

INQUIRER v1.0 – sistem za postavljanje upita bazi podataka na prirodnom jeziku

INQUIRER je sistem koji omogućava prevođenje upita zadatih na prirodnom jeziku (npr. engleskom) na SQL upite. Sastoji se iz 9 modula pisanih u C jeziku, uz koje ide i primer programa za administraciju i primer programa za postavljanje upita. Upiti se zadaju pomoću fraza koje postoje u odgovarajućim datotekama fraza. Postoje tri vrste fraza, prema kojima se upit deli na tri dela: restrikciju, projekciju i sortiranje. Za održavanje jednog ovakvog sistema potreban je administrator sa dobrim znanjem SQL jezika, koji će napuniti bazu najčešće korišćenim frazama. Korisnici dalje mogu definisati sinonime tih fraza i tako prilagođavati sistem svojim potrebama, bez potrebe za poznavanjem brojnih nerazumljivih konstrukcija SQL jezika.

Ključne reči: SQL, prirodni jezik, upiti, baza podataka

Uporedo sa razvojem baza podataka, rasla je i težnja korisnika za efikasnim i jednostavnim postavljanjem upita. Kao rezultat pokušaja tvoraca sistema za upravljanje bazom podataka da udovolje korisničkim potrebama, nastala su razna sredstva za olakšano postavljanje upita, među kojima je i SQL jezik. SQL je do sada najpribližniji prirodnom jeziku zahvaljujući prisustvu mnogobrojnih redundantnih sintaksnih konstrukcija, popularno zvanih syntactic sugar, kao što je BETWEEN ... AND ... operator, ali je i dalje suviše formalan za običnog korisnika. Osim toga, za postavljanje upita na SQL jeziku korisnik mora imati određeno poznavanje relacionog modela. Ovaj sistem je pokušaj da se korisnicima omogući da postavljaju upite bazi podataka bez ikakvog neophodnog teorijskog predznanja.

Realizacija

Sistem je sastavljen od 9 modula pisanih u C jeziku, koji omogućavaju tokenizaciju unetog teksta, pronalaženje fraze u fajlu, zamenu sinonima odgovarajućom frazom i generisanje SQL upita. Ovih 9 modula

*Vojislav Stojković
(1978), Beograd,
Stevana Sremca 4,
učenik 4. razreda
Matematičke gimnazije
u Beogradu
vojas@sezampro.yu*

predstavljaju jezgro koje se može integrisati sa raznim korisničkim interfejsima, čime se dobija kompletan sistem za postavljanje upita. Moduli su sledeći:

1) DICT.C Modul za rad sa rečnikom. Omogućava pretraživanje i dodavanje reči u rečnik, kao i učitavanje celokupnog sadržaja jednog rečnika u drugi.

2) PARSE.C Modul koji služi za tokenizaciju unete fraze. Koristi funkcije modula DICT.C za pronalaženje odgovarajućih reči u frazama. Generiše niz identifikatora reči i parametara, koji predstavljaju promenljive delove u frazama.

3) PARAM.C Modul za rad sa listom parametara. Parametri iz svih unetih fraza se slažu u jedinstvenu listu iz koje se čitaju pri generisanju SQL upita, na osnovu broja parametra unutar fraze i zapamćenog stanja brojača parametara.

4) FROM.C Modul zadužen za rad sa FROM klauzulom SQL upita. Pomoću ovog modula se navode tabele, tipovi spajanja tih tabela i uslovi tih spajanja, što će se na kraju upisati u FROM sekciju dobijenog upita.

5) TABLE.C Modul za rad sa frazama tabela. Ove fraze služe za određivanje tabele za koju su vezane fraze restrikcije.

6) RESTRICT.C Modul zadužen za rad sa frazama restrikcije. Ove fraze sadrže podatke za FROM, WHERE, GROUP BY i HAVING klauzule izlaznog SQL upita i određuju koji će od predikata ALL i DISTINCT stojati pre liste atributa.

7) PROJECT.C Modul za rad sa frazama projekcije. Fraze projekcije određuju listu atributa i predikat ALL/DISTINCT.

8) ORDER.C Modul za rad sa frazama sortiranja. Prema frazama sortiranja konstruiše se ORDER BY klauzula finalnog SQL upita.

9) QUERY.C Modul zadužen za generisanje izlaznog SQL upita na osnovu prikupljenih podataka.

Osim ovih 9 modula unutar ovog sistema nalazi se još 12 header fajlova, od kojih 9 odgovaraju modulima, a ostala 3 sadrže promenljive i konstante zajedničke za ceo sistem. Uz sistem idu i primeri korišćenja, ADMIN.C i USER.C, kao i skriptovi za njihovo kompajliranje i linkovanje.

INQUIRER sa korisničke strane

U ovoj sekciji se objašnjava mogući način korišćenja INQUIRER-a, onako kako je to urađeno u programu USER.C

Upiti se zadaju iz tri dela. U prvom delu se navodi u kojim tabelama se vrši pretraga, tj. fraze odabira tabela. Program daje prompt "find> " na koji se odgovara frazom. U slučaju da ne želimo više da uključujemo tabele, zadajemo prazan string umesto fraze. U drugom delu se određuju kriterijumi pretrage preko fraza restrikcije. Za ovaj deo karakterističan je

prompt "condition> " na koji se odgovara frazom restrikcije, posle čega sledi prompt "operator> ". Kao odgovor unosi se neki od logičkih operatora "and" ili "or", ili prazan string, u slučaju da smo završili sa restrikcijom. U trećem delu se određuje šta se od nađenih podataka želi prikazati, tj. unose se fraze projekcije. Program u ovom delu daje prompt "display> ". Način rada je isti: odgovara se frazom ili praznim stringom. I konačno, u četvrtom delu, ukucavanjem fraza za sortiranje, određuje se kako će dobijeni podaci biti sortirani. Za ovu fazu sastavljanja upita karakterističan je prompt "sort by >". Po završetku sastavljanja upita dobijamo izlaz u obliku odgovarajućeg SQL upita, koji posle možemo proslediti nekom sistemu za postavljanje SQL upita.

Primer zadavanja INQUIRER upita

```
find> products
find>
condition> purchased by customers from "Belgrade"
condition and
condition> made in "Italy"
condition>
display> product names
display> product colors
display>
sort by> product prices
sort by>
```

Odgovarajući SQL upit:

```
SELECT DISTINCT Product.Name, Product.Color
FROM (Product INNER JOIN Order ON
      Product.OrdID = 3D Order.ID)
INNER JOIN Customer ON Order.CustID =3D Customer.ID
WHERE Customer.City =3D "Belgrade" AND
      Product.Origin =3D "Italy"
ORDER BY Product.Price;
```

U slučaju da u nekoj fazi sastavljanja upita odgovarajuća fraza ne postoji, korisniku se nudi mogućnost da kreira sinonim: treba uneti neku postojeću frazu sa istim parametrima, a program će sam utvrditi pravilo po kome se parametri zamenjuju.

Primer zadavanja sinonima:

```
condition > shipped to "Belgrade" city WARNING:
Unrecognized restriction phrase!
Create synonym (Y/N)? Y
Rephrase: purchased by customers from "Belgrade"
```

NAPOMENA: Tekst fraza određuju korisnici i administrator, i od njih zavisi da li će te fraze imati ikakvog smisla u realnom svetu. Dakle, ako korisnik ili administrator definiše frazu „yellow dog eats sea“, tako da ona odgovara nekom delu upita, ona će raditi kako valja – iako ništa ne znači.

Zaključak

Ova verzija INQUIRER sistema sposobna je za generisanje upita selekcije i to bez mogućnosti ugnježđenih SELECT naredbi. Sledeće verzije bi trebalo da podrže pomenuto ugnježdavanje, operaciju unije, INSERT, UPDATE i DELETE upite. Program je i na ovom stupnju razvoja veoma primenljiv, iako nije moguće zadati baš svaki upit.

Literatura

- [1] Alagić, Suad. 1985. *Relacione baze podataka*, drugo izdanje. Sarajevo: Svjetlost.
- [2] *Microsoft Access Help – SQL Reference*

Vojislav Stojković

INQUIRER – The System Translate Queries in Native Language

INQUIRER is a system that translates queries in native language (e. g. English) to SQL queries. It consists of 9 modules written in C, sample administration program and sample query construction program. Queries are built from phrases that are stored in phrase storage files. There are three types of phrases: restriction, projection and order phrases, and the queries are divided in three parts corresponding to those three types. The system is administrated by an administrator with a solid knowledge of SQL, who „feeds“ the system with most commonly used phrases. The users can define their synonyms for those phrases and in that way customize the system without need to know numerous incomprehensible SQL constructions.

Key words: SQL, native language, queries, database.

