

Služba pro televizní program

Vytvořeno pro KolejNet, kolejní studentskou síť Vysokého Učení Technického v Brně.

Cíl projektu

Cílem tohoto projektu bylo vytvořit centralizovanou službu s vysokou mírou spolehlivosti za účelem poskytování televizního programu. Aby byla služba univerzálně použitelná, formát dat jak pro ukládání dočasných souborů, tak pro výstup je XML.

Proč XML?

Použitím XML taky odpadá několik problémů, mezi hlavní přínosy patří:

- nezávislost na systému - XML dokument lze načíst jenom jediným způsobem a většina moderních programovacích prostředí má možnost pracovat s XML dokumenty
- internacionalizace - XML dokument si sice nese informaci o použité kódové stránce ale podpora pro čtení dané stránky je již zakomponována do funkce která nám XML soubor čte, takže uživatel není zatěžován problémem nesprávné znakové sady¹
- rozšiřitelnost - v případě že se povede identifikovat další doplňující údaje příslušející některým pořadům, je možnost je ihned zakomponovat do souboru nebo výsledků hledání bez porušení funkčnosti stávajících systémů

Struktura dat

Struktura poskytovaných dat je dvojího druhu:

- seznam televizních kanálů
- televizní program

Seznam televizních kanálů

Seznam kanálů je jednoduchá, jednoúrovňová struktura (jednoduché pole), která sestává z objektů s následujícími vlastnostmi:

- identifikátor
- krátký název
- dlouhý název
- jazyk

Příklad XML dokumentu s TV kanály:

```
<?xml version="1.0" encoding="Windows-1250"?>
<channels xmlns="http://schemas.rozsnyo.com/kolejnet/tv/channels">
  <channel id="CT1"
    shortname="ČT 1"
    longname="První program České Televize"
    lang="cz" />
  <channel id="CT2"
    shortname="ČT 2"
    longname="Druhý program České Televize"
    lang="cz" />
</channels>
```

XML namespace tohoto dokumentu je <http://schemas.rozsnyo.com/kolejnet/tv/channels> a pod touto URL také nalezneme formálnější specifikaci formátu dat.

Soubor s aktuálními televizními kanály (kterých program se také sleduje) se nachází na adrese: <http://mess.kn.vutbr.cz/tv/program/channels.xml>.

¹ Na místech kde jsou data v jiném formátu používáme kódování UTF-8 které není omezené na 256 různých znaků jako ISO-8859-2 nebo WINDOWS-1250

Televizní program

Struktura programu také není příliš složitá - jedná se jenom o dvouúrovňovou strukturu reprezentující příslušné pole seznamů (nebo také dvourozměrné pole). Dvě složky které jednoznačně identifikují položku programu jsou televizní kanál (jeho identifikátor) a čas.

Čas pořadu se ukládá jenom ten, ve který daný pořad začíná, takže u poslední položky nelze určit kdy skončí (algoritmicky, heuristicky by to šlo podle textu „závěr vysílání“ a pod.). Tento čas je reprezentován v Unix-ovém formátu, který je nezávislý na časové zóně a letním/zimním čase.

Položka programu obsahuje povinně jenom 2 informace:

- čas (time)
- název (label)

Volitelné informace, které nemusí být uvedeny u všech pořadů jsou pak:

- původní název (original)
- detailní popis (detail)
- příznak - premiéra, repríza, přímý přenos (flag)
- počet dílů seriálu (count)
- aktuální díl (current)

Všechny položky jsou textové, kromě příznaku, který je typu výčet (anglická slova premiere, reprise, live). Protože jsou tyto informace textové, může se objevit u aktuálního dílu seriálu třeba hodnota 13-14, která se už nedělí.

Pozor je třeba dát taky na to, že informace jsou na sobě nezávislé, tj. pokud existuje aktuální díl ještě neznamená že určité je definován počet dílů (a stejně to nejde usuzovat opačně).

Fyzicky vypadá televizní program následovně:

```
<?xml version="1.0" encoding="UTF-8" ?>
<program xmlns="http://schemas.rozsnyo.com/kolejnet/tv/program">
  <channel id="CT1">
    <item time="1108944600" label="Otázky Václava Moravce"/>
    <item time="1108948200" label="Sportovci světa: Slavné sportovní duely"
    detail="All Blacks versus Springboks Legendární soupeření ragbistických velmocí
    Nového Zélandu a Jihoafrické republiky. Francouzský dokumentární cyklus"/>
    <item time="1108951500" label="Missis 2005"/>
  </channel>
  <channel id="CT2">
    <item time="1108941300" label="Krvavé léto" original="Summer of Sam"
    detail="New York v obležení nesnesitelného vedra a šíleného sériového vraha.
    Americký film (1999). Hrají: J. Leguizamo, M. Sorvinoová, J. Espositoová, A.
    Brody, M. Rispoli, B. Neuwirthová, S. Guerra... Režie Spike Lee (V anglickém
    znění s titulky)"/>
  </channel>
</program>
```

Jmenný prostor, <http://schemas.rozsnyo.com/kolejnet/tv/program> je znova zároveň i adresou dokumentace sebe samého.

Výše uvedená ukázka pochází z adresy na které se nachází vždy aktuální krátký program (běžící pořad a dva následující) všech kanálů: <http://mess.kn.vutbr.cz/tv/program/current.xml>.

Služby

Služby jsou poskytovány na adrese <http://mess.kn.vutbr.cz/tv/program/> a to jenom přes HTTP GET metodu, protože zatím webové služby přes SOAP nezvládáme. Seznam služeb je následovně:

- zjištění aktuálně běžícího pořadu a N následujících
- zjištění programu na den určený relativně ke dnešnímu (0 dnes, +1 zítra, atd..)
- zjištění programu v určitém intervalu
- televizní kanál je určen buď množinou, nebo požadujeme všechny kanály
- výstup je XML, ale existuj možnost změny formátu na straně poskytovatele služby

Aktuální pořad

Pomocí volby **o=C** volíme zobrazování aktuálně běžícího pořadu. Kromě této volby musíme uvést také, že kolik položek programu nám má služba vrátit, takže v případě že chceme jenom aktuálně běžící, uvedeme **n=1** ale když nás zajímá taky co bude následovat, zadáme **n=2**. Obdobně lze získat i větší počet položek, ale služba v základním nastavení má limit 32 položek.

Denní program

Volíme pomocí **o=D** ale vyžaduje další dva parametry - který den chceme (volitelný parametr, výchozí hodnota je 0, tj. dnešní, aktuální den) a parametr specifikující časovou zónu.

Den se udává pomocí **d=** a čísla, které určuje relativní den k dnešku, tj. **-2** je předevečírem, **-1** je včera, **0** je dnes, **1** je zítra, **2** je pozítří atd.

Časová zóna je potřebná k správné detekci hranic dne, protože požadavek na službu může přijít i z jiné země, pro jejich lokální kanál.

Program v intervalu

Tento způsob určení času se aktivuje pomocí parametru **o=R** a vyžaduje povinně 2 parametry, **rf** a **rt** které reprezentují počátek a konec intervalu. Hodnota těchto parametrů je ve formátu YYYYMMDDHHSS nebo ve zkráceném YYYYMMDD, který se pak doplní pomocí 00:00 u času počátku a 23:59 u času konce intervalu.

Pokud v čase určeném pomocí **rf** (počátek) už běží nějaký pořad, tak je vrácen jako první položka. Opačný případ (ideální) nastává jenom když pořad začíná ve stejném čase jako je počátek intervalu.

Volba kanálů

Pokud chceme, aby výsledek obsahoval všechny kanály, použijeme v dotazu parametr **c=A**, ale pokud nás zajímají jenom některé kanály, uvedeme **c=S**. V případě druhé volby musíme ještě specifikovat o které kanály se jedná a to pomocí parametru **ch**, kterému přiřadíme seznam identifikátorů oddělený čárkami.

Volba formátu

Pokud neuvedeme formát, tak bude výstup ve formátu XML jak bylo ukázáno výše. Pro některé aplikace se ale tento formát nehodí, takže existují konvertory, které převedou program u poskytovatele služby na jiný tvar. V současné době existují dva formáty, jeden se aktivuje pomocí **f=csv** a vrací čárkou oddělený seznam informací a druhý, **f=kn** slouží pro stránky kolejni televize, které běží na zabezpečeném serveru (a je tam starší PHP, bez podpory XML).

Chybové hlášení

Pokud je formátem XML, tak se v případě chyby (špatně zadaný vstup nebo překročení limitů) vrátí jeden element, obsahující popisek chyby:

```
<?xml version="1.0" encoding="UTF-8" ?>
<error>number of items is limited to 32 items</error>
```

Pokud je aktivována konverze formátu (parametr **f**), tak se většinou vrátí prázdný dokument.