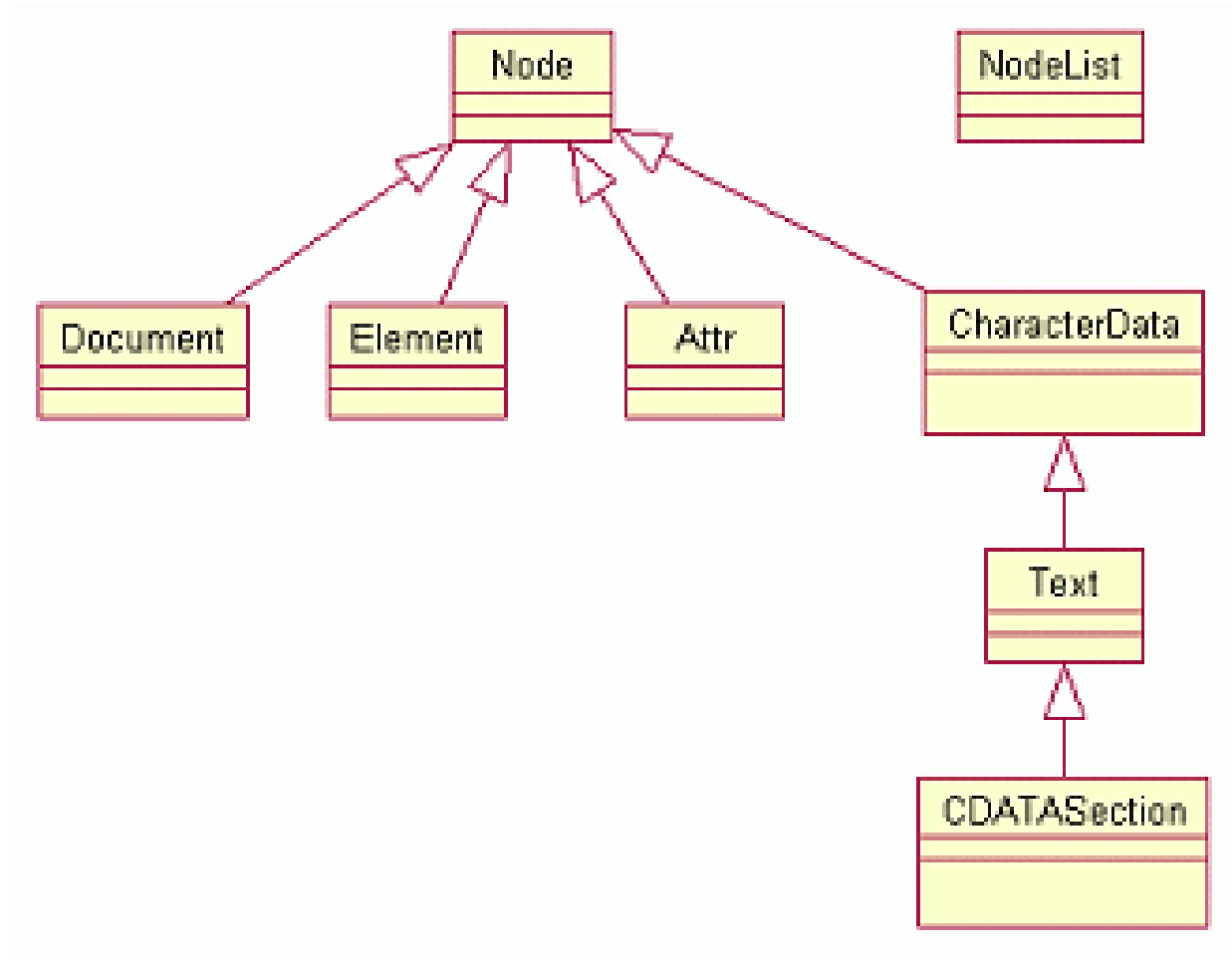
9-6. DOM Core 파트의 구조

구분	인터페이스 종류	사용 용도
	NamedNodeMap	name으로 찾을 수 있는 Node의 Collection에 해당하는 인터페이스이다.
	DOMException	DOM 트리 구조에서 발생하는 예외 클래스이다.

- 확장 인터페이스
 - XML 문서에서만 적용될 수 있는 인터페이스

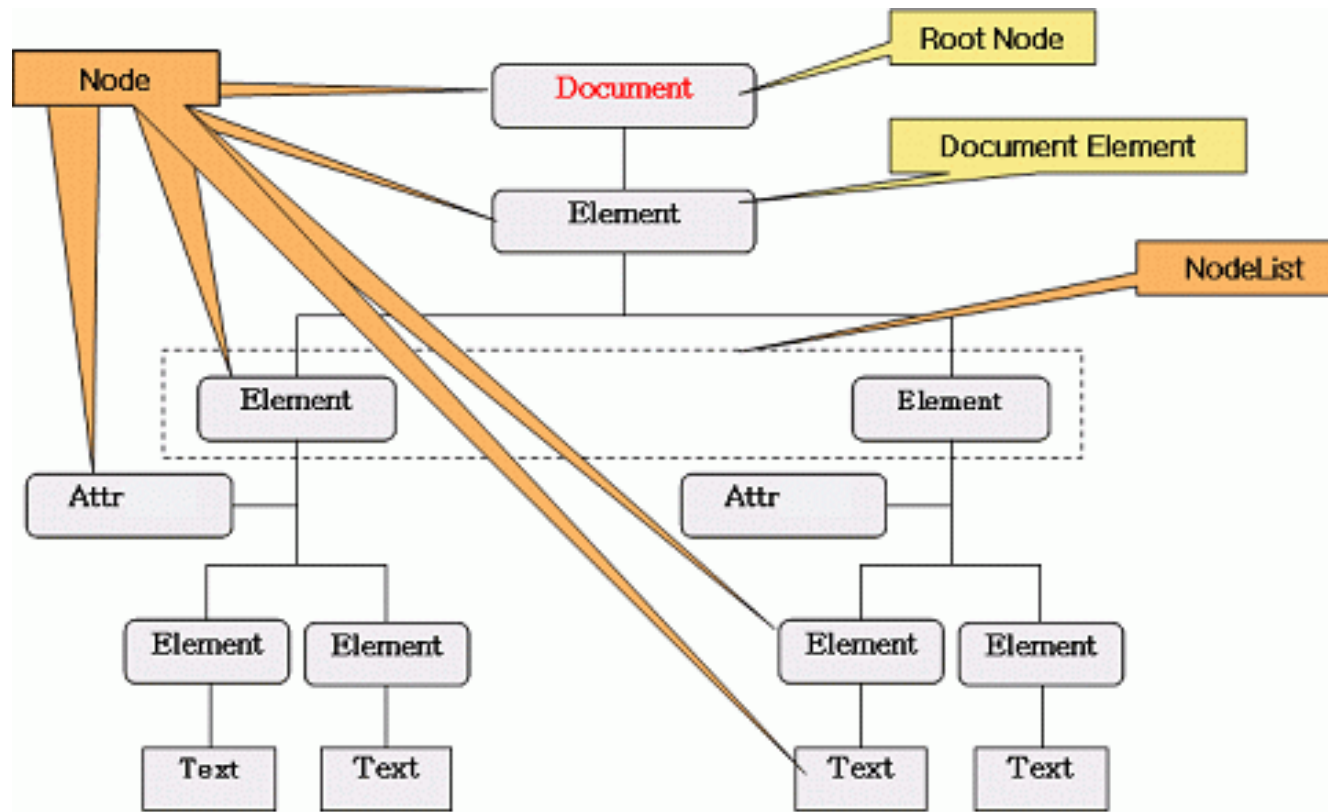
구분	인터페이스 종류	사용 용도
확장 인터페이스	CDataSection	XML 문서에서 CDATA 섹션에 해당되는 객체 유형 인터페이스이다.
	DocumentType	XML 문서에서 DTD 선언에 해당하는 객체 유형 인터페이스이다.
	Notation	DTD에서 Notation 선언에 해당하는 객체 유형 인터페이스이다.
	Entity	DTD에서 Entity 선언에 해당하는 객체 유형 인터페이스이다.
	EntityReference	XML 문서 또는 DTD에서 Entity 참조에 해당하는 객체 유형 인터페이스이다.
	ProcessingInstruction	XML 문서에서 프로세싱 지시자에 해당하는 객체 유형 인터페이스이다.

9-6. DOM Core 파트의 구조



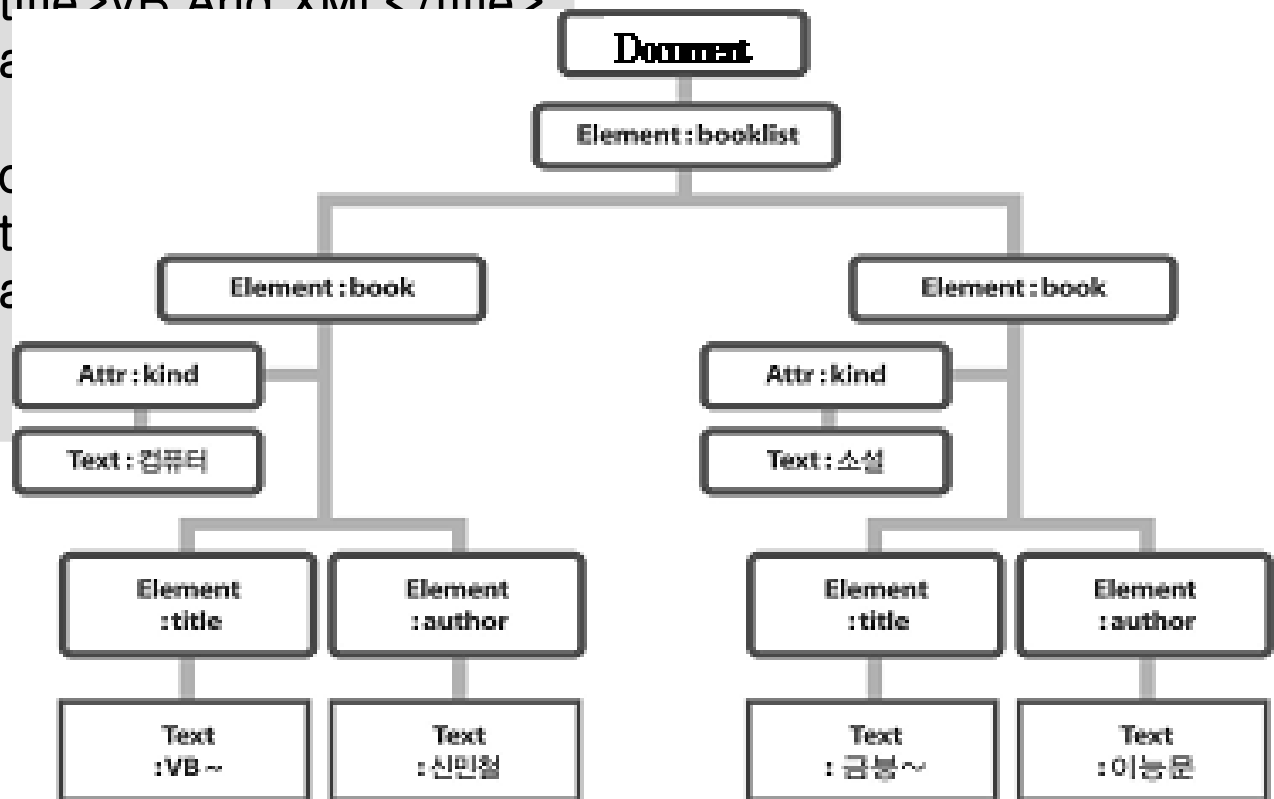
9-6. DOM Core 파트의 구조

- JAVA의 XML 관련 API
 - SUN의 JAXP(Java API for XML Processing)
 - Apache Group의 Xerces



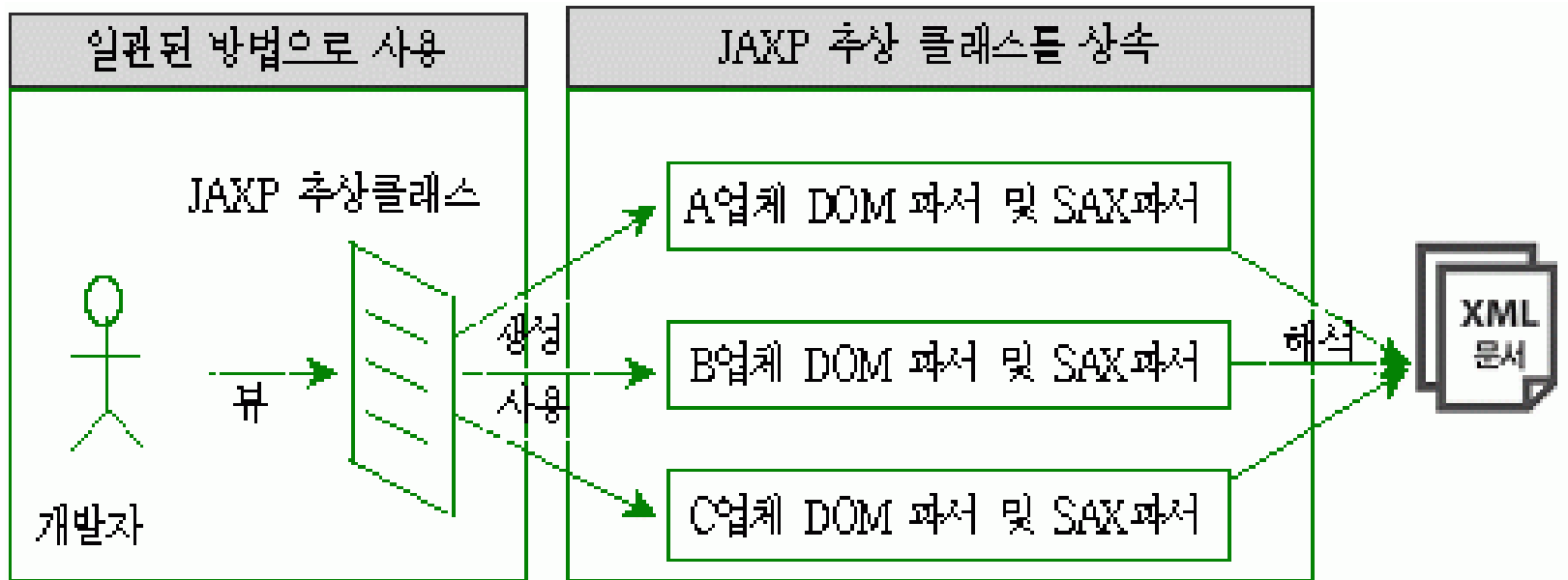
9-6. DOM Core 파트의 구조

```
<?xml version="1.0" encoding="euc-kr"?>
<booklist>
  <book kind="컴퓨터">
    <title>VB And XML </title>
    <author>신민철 </author>
  </book>
  <book kind="소설">
    <title>금붕어 </title>
    <author>이능문 </author>
  </book>
</booklist>
```



9-7. DOM 파서의 생성

- JAXP(Java API for Processing)
 - J2SE 1.4부터 포함.
 - DOM과 SAX 파서 지원



9-7. DOM 파서의 생성

추상 클래스명	설명
DocumentBuilderFactory	DocumentBuilder(DOM 파서)를 생성시키는 Factory 클래스
DocumentBuilder	DOM 파서 객체의 클래스
SAXParserFactory	SAXParser를 생성시키는 Factory 클래스
SAXParser	SAX 파서 객체의 클래스

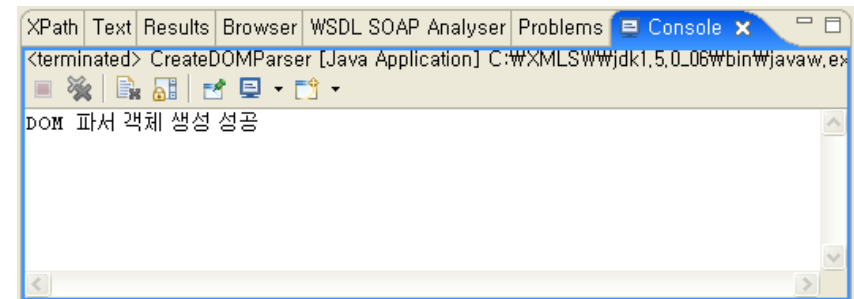
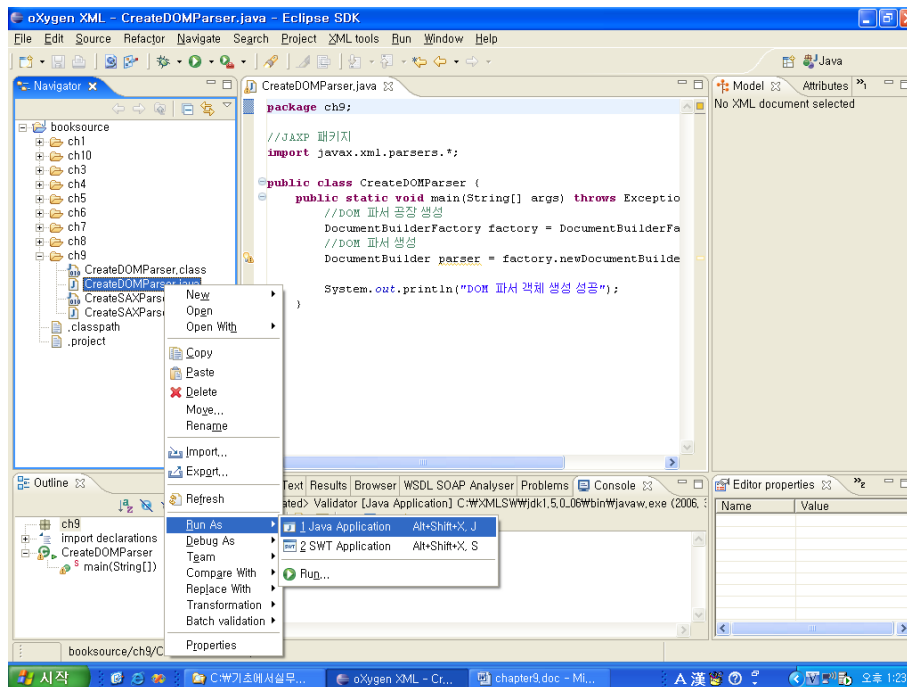
9-7. DOM 파서의 생성

- DOM 파서 생성
 - JAXP의 추상 클래스인 javax.xml.parsers.DocumentBuilder의 작
식 클래스를 javax.xml.parsers.DocumentBuilderFactory 클래스
를 사용하여 객체화

```
import javax.xml.parsers.*;  
...  
// DOM 파서 공장 생성  
DocumentBuilderFactory factory =  
DocumentBuilderFactory.newInstance();  
//DOM 파서 생성  
DocumentBuilder parser = factory.newDocumentBuilder();
```

9-7. DOM 파서의 생성

- Eclipse에서 실행
 - [oXygen XML] Perspective로 변경
 - [Navigator] 창에서 ch9/CreateDomParser.java 파일 실행



9-7. DOM 파서의 생성

- DOM 파서에 다양한 기능 추가하기
 - DocumentBuilderFactory의 Feature 속성값 변경

```
// 파서 공장 생성
DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
// ① WhiteSpace를 무시함(XML문서에 문서유형선언이 있어야 동작함)
factory.setIgnoringElementContentWhitespace(true);
// ② DTD Validation 검사를 하게 함
factory.setValidation(true);
// ③ Namespace를 해석하도록 함
factory.setNamespaceAware(true);
// 파서 생성
DocumentBuilder parser = factory.newDocumentBuilder();
```

9-7. DOM 파서의 생성

- XML 문서에서의 Whitespace

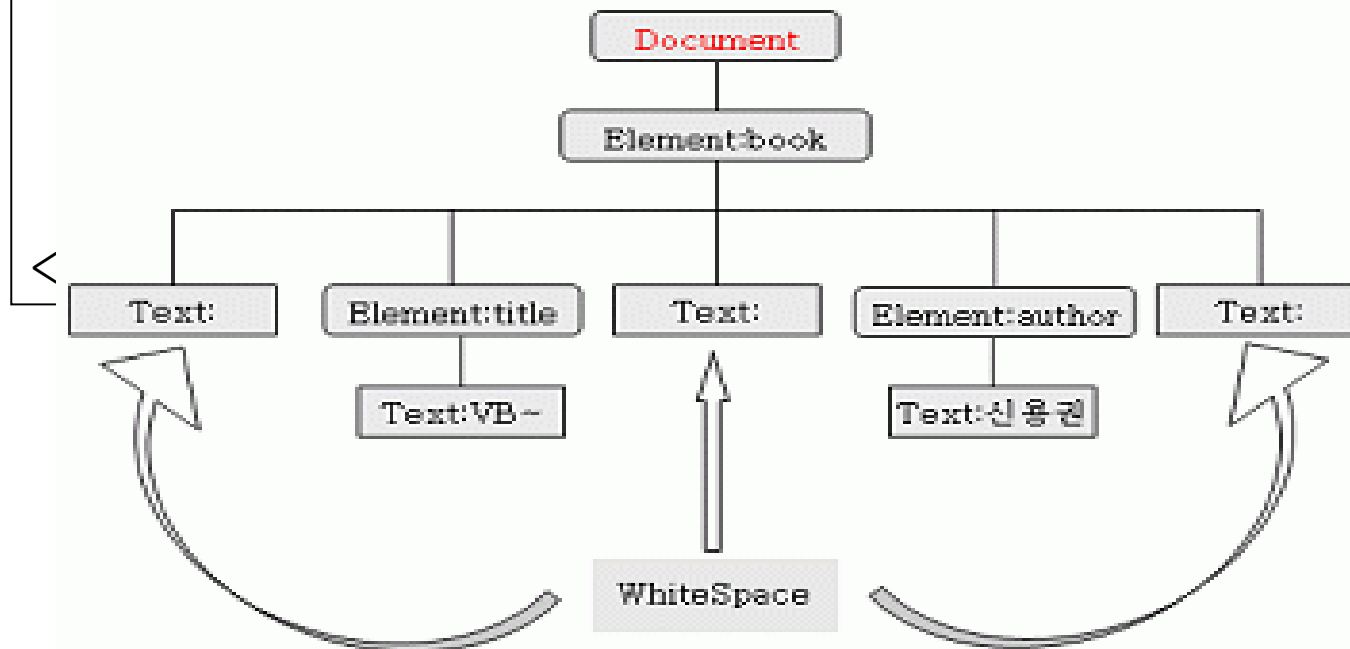
```
<?xml version="1.0" encoding="euc-kr"?>
```

```
<book> Whitespace
```

```
 Whitespace
```

```
 Whitespace
```

```
<title>VB And XML</title>
```



9-8. XML 문서 파싱과 Document 객체 얻기

1) XML 문서 파싱하기

- Document 객체를 반환하는 parse() 메소드 이용

```
// 로컬 하드 디스크에 있는 XML 문서 파싱  
public Document parse(String uri)  
public Document parse(File f)
```

```
// 네트워크를 통해 전달되는 XML 문서 파싱  
public Document parse(InputStream is)  
public Document parse(InputStream is, String systemId)  
public abstract Document parse(InputSource is)
```

- 로컬 하드 디스크에 있는 XML 문서를 파싱

```
// 로컬 하드 디스크 XML 문서 파일일 경우  
Document xmlDoc = parser.parse("C:/Temp/bml.xml");
```


9-8. XML 문서 파싱과 Document 객체 얻기

- 웹 서버에서 XML 문서를 다운로드해서 파싱

```
// 웹서버상의 XML 문서 파일일 경우  
URL url = new URL("http://www.mincheol.com/bml.xml");  
InputStream is = url.openStream();  
Document xmlDoc = parser.parse(is);
```

9-8. XML 문서 파싱과 Document 객체 얻기

2) 루트 엘리먼트 객체 참조 얻기

```
Element eRoot = xmlDoc.getDocumentElement();
```

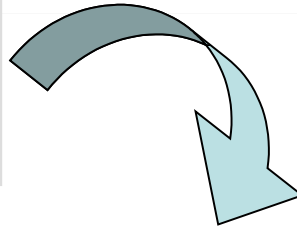
Ex) bml.xml, GetRootElement.java

9-9. XML 문서 검색

1) 첫번째 자식 객체 참조 얻기

```
Node firstChild = parentNode.getFirstChild();
```

```
<booklist>  
  <book kind="컴퓨터">  
    <title>VB And XML</title>  
    <author>신민철</author>
```



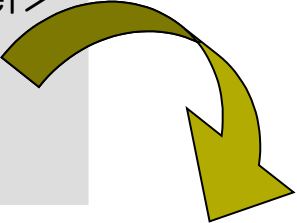
```
Element eBook = (Element) eRoot.getFirstChild();  
Element eTitle = (Element) eBook.getFirstChild();  
Text tTitle = (Text) eTitle.getFirstChild();  
String strTitle = tTitle.getData();
```

9-9. XML 문서 검색

2) 마지막 자식 객체 참조 얻기

```
Node lastChild = parentNode.getLastChild();
```

```
<book kind="컴퓨터">  
    <title>Java And XML</title>  
    <author>홍길동</author>  
    <publisher>영진 출판사</publisher>  
    <price>25000</price>  
</book>  
</booklist>
```



```
Element eBook = (Element) eRoot.getLastChild();  
Element ePrice = (Element) eBook.getLastChild();  
Text tPrice = (Text) ePrice.getFirstChild();  
String strPrice = tPrice.getData();
```

9-9. XML 문서 검색

3) 형제 자식 객체 참조 얻기

```
Node previousNode = currentNode.getPreviousSibling(); → 이전 형제  
Node nextNode = currentNode.getNextSibling(); → 다음 형제
```

Ex) GetFirstBookInfo.java

9-9. XML 문서 검색

4) 엘리먼트 객체의 속성값 얻기

```
String strAttr = objElement. getAttribute("속성명");
```

Ex) GetFirstBookInfo.java

9-9. XML 문서 검색

5) 태그 이름이 동일한 엘리먼트 객체 리스트 얻기

```
NodeList objNodeList = objElement.getElementsByTagName("태그명");
```

```
int items = objNodeList.getLength();  
for (int i=0; i<items; i++) {  
    Element element = (Element) objNodeList.item(i);  
    //~~ 처리 내용 ~~  
}
```

Ex) GetBookTitle.java

9-10. DOM 객체의 수정

1) 텍스트 객체의 문자 데이터 수정

```
objText.setData("바꿀 문자 데이터");
```

2) 속성 객체의 문자 데이터 수정

```
objElement.setAttribute("속성명", "속성값");
```

Ex) ModifyFirstBook.java

9-11. DOM 객체의 추가

1) DOM 객체의 생성

```
// 엘리먼트 객체 생성
Element objElement = objDocument.createElement("엘리먼트명");
// 텍스트 객체 생성
Text objText = objDocument.createTextNode("문자 데이터");
```

2) DOM 객체 붙이기

```
// 텍스트 객체를 부모인 엘리먼트 객체에 붙임
objElement.appendChild(objText);
// 엘리먼트 객체를 부모인 엘리먼트 객체에 붙임
parentElement.appendChild(childElement);
// 속성을 엘리먼트에 추가
objElement.setAttribute("속성명", "속성값");
```

Ex) AppendNewBook.java

9-12. DOM 객체의 제거

1) 부모 엘리먼트 객체에서 자식 엘리먼트 객체 제거

```
parentElement.removeChild(childElement);
```

Ex) RemoveFirstBook.java

9-13. DOM 객체를 파일로 저장

- javax.xml.transform.Transformer 추상 클래스를 상속한 XSLT 변환기를 이용하여 XML 문서를 파일로 저장

```
// 변환기 공장 생성
TransformerFactory factory = TransformerFactory.newInstance();
// 변환기 생성
Transformer transformer = factory.newTransformer();
```

```
// XML 선언과 문서 유형 선언 내용 설정하기
transformer.setOutputProperty(
    OutputKeys.ENCODING, "인코딩 방식");
transformer.setOutputProperty(
    OutputKeys.DOCTYPE_SYSTEM, "DTD 문서 경로");
transformer.setOutputProperty(
    OutputKeys.INDENT, "WhiteSpace 포함 여부");
```

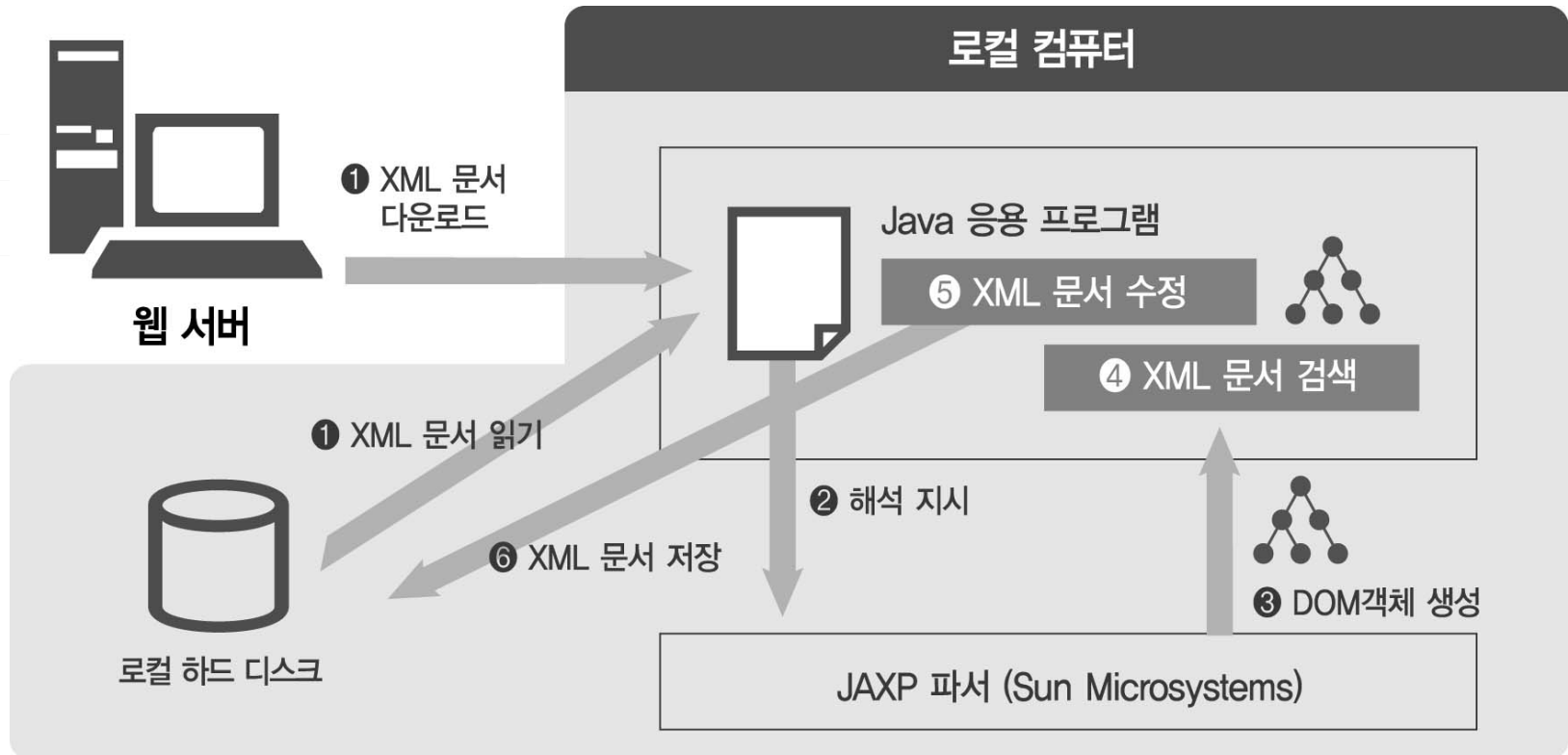
```
DOMSource source = new DOMSource(objDocument);
StreamResult result = new StreamResult(new File("로컬 파일 경로"));
transformer.transform(objSource, objResult);
```

9-13. DOM 객체를 파일로 저장

Ex) SaveDOMToFile.java

9-14. XML 문서 뷰어 개발

1) XML 문서 브라우저 이해



9-14. XML 문서 뷰어 개발

2) 유저 인터페이스 이해

버튼	설명
찾아보기	로컬 컴퓨터의 하드 디스크에 있는 파일을 찾기
로드	XML 문서를 파싱
처음	XML 문서에서 첫 번째 책 정보로 이동
이전	XML 문서에서 이전 책 정보로 이동
다음	XML 문서에서 다음 책 정보로 이동
마지막	XML 문서에서 마지막 책 정보로 이동
새 XML 문서	새로운 XML 문서 생성
수정	XML 문서에 있는 책에 대한 정보를 수정
추가	신규 책에 대한 정보를 추가(처음: 빈칸제공 두 번째: 저장)
저장	수정 및 추가를 한 후, XML 문서 파일로 저장
닫기	응용프로그램 종료

9-14. XML 문서 뷰어 개발

JTextField	객체참조변수명
문서 경로 입력	txtURL
책이름	txtTitle
종류	txtKind
저자	txtAuthor
출판사	txtPublisher
가격	txtPrice

JButton	객체참조변수명
찾아보기	btnSearch
로드	btnLoad

9-14. XML 문서 뷰어 개발

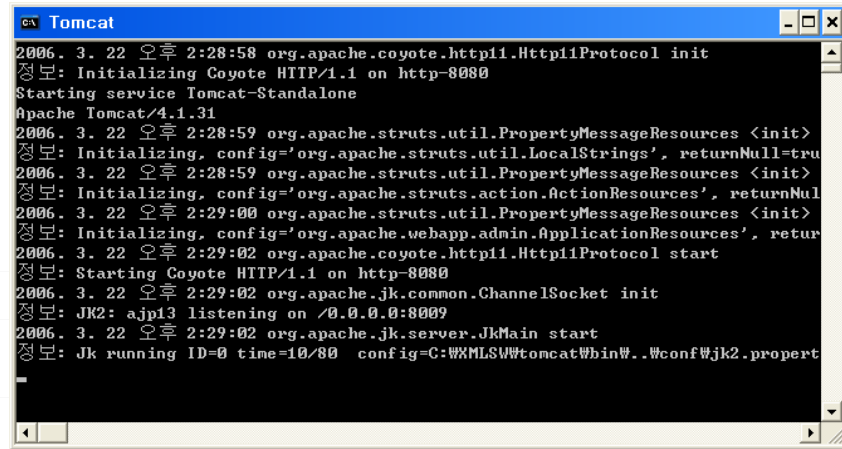
JButton	객체참조변수명
처음	btnFirst
이전	btnPrevious
다음	btnNext
마지막	btnLast
새 XML 문서	btnNew
수정	btnModify
추가	btnAdd
저장	btnSave
닫기	btnClose

Ex) XMLJFrame.java

9-14. XML 문서 뷰어 개발

4) 실행 방법

- 웹 서버 (Tomcat) 실행



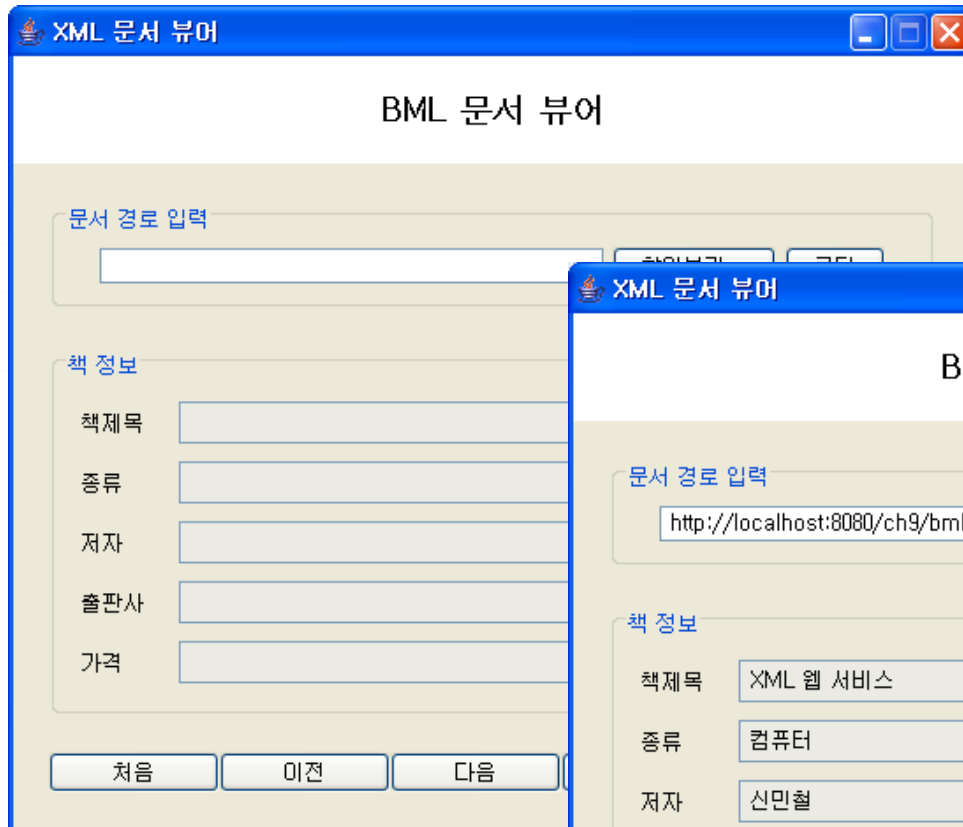
```
Tomcat
2006. 3. 22 오후 2:28:58 org.apache.coyote.http11.Http11Protocol init
정보: Initializing Coyote HTTP/1.1 on http-8080
Starting service Tomcat-Standalone
Apache Tomcat/4.1.31
2006. 3. 22 오후 2:28:59 org.apache.struts.util.PropertyMessageResources <init>
정보: Initializing, config='org.apache.struts.util.LocalStrings', returnNull=true
2006. 3. 22 오후 2:28:59 org.apache.struts.util.PropertyMessageResources <init>
정보: Initializing, config='org.apache.struts.action.ActionResources', returnNull=true
2006. 3. 22 오후 2:29:00 org.apache.struts.util.PropertyMessageResources <init>
정보: Initializing, config='org.apache.webapp.admin.ApplicationResources', returnNull=true
2006. 3. 22 오후 2:29:02 org.apache.coyote.http11.Http11Protocol start
정보: Starting Coyote HTTP/1.1 on http-8080
2006. 3. 22 오후 2:29:02 org.apache.jk.common.ChannelSocket init
정보: JK2: ajp13 listening on /0.0.0.0:8009
2006. 3. 22 오후 2:29:02 org.apache.jk.server.JkMain start
정보: Jk running ID=0 time=10/80 config=C:\XMLS\W\Tomcat\bin\..\conf\jk2.properties
```

- Eclipse > [Navigator] > XMLJFrame.java >> 실행

9-14. XML 문서 뷰어 개발

- [찾아보기...] 클릭

<http://localhost:8080/ch9/bml.xml>



XML 문서 뷰어

BML 문서 뷰어

문서 경로 입력

책 정보

책제목

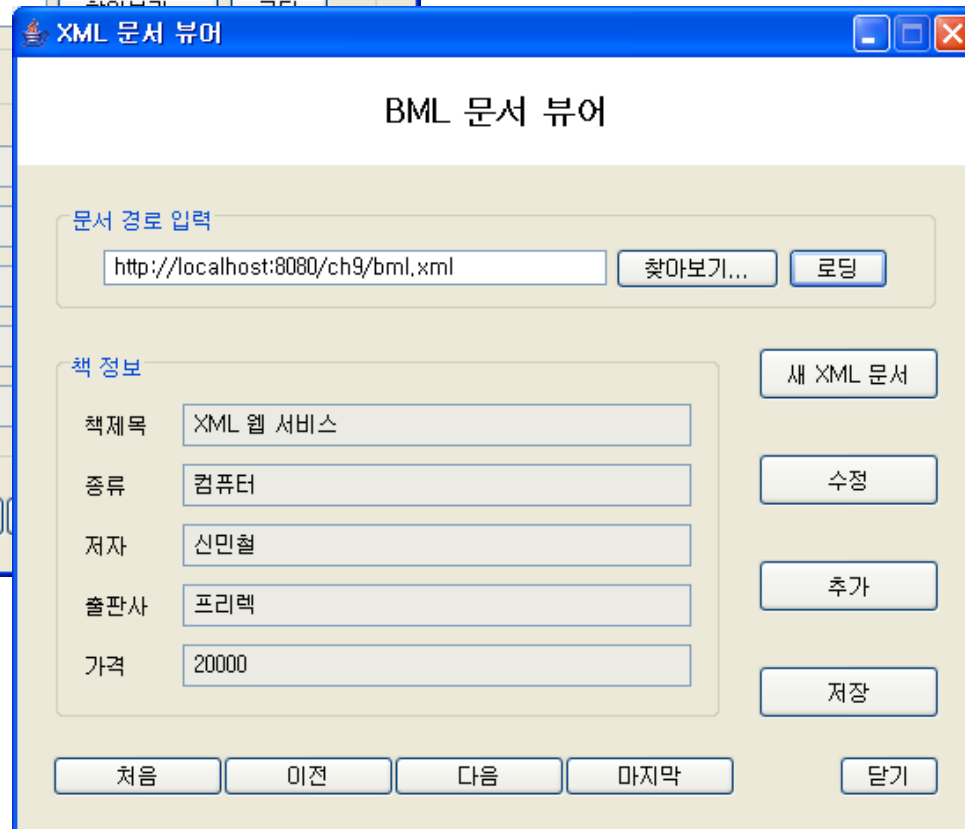
종류

저자

출판사

가격

처음 이전 다음



XML 문서 뷰어

BML 문서 뷰어

문서 경로 입력

책 정보

책제목

종류

저자

출판사

가격

처음 이전 다음 마지막 닫기