

# **Seznam YouTube videí**

**Martin Kapal**

---

## Seznam YouTube videí

Martin Kapal

---

# Table of Contents

Úvod .....	iv
1. XML .....	1
Obsah dokumentu .....	1
Popis elementů .....	1
Údaje o seznamu videí .....	1
Údaje o videu .....	1
Údaje o komentářích .....	2
2. XML Schema .....	3
Struktura schématu .....	3
3. XSL Transformace .....	4
Transformace do HTML .....	4
Transformace do PDF .....	4
4. Schematron .....	5

---

# Úvod

Semestrální práce představuje uživatelský seznam videí z webu YouTube [<http://www.youtube.com>]. Tento web posloužil rovněž jako inspirace pro strukturu, obsah a vzhled výstupních souborů.

Práce se skládá z těchto souborů:

- XML dokument – `playlist.xml`
- XML schéma – `playlist.xsd`
- XSL transformace do PDF – `playlist_FO.xsl`
- XSL transformace do HTML – `playlist_HTML.xsl`
- Schematron pravidla – `playlist.sch`

---

# Chapter 1. XML

## Obsah dokumentu

XML dokument `playlist.xml` představuje samotný seznam videí. Uvnitř kořenového elementu `seznam_videi` se nachází titulek seznamu, údaje o autorovi, datum vytvoření a poslední aktualizace seznamu, a údaj o tom, zda se jedná o veřejný nebo soukromý seznam.

## Popis elementů

### Údaje o seznamu videí

Kořenový element `seznam_videi` obsahuje atribut `lid`, který slouží jako identifikátor. Uvnitř kořenového elementu je jako první uveden element `titulek` obsahující stručný název seznamu.

Následují informace o autorovi seznamu. Element `autor` se váže na konkrétního uživatele služby YouTube, a tak je tento element použit také u autorů videí a komentářů. Element obsahuje tyto atributy a podelementy:

- `chid` – atribut označující ID YouTube kanálu daného uživatele
- `jmeno` – jméno a příjmení nebo uživatelské jméno autora
- `google_plus` – číselné ID Google+ profilu autora
- `overeno` – zda se jedná o ověřeného uživatele

Dále je zde element `vytvoreno`, značící datum a čas vytvoření seznamu, aktualizovano datum a čas poslední úpravy seznamu, a nakonec element `verejne`, který značí, zda je seznam veřejně dostupný.

### Údaje o videu

Hned po údajích týkajících se seznamu videí následuje element `videa` obsahující jeden nebo více elementů `video`. Každý element `video` je jednoznačně identifikován atributem `vid`, pomocí kterého se lze na video odkazovat prostřednictvím URL.

Následuje výčet všech elementů uvnitř `video` a jejich význam:

- `titulek` – název videa
- `autor` – autor videa (viz výše)
- `nahrano` – datum a čas nahrání videa
- `publikovano` – datum a čas zveřejnění videa (může se lišit od data nahrání)
- `delka` – délka videa
- `zhlednuti` – počet zhlédnutí videa
- `likes` – počet lidí, kterým se video líbí
- `dislikes` – počet lidí, kterým se video nelíbí
- `kvalita` – seznam dostupných rozlišení videa
- `kvalita` – kvalita videa (rozlišení, příp. počet snímků za sekundu)

- `titulky` – seznam dostupných jazyků s titulky
- `titulek` – jazyk titulků (obsahuje atribut `lang` představující kód ISO 639-1 daného jazyka)
- `popis` – popis videa (může obsahovat řádkování)
- `komentare` – seznam všech komentářů týkajících se tohoto videa (viz níže)

## Údaje o komentářích

Ke každému videu jsou připojeny i textové komentáře od uživatelů. Komentáře mohou být hodnoceny kladně nebo záporně.

Následuje výčet všech elementů uvnitř `komentare` a jejich význam:

- `komentar` – komentář k videu identifikovaný pomocí atributu `cid`, nepovinný atribut `replyto` označuje ID komentáře, na který komentář reaguje
- `autor` – autor komentáře (viz výše)
- `zverejmeno` – datum a čas zveřejnění komentáře
- `text` – vlastní text komentáře (může obsahovat řádkování pomocí elementu `br`)
- `likes` – počet lidí, kterým se komentář líbí
- `dislikes` – počet lidí, kterým se komentář nelíbí

---

# Chapter 2. XML Schema

Pomocí XML schématu se určuje, které elementy a atributy má XML dokument obsahovat, kolik jich má být a v jakém mají být pořadí. Zároveň XML Schema dovoluje definovat datové typy elementů a atributů.

## Struktura schématu

Soubor schématu `playlist.xsd` byl vytvořen ve stylu “salámová kolečka”. Protože se v dokumentu nachází elementy, které se využívají na více místech, tento styl dovoluje definovat elementy pouze jednou a zpětně se na ně odkazovat z více míst.

Téměř všechny definované elementy jsou povinné. Nepovinné elementy jsou aktualizovány (datum poslední aktualizace seznamu), titulky, popis videa a komentář. Počet elementů `video`, `kvalita`, `jazyk` a `komentář` je neomezený.

Co se týče datových typů, pro délku videa jsem zvolil typ `xs:duration`, který umožňuje zápis ve tvaru např. `PT2M18S` (12 minut a 18 sekund). Číselné hodnoty mají datový typ `xs:int`, elementy typu `ano/ne` používají `xs:boolean`.

Ve schématu je použito několik restrikcí. Identifikátory povolují pouze určitou sadu znaků, a tak byla použita restrikce pomocí regulárních výrazů `[A-Za-z0-9]+` či `[A-Za-z0-9\_\-]+`. Zároveň je u některých omezena délka na pevný počet znaků. Regulární výraz je také použit u obsahu elementu `kvalita`, který musí vyhovovat výrazu `\d+p(60)?` (720p, 1080p60 apod.).

---

## Chapter 3. XSL Transformace

Převod XML dokumentu do formátu PDF a HTML je řešeno pomocí XSL transformace. Pro převod do formátu PDF se využívá formátovacích objektů (FO).

Seznam videí je transformován s využitím dvou režimů – `seznam` a `detail`. Oba výpisy jsou řazeny abecedně podle titulků videa pomocí `xsl:sort`. Obdobně jsou řazeny komentáře podle oblíbenosti.

V transformaci je také využito větvení pomocí `xs:if` a `xs:choose`. Také byly použity pojmenované šablony s parametry pro opakované použití při výpisu délky videa nebo data a času.

### Transformace do HTML

Převod XML dokumentu do podoby HTML stránek se provádí pomocí souboru `playlist_HTML.xsl`. Pro vzhled stránky se využívá externího CSS souboru `playlist.css`, ve kterém jsou definovány kaskádové styly pro jednotlivé části dokumentu.

HTML výstup je rozdělen do více souborů. Hlavní stránka `playlist.html` obsahuje stručný výpis všech videí v seznamu. Po kliknutí na miniaturu nebo název videa se zobrazí stránka s podrobnostmi o daném videu. Na stránce je vložen tag `iframe` s URL videa, takže je možné si video přehrát.

### Transformace do PDF

Převod do PDF je řešen pomocí formátovacích objektů (FO). Definice pro tuto transformaci jsou v souboru `playlist_FO.xsl`.

V záhlaví stránek je uveden autor semestrální práce, v zápatí pak číslo stránky. Na první stránce je stručný seznam videí, který se odkazuje na detailní výpis na dalších stránkách pomocí odkazů.



---

# Chapter 4. Schematron

Schematron představuje soubor pravidel, která musí XML dokument splňovat. Pokud je nesplňuje, zobrazí se při validaci chybová zpráva.

Soubor `playlist.sch` kontroluje:

- datum vytvoření a aktualizace seznamu
- datum nahrání a publikování videa
- datum přidání komentáře
- jedinečnost identifikátorů videí