On Using XML Pull Parsing Java APIs

Prepared for First International Workshop on High Performance XML Processing

15 March 2004

Aleksander Slomiski
Department of Computer Science
Indiana University
www.extreme.indiana.edu

Abstract

This paper provides an overview of XML Pull Parsing Java APIs. We first briefly describe how pull parsing is different from other approaches. Then we introduce existing APIs: kXML1, NekoPull, XNI2XmlPull, XPP1, XPP2, kXML2, XPP3, and JSR-173 StAX and compare them. Finally we look on performance implications and future directions.

Introduction

TODO: high performance motivsation – easy of use vs. memory utilization/performance trade-off

We will use one simple example to compare APIs. This is a simple stream of records – address book containing list of persons, each represented as XML of following structure (name is required, home_address and work_address are optional):

```
<person>
<name>Joe Doe</name>
<home_address>
<street>101 Sweet Home</street>
<phone>333-3333</phone>
</home_address>
<work_address>
<street>303 Office Street</street>
<phone>444-4444</phone>
</work_address>
</person>
```

In this example we want to read XML in streaming manner and extract information into Java specific data types, for example:

```
class Person {
        String name;
        Address homeAddress;
        Address workAddress;
```