



Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Semestrální práce z KIV/DB2

Semestrální práce č.1 - XML (Oracle)

Jakub Schenk

A21N0069P

schenkj@students.zcu.cz

3. dubna 2022

Obsah

1	Zadání	2
2	Úvod	3
3	Struktura XML	3
3.1	Struktura Hráče	3
3.2	Struktura Videohry	4
3.3	Struktura tabulky Hraje	4
4	Dotazy nad databází	6
4.1	Jednoduchý dotaz na jména hráčů	6
4.1.1	Dotaz	6
4.1.2	Výsledek	6
4.2	Dotaz na typ hráče, jeho příjmení a vlastněné videoherní konzole	6
4.2.1	Dotaz	6
4.2.2	Výsledek	7
4.3	Podobný dotaz pomocí XQuery	7
4.3.1	Dotaz	7
4.3.2	Výsledek	7
4.4	Dotaz pomocí XQuery se složitější podmínkou a přístupem k atributu tagu	8
4.4.1	Dotaz	8
4.4.2	Výsledek	8
4.5	Složitější dotaz pomocí XQuery	8
4.5.1	Dotaz	8
4.5.2	Výsledek	9
4.6	Dotaz pro vyzkoušení požadovaných dotazů	9
4.6.1	Dotaz	9
4.6.2	Výsledek	10
4.7	Dotaz obsahující zbylé požadované příkazy	10
4.7.1	Dotaz	10
4.7.2	Výsledek	10
5	Závěr	11
5.1	Body k zamyšlení a diskuzi	11
5.2	Shrnutí	11

1 Zadání

Připravte realistické datové XML soubory týkající se nahlášeného tématu. Vložte soubory do databáze Oracle a uložte k nim vhodná metadata¹. Vytvořte sadu SQL dotazů využívající XPath nebo XQuery a zodpovídající reálné otázky. Vytvořte dokumentaci vysvětlující použitou datovou strukturu a popisující dotazy přirozeným jazykem. Diskutujte svoje řešení. Požadavky:

- XML soubory obsahují elementy zanořené alespoň do 2. úrovně.
- XML soubory obsahují vícenásobnou hodnotu (více elementů se stejným tagem zanořené pod společným rodičovským elementem).
- Některá data jsou uložena jako samostatné elementy, některá jako atributy elementu.
- Alespoň 4 dotazy.
- SQL dotazy pracují s daty z více XML souborů, nikoliv pouze s jedním konkrétním.
- V sadě dotazů je alespoň jednou použita (lze kombinovat):
 - podmínka na atribut elementu;
 - podmínka na pozici elementu;
 - podmínka na číselnou hodnotu;
 - podmínka na existenci podřetězce v textové hodnotě;
 - podmínka využívající logickou spojku A nebo NEBO;
 - podmínka využívající metadata o XML souboru uložené v relační DB;
 - agregace výsledku.

2 Úvod

Za své téma pro tabulky v relaci m:n jsem si vybral Hráč-hraje-Videohry. Toto téma jsem si zvolil, protože je mi relativně blízké, splňuje zadání a přijde mi poměrně názorné.

Jako atributy hráče jsem zvolil id jakožto primární klíč, jeho jméno a příjmení, věk a výpis videoherních konzolí na kterých potencionálně hraje videohry. Hráč je pak nějakého typu. Pro videohry jsem zvolil opět id jako primární klíč, název videohry, vydavatele, věkovou hranici a výčet videoherních konzolí na kterých si hráč může videohru zahrát. Videohry pak můžou nabývat různých kategorií.

3 Struktura XML

Z důvodu problému s kódováním jsou háčky a čárky odebrány z ukázek kódu. V XML souborech tyto znaky jsou. Zde se jedná o slovo 'Vášnivý' u typu hráče a 'Farná' u jeho příjmení.

3.1 Struktura Hráče

Jednotlivé hráče jsem se rozhodl obalit do tagu <playerbase>, který obaluje celý XML soubor a všechny hráče v něm. <playerbase> představuje obecnou komunitu hráčů, v mém případě několik hráčů které jsem si vymyslel pro příklady v této práci.

Tag <hrac> má již zmíněný atribut "typ" který jednotlivé hráče odděluje. Tag samotný pak obaluje instanci hráče, který má být uložen v tabulce.

Samotný hráč má pak několik výše uvedených tagů. <id> jakožto primární klíč a celočíselná hodnota, která by měla být unikátní. <jmeno> je tag, který obsahuje dva podtagy <krestne> a <prijmeni>. Ty předvídatelně obsahují křestné jméno a příjmení hráče. Tag <vek> je celočíselná hodnota, která není povinná a obsahuje věk našeho hráče. Nakonec <typ_konzole> je set tagů, které obalují jednotlivé konzole, na kterých hráč hraje své videohry.

```
<playerbase>
  <hrac typ="Vasnivy">
    <id>12</id>
    <jmeno>
      <krestne>Magdalena</krestne>
      <prijmeni>Farna</prijmeni>
    </jmeno>
    <vek>18</vek>
```

```

    <typ_konzole>PC</typ_konzole>
    <typ_konzole>Android</typ_konzole>
</hrac>
...

```

3.2 Struktura Videohry

Všechny videohry jednoho XML souboru jsem se rozhodl obalit do tagu `<gamestore>`. `<gamestore>` představuje nějaký vymyšlený obchod s videohrami.

Tag `<videohra>` má již zmíněný atribut "category" který představuje kategorii do které videohra spadá. Tag samotný pak obaluje instanci videohry, která má být uložena v tabulce.

Videohra má pak několik tagů. `<id>` jakožto primární klíč a celočíselná hodnota, která by měla být unikátní. `<nazev>` je tag, který obsahuje název titulu. Tag `<vydavatel>` pak obsahuje vydavatele/společnost která vydala tuto videohru. `<vekova_hranice>` je celočíselná hodnota, která není povinná a jedná se o doporučení pro hráče, které říká jakého věku by měl hráč být před tím, než bude tuto hru hrát. Nakonec `<podpora_konzole>` je set tagů, které obalují jednotlivé konzole, na kterých si hráč může zahrát tuto videohru.

Pro jednoduchost jsem rozhodl o jednoduchých a unikátních `<id>` a vydavatelých videoher. Někteří vydavatelé jsou vymyšlení a myslím, že neexistuje seznam všech videoher, s přidělenými unikátními ID.

```

<gamestore>
  <videohra category="FPS">
    <id>4</id>
    <nazev>Doom</nazev>
    <vydavatel>Doom s.r.o.</vydavatel>
    <vekova_hranice>12</vekova_hranice>
    <podpora_konzole>PC</podpora_konzole>
    <podpora_konzole>PSP</podpora_konzole>
    <podpora_konzole>Android</podpora_konzole>
  </videohra>
  ...

```

3.3 Struktura tabulky Hraje

Poslední tabulku skládá XML soubor obsahující informace o tom jaký hráč "hraje" jakou videohru. Tato tabulka obsahuje dvojici cizích klíčů, a to sice primární klíče ostatních tabulek, a to sice id hrace a id hry. Tato dvojice pak

představuje složený primární klíč. Celý soubor je uzavřen do tagu `<session>` a nese informaci o nějakém časovém období během kterého kteří hráči hráli které hry.

```
<session>
  <hraje>
    <id_hrace>1</id_hrace>
    <id_hry>12</id_hry>
  </hraje>
  ...
```

4 Dotazy nad databází

Po nahrání jednotlivých XML souborů do databáze (podle návodu probíraném na cvičení), jsem se začal dotazovat na data v tabulkách následujícími dotazy.

4.1 Jednoduchý dotaz na jména hráčů

4.1.1 Dotaz

```
SELECT krestne , prijmeni
FROM SP1_XML_HRACI,
      XMLTABLE( 'playerbase/hrac/jmeno' PASSING obsah
      COLUMNS krestne VARCHAR2(50) PATH 'krestne ',
                prijmeni VARCHAR2(50) PATH 'prijmeni '
      );
```

4.1.2 Výsledek

	KRESTNE	PRJMENI
1	Aleš	Bodrý
2	Bohdana	Štědrá
3	Cecil	Bureš
4	Daniel	Houžvička
5	Magdalena	Farná
6	Norman	Hrubý
7	Oliver	Liška
8	Pavel	Líný

Tímto dotazem lze získat jména a příjmení všech hráčů, kteří jsou uloženi v tabulce 'SP1_XML_HRACI', což je tabulka kterou jsem si vytvořil pro uložení XML souborů s hráči.

4.2 Dotaz na typ hráče, jeho příjmení a vlastněné videoherní konzole

4.2.1 Dotaz

```
SELECT typ , prijmeni , konzole
FROM SP1_XML_HRACI,
      XMLTABLE( 'playerbase/hrac' PASSING obsah
```

```

    COLUMNS typ VARCHAR2(50) PATH '@typ',
              prijmeni VARCHAR2(50) PATH 'jmeno/prijmeni',
              konzole XMLTYPE PATH 'typ_konzole'
  );

```

4.2.2 Výsledek

	TYP	PRIJMENI	KONZOLE
1	Přilešitostný Bodrý		<typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole>
2	Přilešitostný Štědrá		<typ_konzole>Android</typ_konzole>
3	Betro	Bureš	<typ_konzole>PS4</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzole>iOS</typ_konzole>
4	Vášnivý	Houžvička	<typ_konzole>PSP</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole><typ_konzole>XBox One</typ_konzole>
5	Vášnivý	Farná	<typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole>
6	Přilešitostný Hrubý		<typ_konzole>Android</typ_konzole><typ_konzole>PS3</typ_konzole>
7	Přilešitostný Liška		<typ_konzole>PC</typ_konzole><typ_konzole>iOS</typ_konzole>
8	Vášnivý	Liný	<typ_konzole>PS4</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole>

Tímto dotazem lze získat příjmení všech hráčů, jejich typ a konzole na kterých hrají videohry. Tyto konzole jsou ale vypsány v XML formátu v jedné řádce. To je způsobeno tím, že se jedná o set hodnot, jehož délka je variabilní

4.3 Podobný dotaz pomocí XQuery

4.3.1 Dotaz

```

SELECT dotaz.*
  FROM SP1_XML_HRACI x,
       XMLTABLE( 'for $x at $i in /playerbase/hrac/jmeno
                 let $p := data($x/prijmeni)
                 let $j := data($x/krestne)
                 let $d := data($x/../../id)
                 where $x/../../vek > 12
                 order by $p
return <li>{$d}. { $j }_ { $p } (vek = {data($x/../../vek)})</li>'
  PASSING obsah ) dotaz;

```

4.3.2 Výsledek

Tento dotaz z databáze hráčů získá informace o hráčích a vypíše jejich jméno, příjmení a id, pokud jsou starší než 12 let. Vypsáný obsah bude seřazen podle příjmení.

	COLUMN_VALUE
1	3. Cecil_Bureš (vek = 13)
2	4. Daniel_Houžvička (vek = 25)
3	2. Bohdana_Štědrá (vek = 32)
4	12. Magdalena_Farná (vek = 18)
5	13. Norman_Hrubý (vek = 28)
6	15. Pavel_Líný (vek = 13)

4.4 Dotaz pomocí XQuery se složitější podmínkou a přístupem k atributu tagu

4.4.1 Dotaz

```
SELECT dotaz.*
FROM SP1_XML_HRACI x,
XMLTABLE( 'for $x at $i in /playerbase/hrac/jmeno
let $p := data($x/prijmeni)
let $j := data($x/krestne)
let $d := data($x/../id)
let $t := data($x/../@typ)
where $x/../vek <30 and $t = "Prilezitostny"
order by $p
return <li>{$d}. {$j}_{$p} (vek = {data($x/../vek)} -
{data($t)})</li>' PASSING obsah ) dotaz;
```

4.4.2 Výsledek

	COLUMN_VALUE
1	1. Aleš_Bodrý (vek = 7 - Příležitostný)
2	13. Norman_Hrubý (vek = 28 - Příležitostný)
3	14. Oliver_Liška (vek = 7 - Příležitostný)

Tento dotaz z databáze hráčů získá informace o hráčích a vypíše jejich id, jméno a příjmení, věk a typ hráče, pokud jsou mladší než 30 let a zároveň spadají do kategorie příležitostných hráčů. Vypsany obsah bude seřazen podle příjmení. (Opět jsou v kódu kvůli kódování odebrány háčky a čárky ze slova 'Příležitostný')

4.5 Složitější dotaz pomocí XQuery

4.5.1 Dotaz

```

SELECT dotaz.*
FROM SP1_XML_VIDEOHRY x,
XMLTABLE( 'for $x at $i in /gamestore/videohra
let $n := data($x/nazev)
let $m := data($x/id)
let $o := data($x/@category)
let $p := data($x/vekova_hranice)
where ($o = "FPS" or $o = "RPG") and $p < 10
order by $n
return <li>{$i}. {$n} (id = {$m} | PG = {$p}) = {$o}</li>'
PASSING obsah ) dotaz;

```

4.5.2 Výsledek

	COLUMN_VALUE
1	3. World of Warcraft (id = 21 PG = 8) = RPG
2	2. NeverWinter (id = 13 PG = 8) = RPG

Dotaz si vyžádá z tabulky videoher 'SP1_XML_VIDEOHRY' všechny videohry, které jsou z kategorie 'FPS' nebo 'RPG' a zároveň mají nastavenou věkovou hranici pod 10 let. Výsledek je pak seřazen podle názvu videohry a jsou vypsané některé informace o dané videohře. (PG je zkratka, která doporučuje minimální věk, tzv. Parental Guidance - za ní je napsána věková hranice)

4.6 Dotaz pro vyzkoušení požadovaných dotazů

4.6.1 Dotaz

```

SELECT typ, prijmeni, konzole, prvni
FROM SP1_XML_HRACI,
XMLTABLE( 'playerbase/hrac' PASSING obsah
COLUMNS typ VARCHAR2(50) PATH '@typ',
prijmeni VARCHAR2(50) PATH 'jmeno/prijmeni',
konzole XMLTYPE PATH 'typ_konzole',
prvni VARCHAR(10) PATH 'typ_konzole[1]'
) WHERE konzole LIKE('%PC%');

```

TYP	PRJMEI	KONZOLE	PRVNI
1 Příležitostný Bodrý		<typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole>	PC
2 Retro	Bureš	<typ_konzole>PS4</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzol...	PS4
3 Vášnivý	Houžvička	<typ_konzole>PSP</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzol...	PSP
4 Vášnivý	Farná	<typ_konzole>PC</typ_konzole><typ_konzole>Android</typ_konzole>	PC
5 Příležitostný Liška		<typ_konzole>PC</typ_konzole><typ_konzole>iOS</typ_konzole>	PC
6 Vášnivý	Líný	<typ_konzole>PS4</typ_konzole><typ_konzole>PC</typ_konzole><typ_konzol...	PS4

4.6.2 Výsledek

Vypsání hráče, jeho typ a první z jeho konzolí, ale jen pod podmínkou, že používá i PC.

4.7 Dotaz obsahující zbylé požadované příkazy

4.7.1 Dotaz

```
SELECT dotaz.*
FROM SP1_XML_VIDEOHRY x,
XMLTABLE( 'for $x at $i in /gamestore/videohra
let $avg := avg(for $y in /gamestore/videohra
return number($y/vekova_hranice))
let $n := data($x/nazev)
let $p := data($x/vekova_hranice)
order by $n
return <li>{$n} {$p}({$avg})</li>' PASSING obsah )
dotaz WHERE x.nazev='gamestore1';
```

4.7.2 Výsledek

COLUMN_VALUE
1 Counter Strikel2(8,75E+000)
2 Dooml2(8,75E+000)
3 Minecraft3(8,75E+000)
4 World of Warcraft8(8,75E+000)

Vyžádání si videoher a vypsání průměru věkové hranice, ale pouze pro první z XML souborů. Výsledek je seřazen podle názvu videohry

5 Závěr

5.1 Body k zamyšlení a diskuzi

Členění do souborů – jaký semantický význam má jeden soubor?

Při vytváření souborů pro mě měl jeden soubor význam jednoho většího celku zaobalujícího několik záznamů, které chci do tabulky uložit. Ať už se jednalo o <playerbase>, <gameshop> nebo <session>, vždy jsem to bral jako obalující prvek kolem několika dat, které měli něco společného.

Vhodnost ukládání dat do elementu vs do atributu. Předpokládal jsem, že je vhodné uložit hlavní informaci přímo do elementu a doplňující informace uložit do atributu elementu. Pokud by se nějaká doplňující informace vztahovala na více elementů, vložil bych jí jako atribut do nadřazeného elementu.

Limitace datového formátu XML pro uložení dat z této domény.

XML formát je poměrně zdlouhavý a dotazy nad ním příliš komplexní v porovnání s klasicky uloženými daty a dotazy nad rozumně postavenými tabulkami.

Limitace v dotazování – na jaké otázky není možné nebo je těžké odpovědět?

Nevím jestli správně rozním otázce, ale z mého pohledu je největší limitací uživatel a jeho schopnost vyjádřit co za data od systému potřebuje. To je ale poměrně obecný problém, když uživatel nedokáže zformulovat jasný dotaz.

Jaké metadata je vhodné uložit do relační databáze při vkládání XML souborů?

Název, pro lepší orientaci ve větší databázi s více XML soubory. Datum vytvoření XML souboru (ne vložení do databáze), pro udržování informace o aktuálnosti informací. Nějkou informací o obsahu daného souboru (pokud to není zřejmé z kontextu tabulky). Případně informaci o struktuře XML pro jednodušší dotazování.

Je problematická existence nepovinných údajů? Ano, při vyhledávání nevyplněných (neexistujících) elementů mohou vznikat potíže.

5.2 Shrnutí

Studiem této problematiky jsem získal mnohem lepší povědomí o možnostech XML formátu a dotazů nad databází. I přesto že jsem si myslel, že

většinu dotazů budu provádět pomocí XPath, protože jsem s ním měl drobné zkušenosti, nakonec se XQuery ukázalo být (minimálně pro mě) příjemnější alternativou. Věřím že jsem body zadání splnil a probranou látku v rámci semestrální práce dostatečně zopakoval.