

# Interfaţa în limbaj natural pentru bazele de date ale sistemelor de întrebare – răspuns (QA)

Anatol POPESCU

Universitatea Technică a Moldovei  
[an.popescul@gmail.com](mailto:an.popescul@gmail.com)

Victoria MAXIM

Universitatea Technică a Moldovei  
[maxivica@yahoo.com](mailto:maxivica@yahoo.com)

**Abstract** — A natural language search engine would in theory find targeted answers to user questions (as opposed to keyword search). For example, when confronted with a question of the form 'which U.S. state has the highest income tax?', conventional search engines ignore the question and instead do a search on the keywords 'state, income and tax'. Natural language search, on the other hand, attempts to use natural language processing to understand the nature of the question and then to search and return a subset of the web that contains the answer to the question. If it works, results would have a higher relevance than results from a keyword search engine.

From a commercial standpoint, advertising on the results page could also be more relevant and could have a higher revenue potential than that of keyword search engines.

**Index Terms** — database, interface, natural language, search engine, query.

## I. INTRODUCERE

În trecut, interfeţele în limbaj natural au determinat utilizatorii la antropomorfizarea computerului, sau cel puţin la atribuirea lui a unei ineligenţe mai mult decât justificată. Acest fapt a condus spre aşteptări nerealiste ale utilizatorilor referitoare la capacităţile sistemului. Asemenea aşteptări fac dificilă însuşirea restricţiilor sistemului, dacă i se atribuie prea multe capacităţi, ducând spre o dezamăgire atunci când sistemul eşuează în performanţa scontată.

Un articol din anul 1995 denumit 'Natural Language Interfaces to Databases – An Introduction', descrie unele probleme:[2]

- Ataşarea modificatorilor  
Cererea „Lista tuturor angajaţilor din companie cu un permis de conducere” este ambiguă, doar dacă cunoaşteţi companii care nu pot avea permis de conducere.
- Conjunctia si disjunctia  
“Listaţi toţi solicitanţii care locuiesc în California şi Arizona.” este ambiguă, doar dacă ştiţi că o persoană nu poate locui în două locuri simultan.
- Rezolvarea anaferei  
- într-o întrebare autoreferenţială, rezolvarea a ceea ce utilizatorul a avut în vedere prin „el” sau „ea”.

## II. CREAREA ŞI COMPONENTELE INTERFEŢEI ÎN LIMBAJ NATURAL

Interfeţele în limbaj natural sunt interfeţele în care comunicarea om – maşină se face prin limbajul natural al utilizatorului. [1]

Componentele de bază ale interfeţelor în limbaj natural pentru bazele de date sunt următoarele:

- ✓ Procesorul de limbaj natural, funcţia căruia este reprezentarea internă a întrebării în limbaj natural (de exemplu, arbori sintactici, reţele semantice, structuri-cadru etc).
- ✓ Dicţionarul electronic de cuvinte şi expresii cu informaţie lexicală, gramaticală şi semantică codificată, în dependenţă de nivelul şi metodele analizei limbajului natural.

- ✓ Modelul sursei de date care reflectă structura de bază a bazelor de date, precum şi acele părţi care sunt esenţiale pentru interfaţa în limbaj natural proiectată.
- ✓ Modulul de generare a cererii formale utilizat pentru transferul întrebării de intrare în limbaj natural din reprezentarea internă a sistemului în limbajul formal al sursei de date.

Etapele creării unei interfeţe în limbaj natural sunt:

1. Elaborarea schemei bazelor de date.  
Acest pas depinde în mod considerabil de tipul bazei de date şi include: alegerea BD, construirea modelului BD, selectarea acelor tabele ce vor fi incluse în interfaţă, stabilirea legăturilor între tabele.
2. Construirea modelului domeniului ales.  
Elaborarea modelului unui domeniu constă în crearea diagramelor claselor (entităţilor), construirea schemelor conceptuale, acumularea unui „depozit” de componente semantice, care sunt definite de fragmente ale reţelelor semantice. Componentele pot fi predicate, mulţimi şi hipervalori.
3. Organizarea dicţionarului.  
Conţinutul dicţionarului este alcătuit din trei părţi principale: partea universală, lexicul din baza de date şi cuvintele introduse de proiectantul sistemului. Partea universală apare la elaborarea interfeţei, iar pe parcursul funcţionării sistemului este posibilă completarea dicţionarului de bază cu un vocabular tematic (tezauros).
4. Conectarea câmpurilor textuale ale bazelor de date.  
Acest proces presupune completarea vocabularului cu lexicul din câmpurile bazelor de date. Frecvenţa conexiunii câmpurilor este stabilită la configurarea sistemului; este posibilă conectarea câmpurilor la fiecare pornire a sistemului sau la comanda utilizatorului.

5. Completarea dicţionarului cu cuvinte necunoscute din cadrul întrebării (cererii).  
Atunci când se introduce o întrebare ce conţine cuvinte care nu sunt în dicţionar şi nu se potrivesc unui model specificat, aceste cuvinte sunt afişate utilizatorului într-o listă pentru a fi introduse în dicţionar sau pentru corectare.
6. Testarea interfeţei şi editarea listei de întrebări.  
Pentru etapa de testare sistemul suportă introducerea unei liste de întrebări tipizate în limbaj natural. În rezultatul testării proiectantul sistemului completează dicţionarul cu lexicul din întrebări, atribuind clase semantice articolelor din dicţionar şi stabilind valorile atributelor lor.

### III. AVANTAJELE INTERFEŢELOR ÎN LIMBAJ NATURAL

Avantajele creării interfeţelor în limbaj natural pentru sistemele de dirijare a bazelor de date sunt următoarele:

- ✓ o formare profesională minimă a utilizatorului, având în vedere faptul că limbajul natural este mijlocul cel mai familiar şi convenabil de comunicare,
- ✓ simplitatea întrebării în limbaj natural, deoarece ea uneori este mai scurtă decât formularea întrebării într-un limbaj formal,
- ✓ viteza mare de generare a unei întrebări – utilizatorul poate formula imediat o întrebare corectă în limbaj natural, deoarece această reprezentare este cea mai firească pentru om,
- ✓ nivelul înalt de modelare a domeniului ales; este de remarcat faptul, că interfeţele tradiţionale, de obicei, nu conţin modelul domeniului.

### IV. CONCLUZII

Baza metodei propuse de construire a interfeţei în limbaj natural este abordarea semantic orientată a procesării limbajului natural pe domenii restrânse. Această abordare

permite utilizarea unui număr de principii:

- ✓ Abstractizarea de un limbaj natural concret.
- ✓ Abstractizarea de un anumit domeniu.
- ✓ Abstractizarea de o bază de date specifică.

### REFERENCES

- [1] [http://en.wikipedia.org/wiki/Natural\\_language\\_user\\_interface](http://en.wikipedia.org/wiki/Natural_language_user_interface)
- [2] I. Androutsopoulos, G.D. Ritchie, P. Thanisch, Natural Language Interfaces to Databases – An Introduction, Department of Artificial Intelligence, University of Edinburgh.
- [3] D. Knuth, The Art of Programming. Addison-Wesley, 1973.
- [4] [Chatty, S. \(1994\)](#). "Issues and Experience in Designing Two-Handed Interaction." In *CHI '94: Human Factors in Computing Systems, Conference Companion*. ACM. 253-254.
- [5] [Dix, A., Finlay, J., Abowd, G. and Bealle, R. \(1993\)](#). *Human-Computer Interaction*. Prentice-Hall.
- [6] [Hill, I. \(1983\)](#). "Natural language versus computer language." In M. Sime and M. Coombs (Eds.) *Designing for Human-Computer Communication*. Academic Press.
- [7] [Mayhew, D. \(1992\)](#). *Principles and Guidelines in Software User Interface Design*. Prentice-Hall.
- [8] [Slator, B., Anderson, M. and Conley, W. \(1986\)](#). "Pygmalion at the Interface." *Communications of the ACM* 29, 7, 599-604. B. Shneiderman. 1992. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (Second Edition)*. Addison-Wesley.
- [9] [Véronis, J. \(1991\)](#). "Error in natural language dialogue between man and machine." *International Journal of Man-Machine Studies* 35, 187-217.
- [10] [Zoltan-Ford, E. \(1991\)](#). "How to get people to say and type what computers can understand." *International Journal of Man-Machine Studies* 34, 527-547.