Курсова работа по ООП – проект 13, XML парсер

Павел Григоров, СУ-ФМИ, КН, ФН 82038

Връзка към проекта - <https://github.com/PKGri/OOP_Course_Project_XML_Parser>

1. Въведение

XML парсерът е програма, която дава възможността да се работи с данни записани във файловия формат XML (e**X**tensible **M**arkup **L**anguage - разширяем маркиращ език). В действителност тези файлове представляват текстови файлове, в който данните са записани по определена конвенция. Основната функционалност е прочитането на файла и съответно представянето им в паметта на компютъра по начин, който позволява да бъдат извършвани множество действия, включващи извличане на конкретна информация, редактиране на някои от данните, записване на данните отново в XML файлов формат, както и осъществяването на някои основни XPath заявки.

XPath заявките са изрази, които указват на програмата каква информация да извлече и да представи на потребителя.

Първата основна задача е прочитането на един XML файл. За тази цел трябва да се измисли как трябва да бъде представяна в паметта текстовата информация от файла, за да могат след това да бъдат извършват операции с нея.

Следващата задача е да се определи какви точно операции да бъдат поддържани и по какъв начин да се подходи с тяхното осъществяване в кода. От тези операции, може би най-сложната е извличането на конкретна информация чрез XPath заявки, затова като последна стъпка в разработката оставям точно тази функционалност.

Междинни цели са осъществяването на работа с команден ред, за улеснена работа с всичките заредени данни, както и придаването на атрибут с ключ “*id”* и с уникална стойност съответстваща на всеки елемент, за да могат някои операции да бъдат извършвани върху точно определен елемент.

По-долу тази документация описвам как съм подходил в достигането си на тези цели, някои от проблемите, които са се явили в процеса на разработката и как съм се справил с тях. Ще бъдат описани всичките класове и повечето от по-сложните функции, които са част от кода. Разбира се, на края на този документ се намират хипервръзки към отделните източници, които ми бяха от полза в изготвянето на проекта.

1. Преглед на предметната област

Тук ще представя и дам дефиниции на специфичните термини, които ще бъдат срещани надолу в документацията.

**Общи:**

* **Регулярен израз –** последователност от знаци, която дефинира шаблон за търсене. В кода е използвана стандартната библиотека <regex>.

**XML специфични:**

* ***“Таг“* –** всяко нещо във файла, което се съдържа между скоби <…>
* **Отварящ таг –** таг, чрез който се маркира началото на нов елемент - <…>
* **Затварящ таг –** таг, чрез който се маркира краят на елемент - </…>
* **Самозатварящ се таг** или  **празен таг–** таг, който отговаря на празен елемент - <…/>
* **Таг на версията –** таг в началото на файла, който може да бъде пропускан; Този таг указва каква версия на XML е използвана за файла.
* **Ключ –** името на елемент или атрибут. Намира се в началото на отварящ таг.
* **Елемент –** елементът включва цялата информация от начален таг до съответния затварящ таг
* **Коренов елемент** или **корен –** първоначалният елемент, отворен чрез първия таг след таг на версията, ако има такъв. Файлът може да съдържа само един коренов елемент.
* **Елемент-дете** или **дете –** елемент, който е вложен в друг елемент
* **Родителски елемент** или **родител –** елемент, който има поне едно дете
* **Текстов елемент –** елемент, който няма деца, но между отварящия и затварящия си таг има текстово съдържание.
* **Празен елемент –** елемент, който няма нито деца, нито текстово съдържание
* **Атрибут на елемент –** атрибутът съдържа информация относно специфичен елемент. Атрибутите са представени в отварящия таг на елемента им, след неговия ключ. Всеки елемент може да притежава множество от атрибути.
* **Стойност на атрибут –** информацията, която даденият атрибут пази за съответния си елемент.
* **Идентификатор** или “***id”,***или “***ID”* –** атрибут, който по условие се придава на всеки елемент и има уникална стойност.

**XPath специфични:**

* **Път –** указаният чрез *XPath-*израз път до елемент, наподобяващ път на директория в компютъра
* **Възел –** всяка част от *XPath-*израза, която отговаря на елемент или атрибут
* **Предикат –** условие, което даден възел-елемент трябва да изпълнява, за да се продължи към края на израза. Намира се в скоби […] след ключа на елемента.