

INTERACTIVE VISUALIZATION OF XML DOCUMENTS

Radim Hernych, Daniel Kubicek

Bachelor Degree Programme (3), FIT VUT

E-mail: xherny00@stud.fit.vutbr.cz, xkubic17@stud.fit.vutbr.cz

Supervised by: Petr Chmelař

E-mail: chmelarp@fit.vutbr.cz

ABSTRACT

There is many data stored and exchanged in XML nowadays. However, understanding of the information included in large data-oriented XML instances may not be simple in contrast to the presentation of document-oriented XML instances. Moreover, acquaint the desired knowledge from complex XML data is hard or even impossible.

1. ÚVOD

V dnešní internetové době mají informace obrovský význam. Velmi důležité je proto umět tyto informace rychle a efektivně zpracovat a také umět je vyměnit s okolím. Dříve se pro výměnu informací používaly proprietární formáty, které ale měly řadu nevýhod. To se však s příchodem technologie XML změnilo. Výměna informací v tomto otevřeném textovém formátu přináší celou řadu výhod. XML umožňuje výměnu dokumentů mezi různými platformami, různými operačními systémy a aplikacemi. Také nabízí jednoduché zpracování obsahu a efektivní způsoby vyhledávání v dokumentech. XML však definuje pouze strukturu dokumentu, ale neříká nic o jeho vzhledu a zobrazení.

2. XML

Jazyk XML(eXtensible Markup Language) je jednoduchý, velmi flexibilní a otevřený textový formát odvozený od SGML (ISO 8879).

XML Dokumenty mohou být rozděleny na textově orientované a datově orientované. U textově orientovaných dokumentů převažuje text. Většina značek zde slouží pouze k přidání určité informace k textu a značky nejsou nezbytně nutné pro pochopení podstaty dokumentu. U datově orientovaných dokumentů je tomu jinak. Ty jsou určeny převážně pro počítačové zpracování a nehodí se k přímému čtení uživatelem. Značky zde vytvářejí strukturu dokumentu a přidávají datům určitý interpretační smysl. Kdyby byly z takového dokumentu značky vyjmuty, zbývající data by nedávala žádný smysl. Článek je zaměřen na datově orientované XML dokumenty, které jsou v textové podobě špatně čitelné, a proto vyžadují vhodné techniky pro jejich vizualizaci.

3. ŘEŠENÍ

3.1. PROGRAM VISUALXML

Prvním řešením je program VisualXML, který byl navržen tak, aby dokázal zobrazit obecný XML dokument, a také aby umožnil interaktivní editaci jeho obsahu i XML neznalému uživateli.

Program zobrazuje elementy jako objekty reprezentované svým obsahem a propojením objektů modeluje vztahy mezi jednotlivými elementy. Umožňuje intuitivně editovat XML dokumenty – smazat či vložit nový element, editovat textový obsah elementu a naklonovat element (i s jeho podstromem). Umožňuje také vkládat, editovat a mazat atributy jednotlivých elementů. Kolekci elementů modeluje program jako seznam a umožňuje jej stránkovat a také interaktivně filtrovat v něm zobrazené elementy (viz Obrázek 1).

```
<weather ver="2.0">
  <head>
    ...
  </head>
  <loc id="EZXX0002">
    <dnam>Brno, Czech Republic</dnam>
    <tm>6:47 PM</tm>
    <lat>49.20</lat>
    <lon>16.61</lon>
    <sunr>7:25 AM</sunr>
    <suns>5:51 PM</suns>
    <zone>2</zone>
  </loc>
  <cc>
    ...
  </cc>
  <dayf>
    <lsup>10/22/07 3:21 PM Local Time</lsup>
    <day d="0" dt="Oct 22" t="Monday">
      <hi>N/A</hi>
      <low>3</low>
      <sunr>7:25 AM</sunr>
      <suns>5:51 PM</suns>
    </day>
    <day d="1" dt="Oct 23" t="Tuesday">
      ...
    </day>
    <day d="2" dt="Oct 24" t="Wednesday">
      ...
    </day>
    <day d="3" dt="Oct 25" t="Thursday">
      ...
    </day>
  </dayf>
</weather>
```

Obrázek 1: Zobrazení ukázkového XML dokumentu programem VisualXML

3.2. XMLAD

Řešení, implementováno pod názvem XMLAD, spočívá ve zjednodušení a zpřehlednění stromové struktury XML dokumentů a neomezování se jen na určitý formát XML dat.

XMLAD přináší dva typy pohledů na jednotlivé XML elementy. Základní typ představuje jeden konkrétní XML element. Grafické znázornění základního typu obsahuje dvě tabulky, tabulku atributu a tabulku sub-elementů. AgregáčnÍ typ pohledu přináší zpřehlednění zobrazování více XML elementů stejného jména a entity. Tyto XML elementy jsou

sjednoceny do jednoho logického celku, který je přehledně graficky znázorněn pomocí dvou tabulek. V první tabulce každý řádek zobrazuje hodnoty jednoho XML elementu a jeho atributů. Druhá tabulka obsahuje seznam metadat příslušející tomuto logickému celku – jedná se o všechny sub-elementy. Hodnoty metadat lze interaktivně zobrazovat v tabulce XML elementů. Jednotlivé graficky zobrazené XML elementy lze interaktivně skrývat, přesouvat a zobrazovat.

The image displays the XMLAD interface. On the left is an XML tree view for a weather document. On the right, several tables provide a visual representation of the data:

- weather** table:

Atribut	Hodnota
ver	2.0
- loc** table:

Atribut	Hodnota
id	EZX0002
- dayf** table:

Subelement	Hodnota
isup	10/22/07 3:21 PM
- day (10)** table:

#...	d	dt	t	/low	/hi	...
day 0	0	Oct 22	Monday	3	N/A	...
day 1	1	Oct 23	Tues...	6	6	...
day 2	2	Oct 24	Wed...	6	12	...
day 3	3	Oct 25	Thur...	6	11	...
day 4	4	Oct 26	Friday	6	12	...
day 5	5	Oct 27	Satur...	4	14	...
day 6	6	Oct 28	Sunday	4	14	...
day 7	7	Oct 29	Monday	3	13	...
day 8	8	Oct 30	Tues...	3	12	...
day 9	9	Oct 31	Wed...	4	10	...
- Metadata** table:

Subelement	Hodnota
hi	
low	
sunr	
suns	
part	

Obrázek 2: Zobrazení ukázkového XML dokumentu programem XMLAD

4. ZÁVĚR

Existují i jiná řešení pro zobrazování XML dokumentů jako je XSLT, nicméně nenabízejí uspokojivé řešení. Největším nedostatkem bývá svázanost s určitým formátem XML a jeho definice, nebo nepřehlednost u větších souborů dat. VisualXML a XMLAD přináší nové možnosti řešení, jak tyto problémy odstranit.

LITERATURA

- [1] Meteorologists at The Weather Channel®. 2007. <weather.com>.
- [2] Jelinek, J., Slavik, P.: XML Visualization Using Tree Rewriting. Procs. of 20th Spring Conf. on Computer Graphics SCCG. p. 65 – 72. ACM Press, 2004.
- [3] Chmelar, P., Hernych, R., Kubicek, D.: Interactive Visualization of Data-Oriented XML Documents, In: Advances and Innovations in Systems, Computing Sciences and Software Engineering, Springer, 2008, p. 4.