

Документация на проект

Проект: XML Parser

Студент: Александър Иванов, КН група 4, 81841

Съдържание

Документация на проект

Проект: XML Parser

Студент: Александър Иванов, КН група 4, 81841

Съдържание

Кратко обяснение на имплементацията

Инструкции за употреба на приложението

Граматика за XML документ

Граматика за XPath заявка

Значение на xpath заявките

Кратко обяснение на имплементацията

"Разбирането" на текста от документ(или заявка) се състои от няколко етапа на 'tokenization' и 'parsing'.

В първия етап текстът бива разбит на "тоукени" - идентификатори, специфични символи, литерали и т.н. Това улеснява работата ни в следващия етап. Тази стъпка е (донякъде) обща за xml текст и xpath заявка. Структурата от данни, представяща тоукените е **class token**.

Следва междинен етап за привеждане на тоукените в нормална форма, отделен за xml документите и xpath заявките. Тук се пропускат интервали, нови редове и т.н., а текстовото съдържание на XML елементите се съединява.

Следващият етап е отделен, но със сходна имплементация, за документите и заявките. Тук редицата от тоукени се "разбира" - построява се адекватна структура от данни, нееднозначно представяща документ или заявка. Това са обекти на класовете **xml_node**, **xpath_query**, както и някои други помощни.

След като сме се сдобили с инстанции на **xml_document** и **xpath_query**, можем да изпълним заявката, използвайки класа **xml_context**.

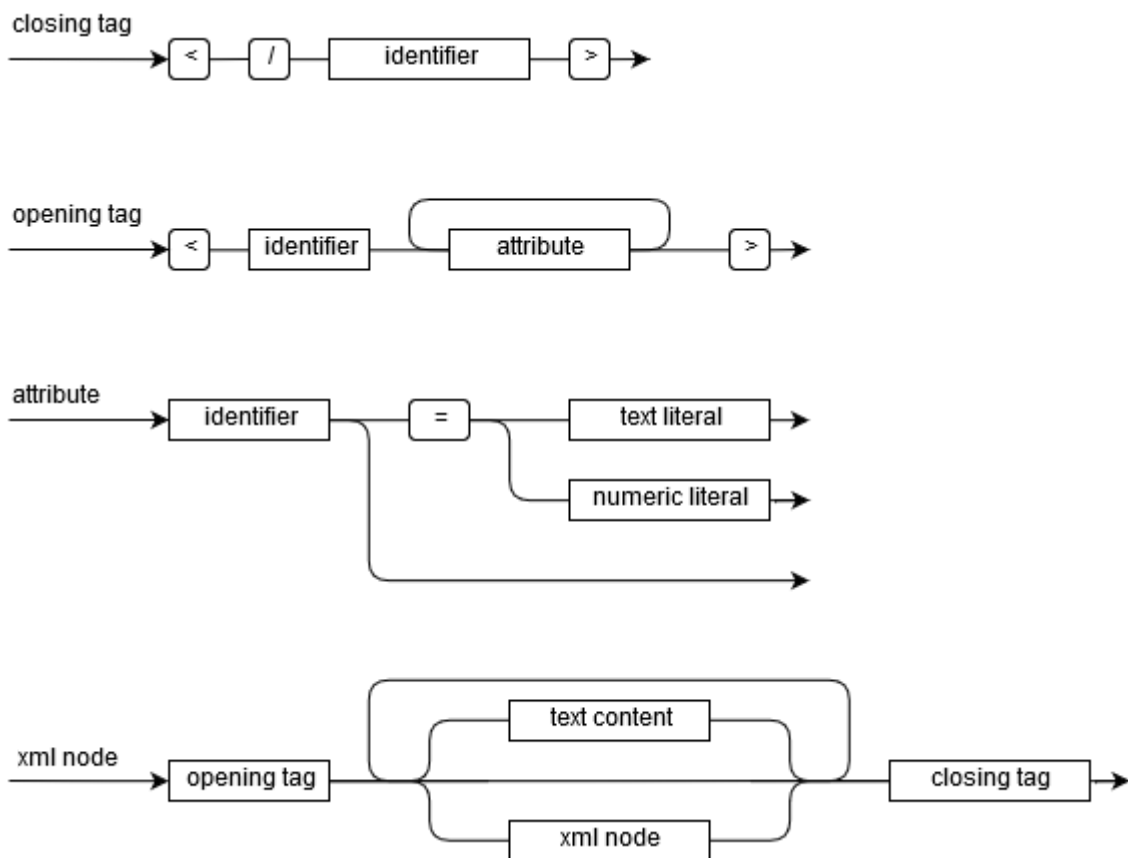
Инструкции за употреба на приложението

Приложението поддържа следните команди

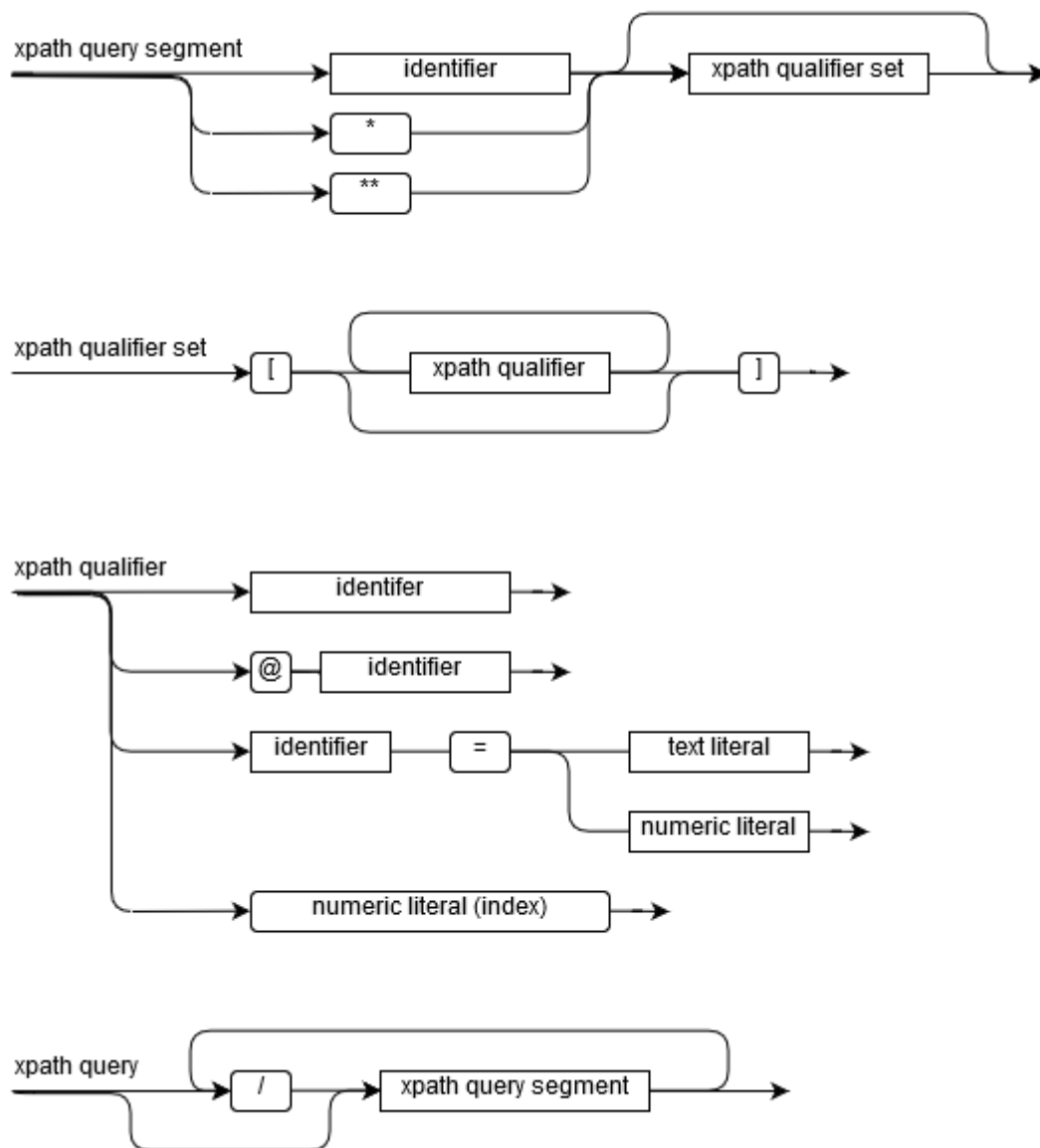
- exit (или quit) - прекратява изпълнението на приложението
- open(или load) [път към XML документ]
 - Зарежда документ от файл и изплюва структурата и съдържанието му на конзолата
- query(или select) [XPath заявка]
 - Изпълнява заявка върху заредения документ и изплюва съдържанието, подобно при open, както и селектираните елементи. Ако има извлечени стойности на атрибуте, те се извеждат накрая.

Опит за зареждането на невалиден документ или заявка най-вероятно ще доведе до принудителен край на процеса :/

Граматика за XML документ



Граматика за XPath заявка



Значение на xpath заявките

- Заявките без префикс '/' се изпълняват относно текущо избраните елементи на документа
- Заявките, започващи с '/' се изпълняват относно елементите от най-горно ниво
- '*' избира всякакъв вид елемент
- '**' избира всякакъв вид елемент или негов пряк или непряк наследник
- Филтри
 - [(число)] филтрира избора до елемент, който е на съответния индекс в списъка с преки наследници на родителя си
 - [@id] извлича стойността на id атрибута на съответния избран елемент
 - [id] ограничава избора до елемент с наличен id атрибут
 - [id (< или > или =) (string literal или numeric literal)]

- Ограничава избора до елемент, с наличен **id** атрибут, имащ стойност, която изпълнява условието
 - Сравнението на стойността се изпълнява числово или лексикографски, в зависимост от вида литерал в заявката
 - Всякакви комбинации от горните филтри, разделени с интервал в квадратните скоби е позволена (както може да се види в горната диаграма). При повтаряне на филтър за индекс важи последния.
-