### dom4j学习报告

本报告为国科大计算机系17级学生胡康智于面向对象程序设计课程上的开源项目阅读学习报告，此报告分析的目标为dom4j——dom4j.org出品的一个开源XML解析包。

在本报告中，我会在exlipse中对dom4j的源代码进行阅读，并使用一些简单的xml文件来测试dom4j对于解析xml文件的效果，尝试从分析与建模、核心流程设计分析、高级设计意图分析等角度分析dom4j的功能与结构。

由于在下才疏学浅，对于dom4j的阅读与理解很可能会有较多的纰漏与误解，我只能尽量保证将在下课上所学皆用于此，不负老师与助教这一个学期的培养。

**一、导言**

1.1 dom4j.org简介

dom4j.org是一个致力于提供用于建设个人网页的教导与服务的免费网站，其域名为http://www.dom4j.org/

其主要功能包括提供一些有用的HTML资源，告诉你如何使用免费的网络资源来制作你的免费网站，以及关于PHP的基础知识。

dom4j为其制作的一个开源的XML解析包。

1.2 dom4j简介

dom4j是dom4j.org出品的一个开源XML解析包，其官网域名为https://dom4j.github.io/

dom4j是一个Java的XML API，是jdom的升级品，用来读写XML文件。它是一个十分优秀的JavaXML API，具有性能优异、功能强大和极其易使用的特点。

dom4j是一个易用的、开源的库，用于XML，XPath和XSLT。它应用于Java平台，采用了Java集合框架并完全支持DOM，SAX和JAXP。

1.3 dom4j主要功能

1.3.1 解析XML文件

解析XML文件是dom4最主要的功能，dom4j通过于org.dom4j.io包读写XML文件，其中提供DOMReader和SAXReader两类不同方式，这两种方式在解析一个XML文件时调用方法是相同的，故这里不做细讲。然后dom4j会遍历XML树，返回每一个节点的属性并将其输出。

1.3.2 使用迭代器

dom4j可以使用返回标准Java迭代器的各种方法来导航文档。这是一种一个容器对象中的各个元素进行访问，而又不暴露该对象容器的内部细节的方式，在各个容器内部结构不同的情况下，迭代器可以较为轻松地遍历容器中的内容，而忽略容器的一些具体细节。

1.3.3 使用Xpath导航

在dom4j中，可以对文档或树中的任何节点(如属性、元素或处理指令)计算XPath表达式。这允许使用一行代码在整个文档中进行复杂的导航。

1.3.4 快速循环

在遍历一个大型XML文档树时，为了提高性能，dom4j使用了快速循环方法，这种方法避免了为每个循环创建迭代器对象的成本。

1.3.5 创建新的XML文档

dom4j做的不仅是解析XML文件，同时它也有创建一个新的XML的功能

1.3.6 将文档写入文件

通过write()方法向写入器写入文档(或任何节点)

1.3.7 转换字符串

如果有对文档或任何其他节点(如属性或元素)的引用，则可以通过asXML()方法将其转换为默认的XML文本。

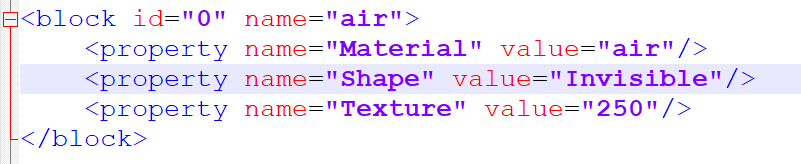
1.3.8 使用XSLT转换文档

这使其可以使用任何XSLT引擎，比如Xalan或Saxon。

1. **实例分析**

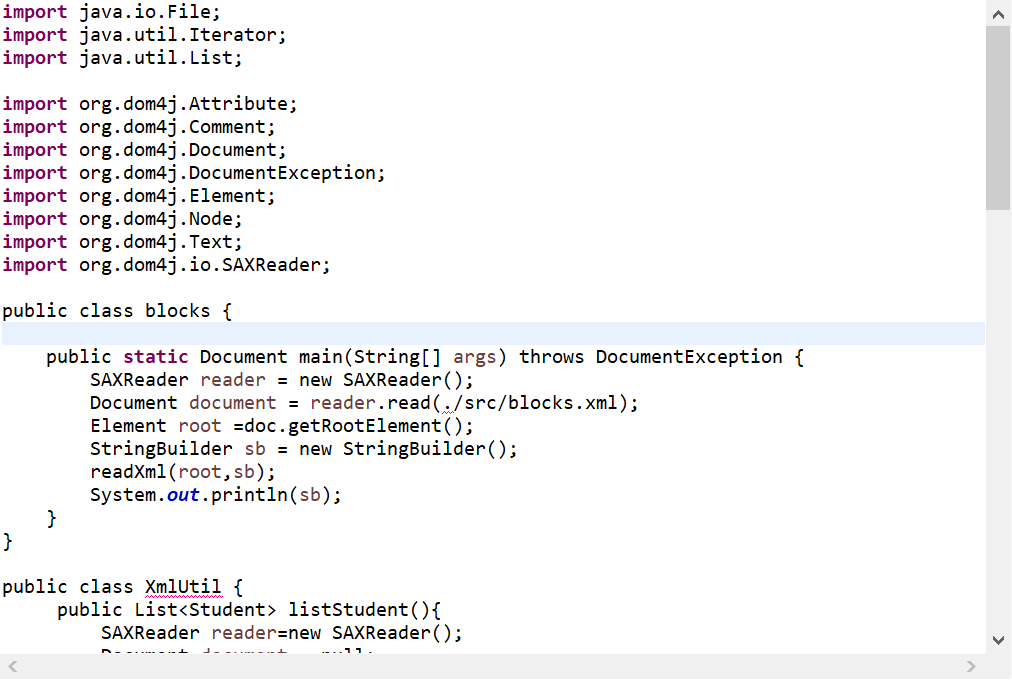
2.1 分析对象

本次使用XML文件节选自一款游戏，具体内容如下：[1]



2.2 分析方法

本次使用环境为eclipse-java-2019-12-R-win32-x86\_64，JDK为jdk-13.0.1\_windows-x64，使用dom4j版本为2.1.1，具体测试代码如下：



2.3 测试结果

测试结果如下：



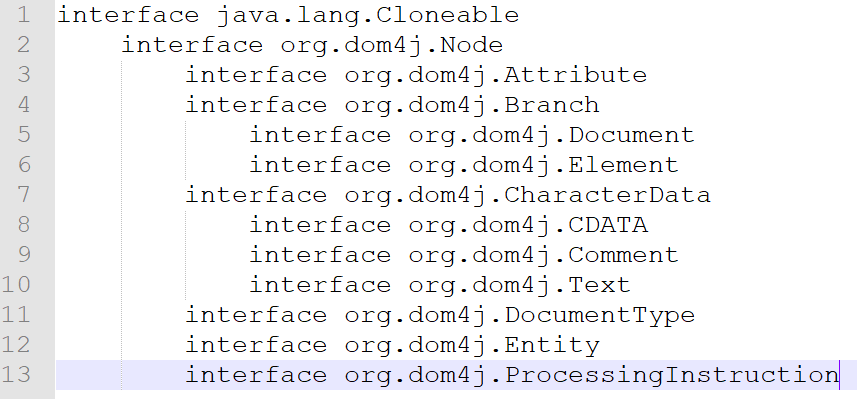
[1]该代码来自游戏“7\_Days\_To\_Die”

1. **Node接口设计**

3.1 接口继承关系

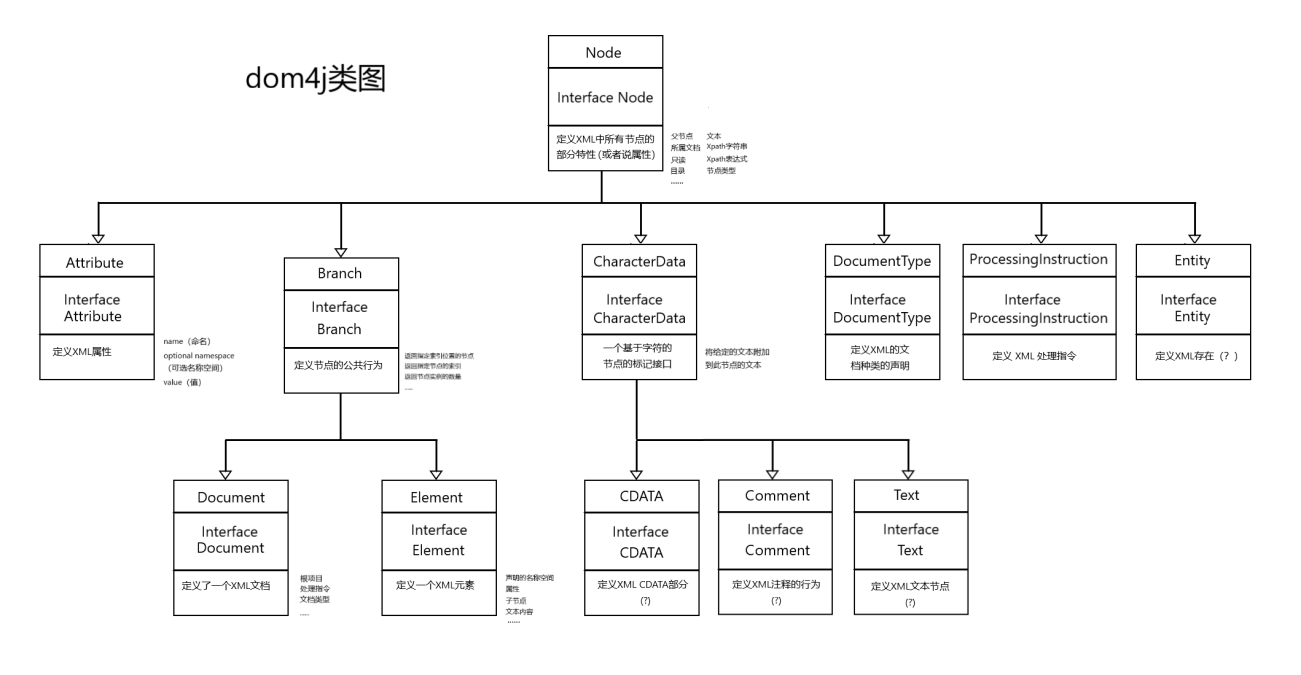
DOM4J 最大的特色是使用大量的接口。它的主要接口都在org.dom4j里面定义，其中大部分用于解析XML文件的接口继承自Node interface。

其主要接口继承关系如下：



3.2 UML类图

将类的继承关系转化为UML图：



3.3 接口具体功能

每个接口的具体作用如下表：

|  |  |
| --- | --- |
| 接口名称 | 功能 |
| Node | 定义XML中所有节点的部分特性（或者说属性） |
| Attribute | 定义XML属性 |
| Branch | 定义节点的公共行为 |
| Document | 定义一个XML文档 |
| Element | 定义一个XML元素 |
| Characher | 一个基于字符的节点的标记接口 |
| CDATA | 定义XML CDATA部分 |
| Comment | 定义XML注释 |
| Text | 定义XML文本节点 |
| DocumentType | 定义XML的文档种类的申明 |
| ProcessingInstruction | 定义XML处理指令 |
| Entity | 定义XML存在 |

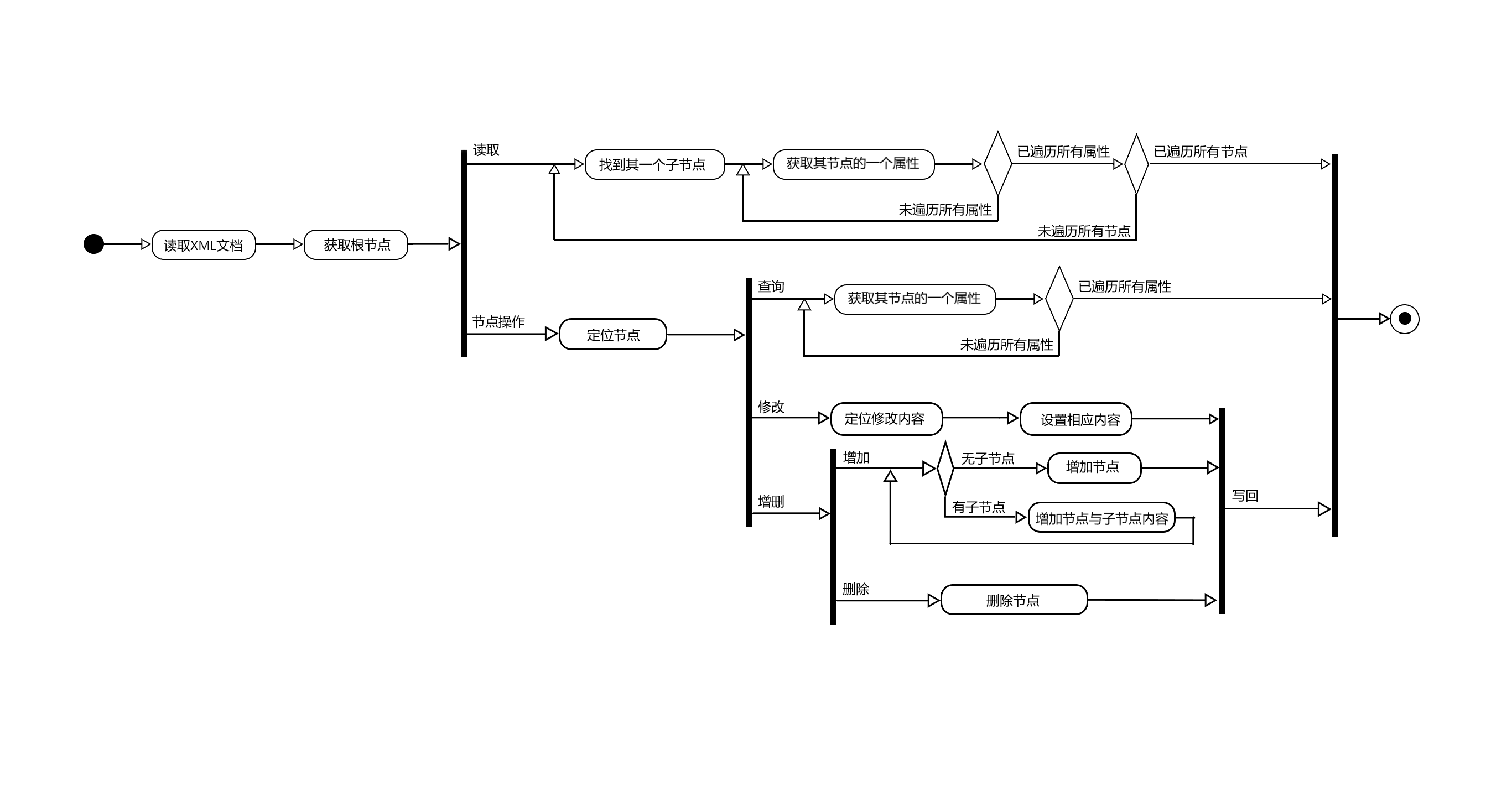
3.4 其他

虽然在dom4j的注释中写明了CDATA、Comment、Text和Entity节点的功能，但是其源代码并没有实现新的功能，简而言之，这四个接口是空接口，没有在父接口的基础上实现新的功能，仅为结构化与模块化的设计而定义。

1. **FactoryMethod工厂设计模式简析**

4.1 流程图

dom4j在处理XML文件时具体流程如下：（可能不够全面）



4.2 工厂设计模式

#当一个类不知道它所必须创建的对象的类的时候。

#当一个类希望由它的子类来指定它所创建的对象的时候。

#当类将创建对象的职责委托给多个帮助子类中的某一个，并且你希望将哪一个帮助子类是代理者这一信息局部化的时候。

#工厂方法不再将与特定应用有关的类绑定到你的代码中。

#代码仅处理Product接口；因此它可以与用户定义的任何ConcreteProduct类一起使用。

#用工厂方法在类中创建对象比直接创建对象更灵活。[2]

4.3 dom4j设计模式

dom4j解析XML文件时用到的接口所定义的并非相同的类，比如Node接口定义了节点的父节点、所属文档、是否只读、目录、文本、Xpath字符串和表达式、节点类型等属性，而Branch接口返回指定位置的节点或者返回指定节点的索引，Document接口定义了XML文档的根目录、处理指令和文档类型。

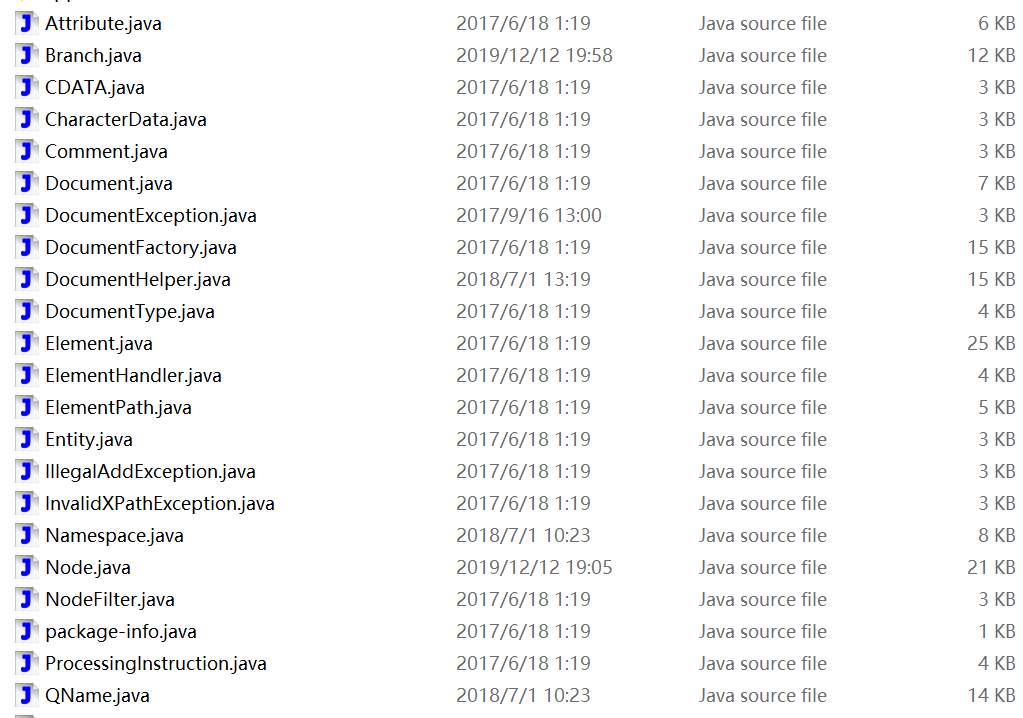
在处理这些接口对定义的要求时，单一的对象的创建显然是不满足要求的，故这里就需要工厂式的设计，将创建对象的任务委托给子类，在dom4j中就体现为每个Node的子接口都有自己的定义，当解析程序需要哪个接口的定义时利用这些接口提供的类即可。

[2]本段内容节选自面向对象程序设计-设计模式-研讨参考素材PPT

1. **面向对象设计原则**

5.1 模块化

dom4j对于XML文件的操作是非常多的，因而也准备了非常多的接口用于处理这些操作：



（图中所示为java源代码，dom4j封装为jar后所有接口均为class文件）

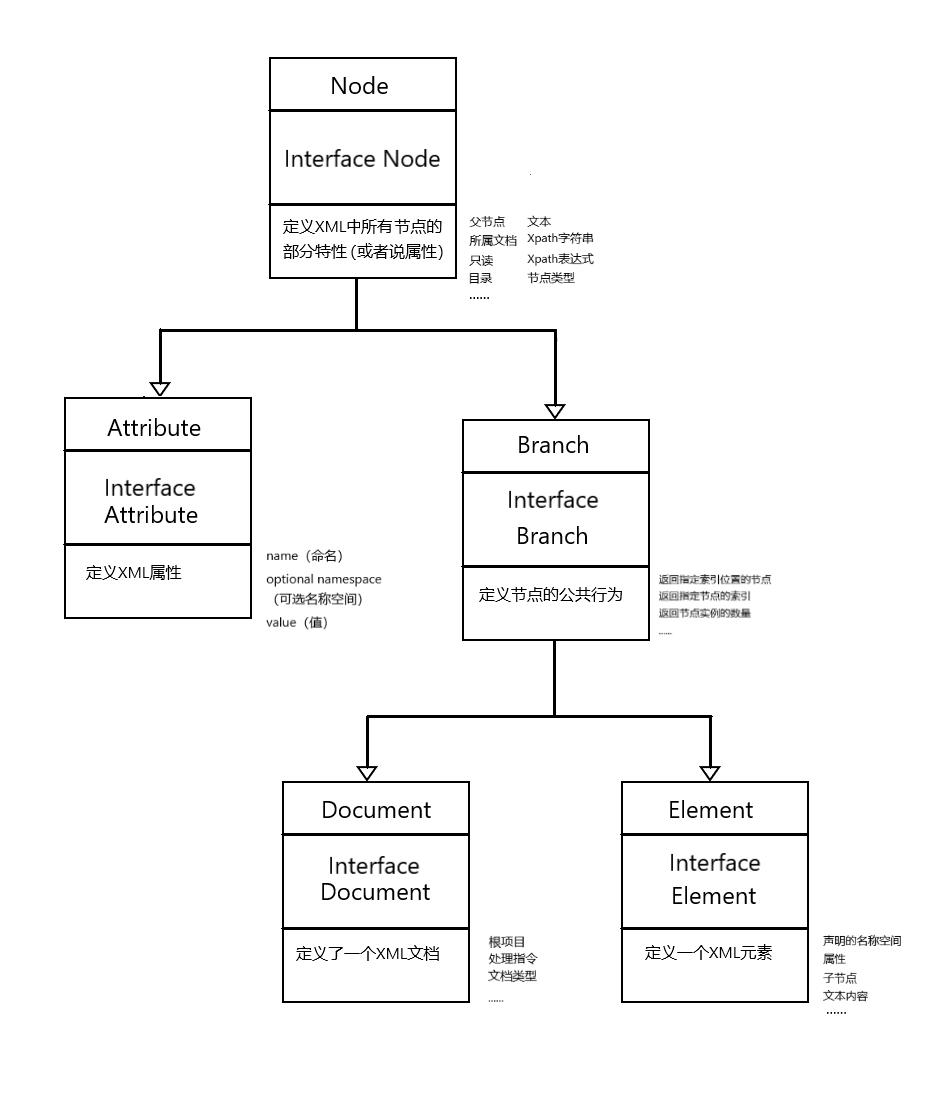
针对每种功能的接口都有清晰的命名与定位，在实现某种功能时通常不会用到过多的接口（例如：在调用Document和Element创建一个新的XML文件时，接口Entity没有必要也不会被调用）



每个结构都被非常好的封装起来。

5.2 层次结构

dom4j核心设计模块：



dom4j的一个很重要的特点就是大量的使用了接口Interface（或者说dom4j有不用接口实现的模块吗？），接口是一种极为抽象的类，只包含常量与抽象方法，并支持多重继承与定义多个类的共同行为。

在dom4j中，所有接口的层次与继承关系都非常清晰，这也使得dom4j的功能与结构分布比较便于理解，基本上只要明确了这些接口的继承关系，对于dom4j功能的理解也基本上是没有问题的。