

Similarity-Based Databases: An Introduction Homework (in Czech)

Vilem Vychodil (Palacky University, Olomouc)

DAMOL

DATA ANALYSIS AND MODELING LAB

Palacky University, Olomouc, Czech Republic



esf
european
social fund in the
czech republic



EUROPEAN UNION



MINISTRY OF EDUCATION,
YOUTH AND SPORTS



OP Education
for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

Zadání

Navrhněte implementaci podobnostního dotazovacího systému popsaného na přednášce, který bude implementován uvnitř klasické relační databáze.

Požadavky:

- Implementaci proveďte pro databázový systém PostgreSQL, <http://www.postgresql.org/docs/8.3>
- podle potřeby využijte vhodně procedurálního jazyka PL/pgSQL <http://www.postgresql.org/docs/8.3/static/plpgsql.html>
- řešení dodejte tak, aby dodržovalo **stanovený formát** (viz dále)
- pošlete **nejpozději 30. ledna 2012** na vilem.vychodil@upol.cz

Hodnocena bude **správnost, elegance a verzatilita řešení.**

Formát

Obecný formát tabulek

- Pokud má tabulka sloupec se jménem *rnk*, pak jeho hodnoty (desetinná čísla) budeme považovat za **ranky**.
- Pokud tabulka tento sloupec nemá, všechny ranky jsou buď 0 nebo 1.

Struktura pravdivostních hodnot a podobnost

- Budeme uvažovat Łukasiewiczzovu strukturu pravdivostních hodnot na intervalu $[0, 1]$.
- Podobnosti na doménách by měly být realizovány jako uložené procedury, přitom by mělo být možné, aby uživatel mohl definovat podobnost na doméně jako
 - vlastní uloženou proceduru nebo
 - informace o podobnostech by byly ukládány do samostatných k tomuto účelu zavedených tabulek (jejich organizace může být libovolná).

Tvar odevzdaných souborů: Úkoly odevzdejte v souboru komprimovaném programem ZIP obsahujícím následující soubory:

- `simildb.sql` – soubor, který po natažení vytvoří potřebná data, procedury aj.
- `demo.sql` – soubor demonstrující dotazování (viz dále)

Obsah souboru demo.sql

Do souboru demo.sql vyjádřete ve vámi navrženém rozšíření následující dotazy (v souboru simildb.sql si připravte všechny potřebné tabulky a podobnosti, do demo.sql dávejte jen dotazy, před každý z nich uveďte číslo pomocí komentáře ve tvaru `/* 1 */`, `/* 2 */`, ...):

- 1 $\pi_{\{y_1, y_2\}}(\mathcal{D}_1)$,
- 2 $\sigma_{y_1 \approx 10}(\mathcal{D}_1)$,
- 3 $\sigma_{y_1 \approx 10}(\pi_{\{y_1, y_2\}}(\mathcal{D}_1))$,
- 4 $0.5 \rightarrow \sigma_{y_1 \approx 10}(\pi_{\{y_1, y_2\}}(\mathcal{D}_1))$,
- 5 $top_3(0.5 \rightarrow \sigma_{y_1 \approx 10}(\pi_{\{y_1, y_2\}}(\mathcal{D}_1)))$,
- 6 $\sigma_{y_1 \approx x_1}(\mathcal{D}_1 \times \mathcal{D}_2)$,
- 7 $\sigma_{y_1 \approx x_1}(\mathcal{D}_1 \times \pi_{\{x_1, x_2\}}(\mathcal{D}_2))$,
- 8 $0.8 \rightarrow \sigma_{y_1 \approx x_1}(\mathcal{D}_1 \times \pi_{\{x_1, x_2\}}(\mathcal{D}_2))$.

Kde \mathcal{D}_1 je na schématu $\{y_1, \dots, y_4\}$ a \mathcal{D}_2 je na schématu $\{x_1, \dots, x_3\}$. Data a podobnosti zvolte tak, aby výsledky žádných dvou různých dotazů (1)–(8) nebyly stejné.