
Schémata a transformace pro srovnávače zboží

Semestrální práce z předmětu 4IZ238

Martin Hujer <mhujer@gmail.com>

Obsah

1. Úvod	1
1.1. Stávající situace	1
2. Cíle práce	1
2.1. Zjednodušit generování feedů pro srovnávače	1
2.2. Snížit množství chyb při vytváření feedů	2
2.3. Zlepšit User-Experience při práci s feedy	2
3. Výstupy práce	2
3.1. Univerzální XML feed	2
3.1.1. Tiskový výstup	2
3.1.2. Transformace pro jednotlivé srovnávače	2
3.2. Schémata pro jednotlivé formáty	3
3.2.1. PHP	3
3.3. XSL transformace pro zobrazení v prohlížeči	3
4. Závěr	3

1. Úvod

Cílem práce je zlepšit situaci kolem XML pro srovnávače zboží a zjednodušit jejich generování.

1.1. Stávající situace

1. Každý zbožíový srovnávač má svůj formát feedu
 - Specifikace Zboží.cz - <http://napoveda.seznam.cz/cz/specifikace-xml.html>
 - Specifikace Heureka.cz <http://sluzby.heureka.cz/napoveda/import/>
 - Jyxo srovnávač <http://admin.vybereme.cz/format-xml-feedu>
2. Jejich XMLka nemají definované schéma – nedá se tedy poznat, zda export z eshopu je OK
3. Je spousta práce s generováním validních feedů pro různé srovnávače

2. Cíle práce

2.1. Zjednodušit generování feedů pro srovnávače

Vytvořením univerzálního XML formátu pro exporty zboží pro srovnávače

Vytvořením transformací, které z jednoho XML vytvoří XML pro jednotlivé srovnávače

2.2. Snížit množství chyb při vytváření feedů

Vytvořením schémat pro jednotlivé srovnávače (podle jejich textového popisu XML feedu)

Jednotlivých srovnávačům navrhnou zahrnutí schématu do popisu služby

2.3. Zlepšit User-Experience při práci s feedy

Vytvořením xsl transformací, které mohou být připojeny ke stávajícím XML exportům pro jednotlivé srovnávače – při zobrazení v browseru bude feed přehlednější

3. Výstupy práce

3.1. Univerzální XML feed

Vytvořil jsem formát pro uložení informací o produktech, který obsahuje nadmnožinu informací o produktech v jednotlivých feedech. Ukázkový soubor s daty je `universal.xml`. Jeho schéma je uloženo v souborech `universal.xsd` a `generic.xsd`.

3.1.1. Tiskový výstup

Z univerzálního XML je možné pomocí transformace v souboru `universal-fo.xsl` vygenerovat PDF s tiskovým výstupem.

3.1.2. Transformace pro jednotlivé srovnávače

Vytvořil jsem XSL transformace v souborech `heureka.xsl`, `jyxo.xsl` a `zbozi.xsl`, které převedou data z univerzálního feedu `fo` formátu pro příslušné srovnávače.

3.1.2.1. PHP

Transformace je možné provést přímo z PHP:

```
<?php
generate(' ../heureka.xsl', 'heureka-out.xml');
generate(' ../jyxo.xsl', 'jyxo-out.xml');
generate(' ../zbozi.xsl', 'zbozi-out.xml');

function generate($inputTransformation, $outputFile)
{
    $xml = new DOMDocument;
    $xml->load(' ../universal.xml');

    $xsl = new DOMDocument;
    $xsl->load($inputTransformation);

    $proc = new XSLTProcessor();
    $proc->importStylesheet($xsl);
    $newXml = $proc->transformToXML($xml);
    file_put_contents($outputFile, $newXml);
}
```

Kód je k dispozici v souboru `php/transform.php`.

3.2. Schémata pro jednotlivé formáty

Vytvořil jsem schémata `heureka.xsd`, `jyxo.xsd` a `zbozi.xsd`, která jsou převedením textového popisu formátu jednotlivých služeb do XML schématu, které lze využít pro automatickou validaci vygenerovaného XML.

3.2.1. PHP

Validace proti schématům je možné spustit přímo z PHP:

```
<?php
chdir('../');

validate('heureka.xml', 'heureka.xsd');
validate('heureka-gen.xml', 'heureka.xsd');
validate('jyxo.xml', 'jyxo.xsd');
validate('jyxo-gen.xml', 'jyxo.xsd');
validate('zbozi.xml', 'zbozi.xsd');
validate('zbozi-gen.xml', 'zbozi.xsd');

function validate ($xml, $schema)
{
    $doc = new DomDocument();
    $doc->Load($xml);
    if (@$doc->schemaValidate($schema)) {
        printf('%s is valid against %s' . "\n", $xml, $schema);
    } else {
        printf('!!! %s is not valid against %s' . "\n", $xml, $schema);
    }
}
```

Kód je k dispozici v souboru `php/validate.php`.

3.3. XSL transformace pro zobrazení v prohlížeči

Vytvořil jsem transformace `heureka-web.xsl`, `jyxo-web.xsl` a `zbozi-web.xsl`, které po připojení k XML feedu (i stávajícímu) převedou při zobrazení v prohlížeči nepřehledné XML na přehledný a formátovaný HTML výstup.

Transformace lze k souboru připojit použitím buď:

```
<?xml-stylesheet type="text/xsl" href="heureka-web.xsl"?>
```

nebo

```
<?xml-stylesheet type="text/xsl" href="jyxo-web.xsl"?>
```

nebo

```
<?xml-stylesheet type="text/xsl" href="zbozi-web.xsl"?>
```

4. Závěr

Pokud máte k projektu jakékoliv dotazy nebo připomínky, rád je uvítám na adrese `<mhuje@gmail.com>`. Ještě raději je ovšem uvidím jako Pull Requesty na Githubu [<https://github.com/mhuje/xml-feeds-schemas>].