

Studiul tipurilor de întrebări din sistemul de întrebare-răspuns

Raisa BOTNARU Victoria BOBICEV

Universitatea Tehnică a Moldovei

victoria.bobicev@rol.md

Rezumat — În articolul dat se descrie clasificarea întrebărilor pentru sistemul de tip întrebare-răspuns ce se crează în baza documentelor IDSI. Au fost elaborate 29 tipuri de întrebări care se includ în 8 clase majore propuse la nivel internațional. Majoritatea întrebărilor au semantică concretă și se clasifică comparativ ușor iar unele întrebări au semantică ambiguă și pot fi incluse în două sau mai multe clase. Aceste întrebări necesită o analiză mai detaliată. Etapa următoare descrisă este analiza întrebărilor în scopul pregătirii datelor necesare pentru motorul de căutare care va extrage răspunsul din textele documentelor. În cadrul acestei analize se determină tipul întrebării ce definește tipul răspunsului și cuvintele-cheie care sunt transferate motorului de căutare cu scopul găsirii răspunsului.

Index Terms — analiza întrebărilor, analiza lingvistică, clasificarea întrebărilor, cuvinte-cheie, sistem întrebare-răspuns.

I. INTRODUCERE

Sistemele de întrebare-răspuns (SÎR) vor fi utilizate de publicul larg numai atunci când el va putea răspunde satisfăcător la toate tipurile de întrebări și, în deosebi, la cele simple. Aceste sisteme reprezintă o continuare în evoluția mijloacelor de căutare a informației. Particularitatea acestor sisteme constă în extragerea răspunsurilor la întrebările în limbaj natural puse pe seturile de documente.

Sistemele de extragere a răspunsurilor la întrebări în limbajul natural uman se înscriu în categoria sistemelor de achiziție a informației.

Cunoașterea are loc în felul următor: se unește un sistem informațional cu un ambient. Drept sistem informațional poate servi un individ care interacționează în mod direct și conștient cu ambientul și care își construiește prin observație universul de cunoaștere, dar și un sistem tehnic care primește informația de la diverși agenți umani și o procesează sau o pune, în continuare, la dispoziția altor persoane. Internetul servește ca un astfel de sistem informațional care, acționând reciproc cu mediul agenților umani, își construiește o bază proprie de cunoaștere.

Deseori, cunoștințele căpătate prin intermediul internetului sunt în forma în care agenții umani o prezintă sistemului, deci în limbaj natural uman.

În prezent, internetul este cea mai mare sursă de cunoștințe ce se extinde și se actualizează încontinuu. Internetul mai este și una din cele mai accesibile locații în care aceste cunoștințe pot fi consultate. Însă dezvoltarea rapidă a internetului nu are loc fără aspecte negative, și anume: din cauza volumului mare de informații disponibile, găsirea informației necesare poate fi, uneori, dificilă sau nesigură.

Cele mai eficiente metode de găsire și de acumulare a informației o reprezintă, la ora actuală, motoarele de căutare, scopul cărora este de a oferi utilizatorului un set de articole sau pagini web în care aceasta să poată găsi informația necesară. Deseori articolele propuse de

motoarele de căutare nu îndeplinesc dezideratul utilizatorului de a căpăta un răspuns satisfăcător. În plus, ele nu oferă răspunsul concret la întrebarea utilizatorului, dar numai un set de pagini web, și utilizatorul e nevoit să-și extragă informația necesară.

Etapa următoare în domeniul achiziției informației este alcătuită din dezvoltarea sistemelor capabile să răspundă la chestiunile formulate de utilizator în limbajul natural. 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 fi suficient de succint

Un sistem de răspunsuri la întrebări necesită o procesare a limbajului natural mult mai complexă decât sistemele de achiziție de documente.

În teorie, procesarea limbajului natural este un subiect foarte atractiv fiindcă se aplică în așa domenii ca cel al interacțiunii dintre om și mașină. În practică se observă o serie de dificultăți importante ce se datorează modului diferit în care o afirmație în limbaj natural poate fi interpretată și a multitudinii de sensuri pe care cuvintele ce-l alcătuiesc le pot căpăta. Sistemele de răspuns la întrebări posedă problemele acestea.

Metode folosite în dezvoltarea sistemelor de tip întrebare-răspuns. Pentru alcătuirea unui sistem de răspuns la întrebări există două variante:

Abordarea de tip shallow, bazată pe cuvinte cheie. În această metodă se folosesc cuvinte cheie pentru a găsi fragmente și propoziții în text care ar putea reprezenta răspunsuri satisfăcătoare la întrebări. Aceste potențiale răspunsuri urmează să fie analizate mai profund pentru a se aprecia dacă răspunsurile sunt reale sau nu. Această metodă poate fi folosită cu succes în cazul întrebărilor scurte, facile, când se caută nume, date, localități, cantități, cifre.

Abordarea de tip deep. Această metodă cere o analiză mai sofisticată, o procesare sintactică, semantică și contextuală. În această categorie pot fi încadrate mai multe metode: abduction named-entity, recognition, relation detection, etc.

Alegerea unuia dintre cele două modele depinde de complexitatea întrebărilor formulate și de gradul de performanță pe care-l așteptăm de la sistem. Sistemele din categoria a doua sunt superioare primelor.

Schema generală a unui sistem de tip întrebare-răspuns. Actualmente, dezvoltarea internetului și succesele câpătate în achiziționarea informației și a tehnicilor de procesare a limbajului natural și cererea pentru acces facil la informație a dus la creșterea interesului pentru sisteme care să ofere răspunsuri din diferite domenii.

Un sistem de răspuns la întrebări bazat pe un set de documente are trei componente principale [1]:

1. Modulul de analiză a întrebării care transformă întrebările formulate în limbaj natural uman în întrebări pentru motorul de acumulare de documente.

2. Modulul de achiziție de articole care caută în colecția de articole relevante pentru întrebarea formulată de utilizator pe baza datelor primite de la modulul de analiză a întrebării.

3. Modulul de extragere a răspunsului extrage, din colecția de articole, un răspuns succint în limbaj natural la întrebarea utilizatorului.

Internetul cuprinde informații din toate domeniile, însă a găsi un răspuns la o întrebare simplă nu e totdeauna ușor. Printre dificultățile ce pot apărea în dezvoltarea unui astfel de sistem întâlnim următoarele:

Formularea corectă a întrebărilor. A transforma o întrebare din limbaj natural într-o întrebare pentru un motor de căutare este o sarcină dificilă. În cazul când întrebarea este prea generală, vom extrage un număr mare de documente. Și subiectele descrise în setul de documente extrase e îndoielnic că vor conține tocmai răspunsul la întrebarea utilizatorului. În cazul când extragem prea multe documente și timpul de procesare este mai îndelungat. Iar dacă extragem un set mic de cuvinte, se poate să nu fie găsit articolul care răspunde la întrebare. De aceea se cere ca setul de cuvinte utilizate în întrebări să fie bine formulat, pentru a fi returnate, pe cât posibil, doar documente ce conțin informații utile.

Un sistem întrebări-răspunsuri caută expresia sau enunțul (pasajul, documentul, setul de documente) care este răspunsul exact la o întrebare:

- întrebările au o multitudine de nuanțe;
- majoritatea căutărilor sunt axate pe întrebări factice;

II. SISTEMUL IDSI

Sistemul ce se crează va fi utilizat la IDSI (Institutul de Dezvoltare a Societății Informaționale). IDSI ne-a prezentat o listă de documente pentru lucru, și anume:

1. L E G E Codul cu privire la știință și inovare al Republicii Moldova
2. H O T Ă R Î R E cu privire la aprobarea Acordului de parteneriat între Guvern și Academia de Știință a Moldovei pentru anii 2009-2012
3. Cu privire la parcurile științifico-tehnologice și incubatoarele de inovare
4. Știință și inovare

Aceste documente au servit la alcătuirea întrebărilor pentru sistemul nostru. Etapa următoare este clasificarea întrebărilor din punct de vedere semantic.

III. TIPOLOGIA ÎNTREBĂRILOR

Tipurile lor pot fi mai generale sau foarte detaliate, în dependență de scopul pe care și-l fixează cercetătorul: întrebări factuale, definiții, opinii, explicații, etc. Întrebările pot fi modificate. De exemplu, "Din ce constă fondul de informații științifico-tehnologice?", poate avea și o altă variantă, ca: "Explicați din ce constă fondul de informații științifico-tehnologice".

Organizatorii competiției internaționale QA@CLEF [2] au definit 8 tipuri de întrebări majore:

1. Timp (TMP)
2. Interval de timp (ITMP)
3. Definiție (DEF)
4. Măsură (MES)
5. Listă (LST)
6. Locul (LOC)
7. Nume/persoană (N)
8. Explicație/descriere (EXP)

Au fost elaborate 29 tipuri de întrebări:

1. Agentul (N). Exemplu: Cine poate modifica cerințele acreditării?

Variantele întrebărilor de acest tip pot fi : De către cine ..., De care categorii de personal ..., Prin intermediul cui ...,

2. Atribuții (EXP, LST). Exemplu: Ce fel de criterii generale de evaluare a activității organizațiilor din sfera științei și inovării există?

3. Cantitatea (MES). Exemplu: De câte ori a fost modificată legea cu privire la știință și inovare al RM?

Variantele întrebărilor de acest tip pot fi : Câte ..., Cât..., Câți...

4. Condiții (EXP). Exemplu: Care sunt condițiile principale de eligibilitate a organizațiilor-executor ai proiectelor?

Variantele : În ce caz ...

5. Consecință (EXP, DA/NU) . Exemplu: Credeți că consecințele neacreditării pot fi grave?

Variantele. Exemplu: Care pot fi consecințele neacreditării?

6. Componenta (LST). Exemplu: 1.Ce intră în resursele tehnice și infrastructurile disponibile?

Variantele : Ce informație cuprinde..., Ce conține..., Ce include..., Din care puncte este compus ..., Care este componenta..., Care ... intră în componenta ..., Ce include..., Din ce se formează ...

7. Cunoștințe. (EXP) Exemplu:1.Ce știți despre direcțiile strategice ale activității din sfera științei și inovării?

8. Da-Nu. (EXP) Exemplu: 1.Funcțiile autorităților publice sunt unice în promovarea politicii de stat în sfera științei și inovării?

Variantele : Sunt oare garanții ..., ..., sunt ele severe, grele, exagerate..., Vă pare că consecințele neacreditării pot fi grave..., Credeți că consecințele..., Cunoașteți care sunt principalele funcții...

9. Date. (TMP) Exemplu:1.Când a fost publicată legea cu privire la știință și inovare al RM?

Variantele : 2.Când ..., În ce ani ...

10. Definiții. (DEF) Exemplu 1.Cum definim banca de date științifico-tehnologice?

Variantele : Care este definiția..., Cum definim noțiunea de..., Ce înseamnă ..., Cum se descifrează... Cum definiți noțiunea de Academie de Științe?

11. Destinatar (N). Exemplu : 1.Cui trebuie să fie restituite proiectele și fișele de evaluare?

Variantele : Cui îi revine responsabilitatea pentru... Cum se numeşte...

12. Durata (TMP). Exemplu : 1.Pentru care ani a fost încheiat acordul de parteneriat între Guvern şi Academia de Ştiinţe a Moldovei?

13. Enumărări (LST). Exemplu :1.Care sunt noţiunile principale ale legii cu privire la parcurile ştiinţifico-tehnologice şi incubatoarele de inovare?

Variantele : Ce conţine ..., Din ce constă..., Numiţi..., Ce..., Care etape..., Care sunt funcţiile..., Care sunt funcţiile..., Care sunt obligaţiile..., Ce date conţine..., Ce reprezintă..., Câte puncte conţine...

14. Explicaţii (EXP). Exemplu :1.Care este politica de stat în sfera ştiinţei şi inovării?

Variantele : Cum explicaţi ..., Ce domenii cuprinde..., În ce constă ..., Prin ce se deosebeşte ..., În ce constă...,

15. Fapţice (DEF). Exemplu :1.Cum se numeşte activitatea orientată spre dobândirea de noi cunoştinţe ştiinţifice spre formularea şi verificarea de noi ipoteze şi teorii?

Variantele : Ce este o cercetare aplicată..., Cum se numeşte ..., În baza cărui act este constituită..., Care este activitatea..., Cine este finanţat de la bugetul de stat prin concurs..., Care sunt obiectivele ..., Cine poate deveni..., De care bunuri poate dispune ..., Ce este produsul informaţional, resursele, sistemul ştiinţifico-tehnologice, Care sunt serviciile informaţionale, Care sunt funcţiile..., Care domenii din viaţa socială sunt o prioritate strategică în politica de stat a Moldovei, Care este politica de stat în sfera ştiinţei şi inovării, Ce proiecte din sfera ştiinţei şi inovării conţine articolul 60, Care obiective conţine articolul 62, Cine are atribuţii, Pe cine întrupeşte Academia de Ştiinţe, Care este organul suprem de conducere al Academiei de Ştiinţe, Care sunt atribuţiile..., De cine este atestat..., Care sunt împuternicirile..., În ce constă funcţiile agenţiei de stat pentru proprietatea intelectuală, Care sunt..., Cine face parte din ..., Care este ..., Care-i repartizarea alocaţiilor prevăzute în bugetul de stat pe anul 2009 pentru programele de stat, Care a fost planul de admitere la studii pentru doctorat şi postdoctorat (cu finanţare bugetară) pentru anul de studii 2009-2010, Ce informaţie conţine fişa biografică, Ce se indică în experienţa ştiinţifică şi profesională, Ce conţine o declaraţie, Care este activitatea expertului independent, Care sunt funcţiile şi atribuţiile comisiei de acreditare,

16.Frecvenţa. (ITMP) Exemplu :1.De câte ori pe an se convoacă în şedinţă comisia de acreditare?

17.Indicaţii (EXP). Exemplu :1.Fă modelul planului calendaristic de realizare a proiectului.

Variantele : Indică..., Formează CV ...

18. Liste (LST). Exemplu :1.Cine poate deveni membru de onoare al Academiei de Ştiinţe?

Variantele :Indică lista ..., Indică punctele..., Prezentaţi lista...

19. Locul (LOC). Exemplu :1.Unde pot fi plasate materialele transmise AGEPI?

Variantele : Unde a fost publicată legea cu privire la ştiinţă şi inovare al RM, În ce număr al Monitorului a fost publicată..., În care articol se vorbeşte despre statutul juridic al personalului aparatului administrativ, Unde sunt fixate ..., În care articol se descriu ..., În care număr al Monitorului Oficial a fost publicată hotărârea cu privire la

aprobarea Acordului de parteneriat între Guvern şi Academia de Ştiinţe pentru anii 2009-2012

20. Modalitatea (EXP). Exemplu : 1.În ce mod se realizează implementarea rezultatelor ştiinţifice, invenţiilor, noilor tehnologii, utilaje, echipamente, soiuri, hibridi, preparate farmaceutice şi a altor obiecte de proprietate intelectuală?

Variantele :Cum se formează piaţa de produse informaţionale ştiinţifico-tehnologice, Cum s-a format complexul patrimonial al Academiei de Ştiinţe, Cum se efectuează ..., Cum sunt finanţate ..., Cum se soluţionează litigiile apărute în legătură cu înregistrarea rezultatelor cercetărilor ştiinţifice.

21. Motivul (EXP). Exemplu : 1.De ce avem nevoie de un transfer tehnologic?

Variantele :De ce trebuie să respectăm legile

22. Numele (N). 1.Cum se numeşte documentul Uniunii Europene pentru finanţarea cercetării şi dezvoltării pe perioada 2007-2013, adoptat prin decizia nr. 1982/206 EC a Parlamentului European şi Consiliului European, publicată în Official Journal al UE din 30 decembrie 2006?

Variantele : Cum se numeşte regulamentul

23. Obiectul (N). Exemplu : 1.Despre ce este vorba în articolul 67?

Variantele : Despre ce este vorba în art.74 lit. (h) introdusă prin LP192 din 30.06.06, Despre care politică este vorba în articolul 116, Care obligaţiuni principale îşi asumă Academia de Ştiinţe, De ce este responsabil guvernul, Care au fost organizaţiile din sfera ştiinţei şi inovării şi a instituţiilor de învăşământ ale căror terenuri agricole vor fi folosite în anul 2009 în scopuri ştiinţifice şi instructive, Ce conţine anexa nr.2, Ce include devizul, Ce conţine contractul încheiat între părţi, Care categorii de activităţi pot servi drept obiect al programelor şi proiectelor de inovare şi transfer tehnologic

24. Opinii (EXP). Exemplu : 1.Care domenii din viaţa socială sunt o prioritate strategică în politica de stat a Moldovei?

Variantele : Cine ştie care sunt principalele direcţii ale activităţii Academiei de Ştiinţe, Ce reprezintă resursele de informaţii ştiinţifico-tehnologice

25.Proces (EXP). Exemplu : 1.Cum se descifrează sigle CCTM?

Variantele :Cum se soluţionează litigiile apărute în legătură cu înregistrarea rezultatelor cercetărilor ştiinţifice, Cum este ales preşedintele Academiei de Ştiinţe, Cum evoluează procesul de acreditare

26.Responsabilitatea (N): Exemplu : 1.Cine sunt responsabilii de valorificarea eficientă şi conform destinaţiei a mijloacelor bugetare alocate?

Variantele : Cui îi revine responsabilitatea pentru..., Ce responsabilitate are organizaţia-executor

27. Scopul (EXP). Exemplu :1.Cu ce scop este fondată organizaţia de drept privat din sfera ştiinţei şi inovării?

Variantele : Numiţi scopul şi sarcinile politicii de stat în sfera ştiinţei şi inovării, Cu ce scop se efectuează..., Cu ce scop se foloseşte ..., Cu ce scop a fost creată ..., În ce scop sunt utilizate ..., Care este scopul organizării şi efectuării expertizei

28. Subiectul (N). Exemplu :1.Care sunt obiectivele principale din cadrul direcţiilor strategice ale activităţii din sfera ştiinţei şi inovării pentru anii 2006-2010 aprobate prin hotărârea parlamentului nr. 160-XVI din 21 iulie 2005?

29. Surse (EXP). Exemplu :1.Din ce mijloace se face asigurarea financiară a activității de acreditare?

Variantele: Reorganizarea sau dezvoltarea organizațiilor de drept public din sfera științei și inovării se efectuează în modul stabilit de legislația în vigoare sau prin alte mijloace?

IV. ANALIZA ÎNTREBĂRILOR

Interpretorul de întrebări este o unitate importantă a sistemului întrebare-răspuns care efectuează următoarele acțiuni:

- analiza lingvistică a întrebării;
- definirea tipului întrebării în raport cu o tipologie predefinită;
- determinarea tipului de răspuns așteptat;
- generarea unei cereri de căutare în limbajul de specificare al motorului de căutare utilizat.

Analiza lingvistică presupune determinarea grupurilor nominale, verbale, prepoziționale ale întrebării; definirea legăturilor dintre ele. De exemplu:

(grup prepozițional: În care număr al Monitorului Oficial) (grup verbal: a fost publicată) (grup nominal:hotărârea cu privire la aprobarea Acordului de parteneriat între Guvern și Academia de Științe pentru anii 2009-2012)?

Ultimul grup nominal conține grupuri prepoziționale: “la aprobarea Acordului”, “de parteneriat”, “între Guvern și Academia de Științe”, “pentru anii 2009-2012” care sunt atașate unul la altul și formează un enunț întreg.

Definirea tipului întrebării se face prin compararea cu modelele prezentate în punctul III. De exemplu, întrebarea: **Unde pot fi plasate materialele transmise AGEPI?** începe cu adverbul de loc “Unde” care indică locul, iar întrebarea **De câte ori pe an se convoacă în ședință comisia de acreditare?** începe cu grupul “De câte ori” ce indică frecvența. Astfel, prima parte a întrebării este indicatorul tipului întrebării.

Tipul de răspuns așteptat se determină în conformitate cu tipul întrebării.

Generarea cererii de căutare în limbajul de specificare al motorului de căutare este cea mai dificilă sarcină și depinde de sursa în care va fi căutat răspunsul și motorul de căutare utilizat. În cazul nostru răspunsurile sunt căutate în documentele specificate. Aceasta presupune căutare în text utilizând metodologia cuvintelor și frazelor – cheie. De exemplu, din întrebarea **Din ce mijloace se face asigurarea financiară a activității de acreditare?** sunt extrase următoarele cuvinte-cheie: **mijloace, asigurarea financiară, activității de acreditare.** Cuvintele cheie pot păstra forma din întrebare sau pot fi schimbate conform formei inițiale din dicționare. În cazul grupului nominal numai primul cuvânt este schimbat. De exemplu: **mijloc, asigurare financiară, activitatea de acreditare.** În tabelul 1 sunt prezentate exemple de întrebări, elemente ce indică tipurile lor și cuvintele-cheie extrase pentru motorul de căutare.

TABEL 1. EXEMPLE DE ÎNTREBĂRI, ELEMENTE CE INDICĂ TIPURILE LOR ȘI CUVINTELE-CHEIE EXTRASE PENTRU MOTORUL DE CĂUTARE

Întrebare	Tipul	Cuvinte-cheie
Pentru care ani a fost încheiat acordul de parteneriat între Guvern și Academia de Științe a Moldovei?	Tipul:Durata (ITMP) Indicatorul : Pentru care ani	ani, acordul, parteneriat, Guvern, Academia de Științe a Moldovei
Cine poate modifica cerințele acreditării?	Tipul: Agentul (N) Indicatorul : Cine	poate modifica, cerințele, acreditarea
Ce informație cuprinde fișa biografică?	Tipul: Componenta: (LST) Indicatorul : Ce informație	informație, fișa biografică
Când a fost modificat articolul 117 al (5)?	Tipul: Timp (TMP) Indicatorul : Când	modificat, articolul 117 al (5)
Cum explicați noțiunea de monitorizare?	Tipul: Explicație/descriere (EXP) Indicatorul : Cum explicați	explica, noțiunea de monitorizare
Ce este o cercetare aplicată?	Tipul: Definiție (DEF) Indicatorul : Ce este	cercetare aplicată
În ce articol se descriu tipurile și structura organizațiilor din sfera științei și inovării?	Tipul: Loc (LOC) Indicatorul : În ce articol	descrie, tipuri, structura, organizație, sfera științei și inovării
Câte criterii generale de evaluare a activității organizației din sfera științei și inovării există?	Tipul: Măsură/cantitate (MES) Indicatorul : Câte	criterii generale de evaluare, activitatea organizației, sfera științei și inovării

V. CONCLUZIE

În articolul dat se descrie clasificarea întrebărilor pentru sistemul de tip întrebare-răspuns ce se crează în baza documentelor IDSİ. Au fost elaborate 29 tipuri de întrebări care se includ în 8 clase majore propuse la nivel internațional. Majoritatea întrebărilor au semantică concretă și se clasifică comparativ ușor iar unele întrebări au semantică ambiguă și pot fi incluse în doua sau mai multe clase. Aceste întrebări necesită o analiză mai detaliată. Etapa următoare descrisă este analiza întrebărilor în scopul pregătirii datelor necesare pentru motorul de căutare care va extrage răspunsul din textele documentelor. În cadrul acestei analize se determină tipul întrebării ce definește tipul răspunsului și cuvintele-cheie care sunt transferate motorului de căutare cu scopul găsirii răspunsului.

REFERINȚE

- [1] Iftene, A., Pistol, I., C., Forăscu, C., Trandabăț, D., Balahur-Dobrescu, A., Cotelea, D., Drăghici, I. Construirea unui sistem de tip Întrebare-Răspuns pentru limba română, în *Lucrările atelierului de lucru Resurse lingvistice și instrumente pentru prelucrarea limbii române*, 2007.
- [2] Tufiș, D., Ștefănescu, D., Ion, R., Ceaușu, A. RACAI's Question Answering System at QA@CLEF 2007, CLEF2007 Workshop, 15 pp., September, 2007.