



Eine neue Notation zur Darstellung von XML-Texten samt Entwicklung eines Editors für sie

Diplomaufgabe für Christian Danner
Matrikelnummer: 0055085
Email: christian.danner@students.jku.at

Die eXtended Markup Language (XML) ist eine Notation zur Beschreibung hierarchisch strukturierter Daten. Sie setzt sich heute in immer mehr Bereichen der Informatik durch, z.B. im Bereich der Web-Services, Plugin-Spezifikationen, Konfigurationsdateien, als Datenaustauschformat oder allgemein zur Serialisierung von Daten.

XML-Texte lassen sich zwar maschinell leicht verarbeiten, sind aber für Menschen schwer lesbar, da sie einigen notationellen Overhead enthalten (z.B. Endmarken oder spitze Klammern). Alternative Notationen wie zum Beispiel *json* (<http://www.json.org>) sind zwar leichter lesbar, haben sich aber noch nicht durchgesetzt, weil viele Programme XML-Daten erwarten.

Ziel dieser Diplomarbeit ist die Entwicklung einer zu XML alternativen Notation, in der man XML-Daten übersichtlicher als bisher lesen und editieren kann. Die beiden Notationen sollen auf einfache Weise ineinander konvertierbar sein.

Ein XML-Text ist nichts anderes als eine hierarchisch strukturierte Menge von Name/Wert-Paaren. Sie bildet einen Baum, der im einfachsten Fall durch einen Text mit Einrückungen visualisiert werden kann. Der XML-Text

```
<menu id="file" value="File" >
  <popup>
    <menuitem value="New" onclick="CreateNewDoc()" />
    <menuitem value="Open" onclick="OpenDoc()" />
    <menuitem value="Close" onclick="CloseDoc()" />
  </popup>
</menu>
```

kann zum Beispiel folgendermaßen dargestellt werden:

```
menu:
  @id:    file
  @value: File
  popup:
    menuitem:
      @value:    New
      @onclick:  CreateNewDoc()
    menuitem:
      @value:    Open
      @onclick:  OpenDoc()
    menuitem:
      @value:    Close
      @onclick:  CloseDoc()
```

Ausgehend von diesem Beispiel als Anregung ist eine möglichst übersichtliche und kompakte Notation zu entwickeln, die alle Sprachelemente von XML umfaßt, also auch Namensräume, XML-Kommentare und XML-Direktiven, so daß eine jederzeitige Umsetzung in XML möglich wird. Die Vollständigkeit und Brauchbarkeit der neuen Notation ist an Hand möglichst vieler realer XML-Texte zu erproben.

Es ist ferner ein Editor zu entwickeln, mit dem man XML-Dateien in der neuen Notation anzeigen, editieren und wieder als XML-Dateien abspeichern kann. Der Editor soll außerdem folgende Eigenschaften aufweisen:

- Validierung des Textes gegen ein XML-Schema (möglichst bereits während des Editierens)
- Folding von Elementen
- Syntax-Coloring und Auto-Indentation

Der Editor soll als eigenständiges Programm benutzbar sein. Wenn Zeit bleibt, soll aber auch versucht werden, ihn als Plugin in Visual Studio.NET zu integrieren.

Die Implementierung soll in C# unter .NET erfolgen, wobei größter Wert auf Robustheit und einfache Bedienbarkeit gelegt werden soll. Achten Sie auf guten Programmierstil und ausführliche Kommentierung, damit der Editor später auch von anderen Personen gewartet werden kann. Ziel ist es, den Editor als Open-Source-Produkt zur Verfügung zu stellen.

Die Arbeit ist in regelmäßigen Abständen mit dem Betreuer zu besprechen.

Betreuer: o.Univ.-Prof. Dr. Hanspeter Mössenböck

Beginn: Oktober 2005