MASARYKOVA UNIVERZITA
INSTITUT BIOSTATISTIKY A ANALÝZ

Databázové systémy a SQL

Lekce 10 – Úvod do XML

Daniel Klimeš

XML - Co to je?

- Formát pro přenos strukturovaných dat
- Text se značkami (tagy)
- Připomíná HTML
- Samotný standard specifikuje jen minimum značek
- Položky
  - Elementy
  - Atributy
- Zanořená struktura
  - 1 kořenový element (root)
  - N zanořených elementů

Jak to vypadá?

- Hlavička
  - <?xml version="1.0" encoding="UTF-8"?>
- Tagy
  - Elementy
    - <tag>data</tag>
    - <question>Datum narození</question>
  - Atributy
    - <element atribut=„text">
    - <name lang="cz">
- Hierarchie elementů
  - Root element
  - Zanořené elementy
    - <el_root>
      <el_nest>text</el_nest>
    - </el_root>
    - <cluster>
      <question>Datum vyšetření</question>
    - </cluster>

Jak pracovat s XML?

- Textový editor – Notepad
- XML editor – XML Spy
- Moderní databáze – ORACLE
  - Nadstavba SQL
- Programovací jazyky
  - Parsery

Související technologie

- XML schéma – šablona XML dokumentu
- XSLT – transformace XML do jiného formátu (jiné XML, prostý text , HTML)
- XPATH – formát vyhledávacích dotazů do XML dokumentů
- Xquery – pokročilé vyhledávání v XML

Zdroje informací o XML

- Internet
  - *www.w3.org*
  - *www.w3schools.com*

XML a ORACLE

- Podpora ukládání XML dat
  - Datový typ XMLTYPE
- Podpora importu dat v XML formátu
  - Přímé vložení pomocí INSERT příkazu
  - XML repository
    - Propojení úložiště s tabulkou
- Podpora exportu dat v XML formátu
  - Funkce DBMS_XMLGEN.getXML
    - Výstup SQl dotazu ve formátu XML
  - Funkce sys_XMLGen
    - Export jednoho sloupce tabulky jako jednoduché XML
  - Speciální SQl funkce - skládání XML stromu
    - XMLELEMENT
    - XMLATTRIBUTES
    - XMLAGG
    - …

Práce s XML

- Vytvoření tabulky
  - Datový typ CLOB pro velké textové řetězce
  - Datový typ XMLTYPE pro XML formát
  - CREATE TABLE jmeno_tabulky
    (velky_text CLOB,
    xml_sloupec XMLTYPE);
  - SELECT DBMS_XMLGEN.getXML('SELECT study_id, study_name FROM studies') FROM DUAL

Práce s XML

- Generování XML
  - SELECT
    DBMS_XMLGEN.getXML('SELECT study_id, study_name FROM studies')
    FROM DUAL
  - INSERT INTO jmeno_tabulky (velky_text)
    SELECT
    DBMS_XMLGEN.getXML('SELECT study_id, study_name FROM studies')
    FROM DUAL
  - UPDATE jmeno_tabulky SET xml_sloupec = XMLTYPE(velky_text)
  - INSERT INTO jmeno_tab (xml_sloupec)
    SELECT SYS_XMLGEN(study_id) FROM studies

Dotazování XML

- SELECT COUNT(*) FROM jmeno_tabulky WHERE existsNode(xml_sloupec, '//STUDY_NAME') = 1
- SELECT COUNT(*) FROM jmeno_tabulky WHERE existsNode(xml_sloupec, '//ROW[STUDY_ID=3]') = 1
- SELECT extract(xml_sloupec, '//ROW[STUDY_ID=3]/STUDY_NAME').GetClobVal()
  FROM jmeno_tabulky WHERE existsNode(xml_sloupec, '//ROW[STUDY_ID=3]') = 1
- SELECT extractvalue(xml_sloupec, '//ROW[STUDY_ID=3]/STUDY_NAME')
  FROM jmeno_tabulky WHERE existsNode(xml_sloupec, '//ROW[STUDY_ID=3]') = 1

Dotazování XML

- SELECT extractvalue(value(t), '//STUDY_NAME')
  FROM jmeno_tab,
  table(xmlsequence(extract(jmeno_tab.xml_sloupec, '//ROW'))) t
  WHERE extractvalue(value(t), '//STUDY_ID') < 100

Zkouškové otázky - příklady

- Který formulář má nejvíce otázek? (Cluster_description)
- Která studie má nejvíce pacientů? (study_name)
- Do kterých studií se zařazovali pacienti v roce 2008 (study_name)
- Vypište názvy všech otázek ve skupině XY
- Vypište pracoviště, která nezařadila žádného pacienta
- Vypište pracoviště, která se účastní více studií
- Vytvořte pohled study_name, počet zařazených pacientů
- Zjistěte průměrnou hodnotu, minumum , maximum otázky XY
- Vypište věk pacientů při zařazení do studie XY
- Vypište hodnotu otázky XY pro všechny pacienty (včetně nevyplněných)
- Jaký je průměrný počet vyplněných formulářů na pacienta u studie XY