

ETAPELE DE IMPLEMENTARE A SISTEMULUI DE „ÎNTREBARE-RĂSPUNS”

LAZU Victoria

Universitatea Tehnică a Moldovei

Recenzent: BOBICEV Victoria dr., conf. univ.

Cuvinte cheie: *căutarea și extragerea informației, sistem de întrebare-răspuns*

Sistemele de extragere a răspunsurilor la întrebări în limbajul natural uman se înscriu în categoria sistemelor de achiziție a informației. Dezvoltarea sistemelor capabile să răspundă la întrebările formulate de utilizator în limbajul natural este una din sarcinile domeniului regăsirii informației. Scopul de bază al unui astfel de sistem constă în a asigura un răspuns la întrebarea utilizatorului care ar îndeplini următoarele trei condiții: să fie corect; să fie formulat tot în limbaj natural; să fie suficient de succint.

Crearea unui sistem de întrebare-răspuns reprezintă un proces complex care implică multe componente și resurse importante. Sistemul ce este dezvoltat în cadrul proiectului „Cercetarea în domeniul de Regăsire a Informației în scopul creării sistemului electronic de informare publică” este lansat în cooperare cu Institutul de Dezvoltare a Societății Informaționale (IDSI) și urmărește scopul de a cerceta problemele de regăsire a informației și de a crea un sistem automat de Întrebare-Răspuns pentru vorbitorii de limbă română. Sistemul va procesa interogările și va extrage răspunsurile la acestea din documentele relevante. Prima etapă de implementare a proiectului prevede elaborarea unui sistem de întrebare-răspuns de tip închis din punct de vedere al informației vehiculate în baza unei colecții de documente legislative. Sistemul format cuprinde un modul de procesare a întrebării și un modul de căutare a fragmentelor de text care conțin un potențial răspuns[1].

Procesarea întrebării se efectuează în trei etape consecutive în rezultatul cărora are loc formalizarea interogărilor limbajului natural. Astfel, etapele includ analiza părților întrebării, formarea listei de Cuvinte cheie și reformularea interogării în răspuns. Din punct de vedere semantic întrebările se clasifică în setul predefinit de tipuri: DEF-definiție, N-nume/persoană, TMP-timp, LST-listă, LOC-loc[2].

După finalizarea operațiilor din cadrul primului modul are loc căutarea răspunsului propriu-zis în documentele disponibile. Căutarea este efectuată pe fragmente de texte în baza cuvintelor cheie obținute în cadrul procesării întrebării. În cazul în care sistemul identifică un cuvânt cheie, fragmentul de text în care se găsește acesta este memorat în lista răspunsurilor posibile. Pentru a determina și a prezenta utilizatorului cel mai potrivit răspuns la întrebarea sa, sistemul calculează ponderea pentru fiecare fragment memorat cu ajutorul formulei care ia în considerare paragraful în care se efectuează căutarea, ponderea paragrafului, numărul de ordine a frazei-cheie în lista frazelor-cheie.

Prin urmare, sistemul afișează răspunsul care are ponderea maximă, însă celelalte răspunsuri sunt înregistrare într-un fișier disponibil pentru utilizator prin accesarea link-ului apărut pe aceeași pagină. Sistemul afișează în același timp titlul, capitolul și articolul din actul legislativ în care a fost găsit răspunsul. În cazul în care utilizatorul este satisfăcut de răspunsul oferit, sistemul îl memorează în lista întrebărilor și răspunsurilor care au fost deja căutate de alți utilizatori. În urma analizei răspunsurilor întrebărilor s-a observat că sistemul răspunde corect la 35.8% de întrebări. Iar cea mai mare acuratețe au răspunsurile la interogările de tip „Definiție” și anume 0,54[2].

Bibliografie:

1. V.Bobicev Preprocesarea textelor în sistemele de tip „întrebare-răspuns”. 7th International Conference on Microelectronics and Computer Science, Chișinău, 2011
2. V. LAZU, T. PRODAN Procesarea și analiza întrebărilor într-un sistem de „ÎNTREBARE-RĂSPUNS”, In International Conference ICTEI 2012, 17-20mai, Chișinău. pp.376-380 Vol. I