

Regăsirea informațiilor. Aspecte generale

Lector univ. dr. Rodica Mandeal

Abstract: *The article is an attempt to review how does information retrieval work. The base model of the retrieval process including the User sequence- the user's need for information, formulating a search strategy and a query- and the Database producer sequence- selection of the information sources, creation of the content of individuals records, query processing- and also the three primary classes of groups involvd in the process of information retrieval are discussed.*

Key-words: *Information Retrieval, Information Retrieval System, Databases Producer, Users, Search Strategy, Query, Information Need.*

Regăsirea informației este în primul rând un proces de selectare. În mod succint, prin sintagma „regăsirea informației” înțelegem identificarea informației dorite dintr-un stoc/colecție de informații iar acest proces conține implicit conceptul selectivității. Deci regăsirea informației este în primul rând un proces de selecție.

Recuperarea informațiilor nu înseamnă același lucru cu regăsirea lor. Activități ca urmărirea unui program de știri la televizor sau copierea unui fișier nu sunt activități de regăsire, deoarece spectatorul sau cel care copiază fișierul nu are nici un control asupra conținutului informațiilor pe care le va obține.

Regăsirea informației este un proces de comunicare indirectă între autorii unui mesaj, indiferent de forma de prezentare, respectiv între creatorii unor înregistrări (text, imagine, sunet) și utilizatorii informațiilor conținute în acest mesaj.

Mesajele transmise de autori ajung la utilizatori după o perioadă relativ lungă de timp; ele vor trebui, din această cauză, să fie prelucrate pentru o utilizare ulterioară fără a se cunoaște *a priori* cine le va citi și în ce circumstanțe.

Alături de cei doi actori principali ai acestui proces de comunicare – *autorii și utilizatorii* – sunt implicați direct sau indirect cei care decid ce aspecte ale mesajelor vor fi reprezentate în înregistrări, cei care prelucrează informația pentru a crea înregistrarea, cei care ajută utilizatorul să desfășoare o cercetare, cei care proiectează și implementează programele de calculator, precum și cei care decid ce servicii vor oferite utilizatorilor și în ce condiții.

Cel mai bun exemplu de organizație/instituție concepută pentru a regăsi informația este un centru de informare (bibliotecă sau centru de documentare).

Nimeni nu intră într-o bibliotecă pentru a citi întreaga colecție de documente, ci doar pentru a căuta ceva selectiv care de cele mai multe ori va satisface un set de nevoi de informare foarte personalizat.

Sintagma *regăsirea informației*, așa cum a fost utilizată prima dată în anul 1957, în Anglia era sinonimă cu *cercetarea literaturii științifice*, respectiv a fost definită ca procesul de căutare într-o colecție de documente (termenul document fiind acceptat în sensul cel mai larg) pentru a identifica pe cele care tratează un anumit subiect sau răspund la o anumită întrebare; în consecință orice sistem proiectat să faciliteze această activitate poate fi numit sistem de regăsire a informației.

Regăsirea informației nu este o activitate legată strict de calculator. Principiile ei s-au aplicat cu mult înainte de a fi fost inventate calculatoarele și astăzi, uneori, această activitate se desfășoară fără asistența lor. Căutarea într-un index tipărit, într-un catalog tradițional sau într-o enciclopedie pot fi considerate tehnologii mai vechi ale aceleiași activități.

Totuși sintagma „regăsirea informației” este în prezent utilizată pentru a descrie procesul căutării și identificării informațiilor care au anumite atribute într-o bază de date, proces asistat de calculator, și desfășurat într-un mod interactiv. Deoarece inițial fișierele bazelor de date conțineau înregistrări descriptive ale documentelor – text, până în anii '90, problema centrală a procesului de regăsire a informației era considerată cea a modului în care documentele erau reprezentate pentru a fi regăsite (respectiv problema indexării și a vocabularului controlat).

O dată cu apariția bazelor de date conținând textul complet al documentelor, precum și a utilizării crescute a limbajului natural în ecuațiile de căutare, punctul cheie al procesului a devenit compararea sau punerea în relație a cererii de informare cu textul stocat în baza de date.

Totuși, dacă considerăm că ecuația de căutare nu este altceva decât reprezentarea documentului pe care dorim să-l regăsim, putem să apreciem că formularea acesteia rămâne o chestiune de importanță majoră în obținerea rezultatelor dorite de utilizator.

Dificultățile pe care le implică procesul de regăsire se datorează de fapt, încercării de a compara nevoile de informare ale utilizatorilor cu mesajele pe care autorii doresc să le transmită.

Majoritatea mesajelor sunt transferate sub forma unor texte, iar nevoile de informare sunt reprezentate prin cereri de informare.

Procesul căutării selective a informației într-o bază de date implică două abordări diferite:

a) Un punct de plecare este *utilizatorul* sau potențialul utilizator al informației, adică o persoană care are nevoie de informații și care utilizează un sistem de regăsire a informației pentru a-și satisface această nevoie.

b) La celălalt capăt, o organizație, *producătorul* bazei de date, de exemplu, care ia decizia colectării, organizării și stocării informației pentru o regăsire ulterioară.

Producătorul bazei de date creează, în acest scop, înregistrări-reprezentări ale textelor/documentelor (care pot fi reproduceri exacte, părți din texte sau alte forme construite de oameni sau computer), le stochează și pune la dispoziție dispozitive, instrumente care permit căutarea în baza de date. În cazul informației electronice „dispozitivul” este un set de programe.

Producătorul creează, de asemenea, și reprezentări ale cererilor de informare înaintate de utilizatori pe care le compară apoi cu înregistrările stocate în baza de date, regăsind acele reprezentări ale documentelor care corespund reprezentării cererilor.

Problema care apare de regulă constă în faptul că documentele pot să nu fie reprezentări perfecte ale mesajelor (ceea ce deși este o problemă de comunicare nu intră sub controlul unui sistem de regăsire a informației), iar reprezentările acestor documente pot să nu reflecte cu exactitate conținutul documentelor (de exemplu, din cauza indexatorilor sau a limbajelor de indexare).

În mod identic cererile de informare sunt rareori reprezentări fidele ale nevoilor de informare, iar ecuațiile de căutare pot să nu reflecte univoc și precis, cererile de informare.

Etapale procesului de regăsire a informației

Începutul oricărui proces de regăsire a informației, dar și succesul acestui proces pleacă de la recunoașterea de către utilizator a unei nevoi de informare, adică lipsa unor date, informații, cunoaștere.

Pentru a-și satisface această nevoie, utilizatorul face *al doilea pas* și își exprimă această nevoie printr-o cerere de informare, cu ajutorul unui intermediar – un specialist în informare- (într-o anumită măsură această funcție poate fi asigurată de un program de calculator).

Exprimarea nevoilor de informare este indicat să se facă într-un limbaj pe care atât intermediarul, cât și utilizatorul îl cunosc – deci în limbaj natural, pe cât posibil neîncărcat de jargoane tehnice și nu în coduri, descriptori sau cuvinte cheie.

Dacă această etapă de comunicare interumană nu decurge bine întregul proces de regăsire va fi compromis.

A treia etapă a procesului de regăsire constă în elaborarea unei strategii de căutare, adică în stabilirea unui plan general de regăsire a

informației dorite. Strategia de cercetare poate fi elaborată la diferite nivele, însă foarte mulți specialiști subliniază importanța aspectelor explicite și detaliate ale strategiei de căutare.

Următoarea etapă, formularea ecuației de căutare, constă în exprimarea precisă a nevoii de informare în termenii specifici sistemului de regăsire. Există două aspecte importante ale acestei etape:

- a) atribuirea de termeni, cuvinte, coduri, utilizate deja în descrierea înregistrărilor existente în baza de date și pe care utilizatorul dorește să le regăsească;
- b) legarea logică a diferitelor seturi de valori.

Formularea ecuației de căutare constă în alegerea termenilor de indexare, reprezentând conținutul cererii de informare și corelarea lor logică sau, o dată cu dezvoltarea programelor de interogare, poate reprezenta conținutul real și nu „tradus” al documentului pe care dorim să-l regăsim.

În concluzie, este important să reținem că atunci când ne referim la regăsirea informației electronice, calculatorul procesează ecuația de căutare și nu exprimarea nevoii de informare (adică cererea de informare) sau nevoia reală de informare.

În ceea ce privește *ieșirile dintr-un sistem* de regăsire a informației acestea pot fi diferite: referințe bibliografice (titlul, autorul, sursa, de exemplu) însoțite sau nu de un rezumat sau o listă de termeni prin care a fost indexat documentul; informații factice sau așa numitele referințe rapide, date numerice care dau răspunsul direct la o anumită întrebare; textul întreg al unui document sau paragrafe ale acestui text în care apar combinații ale unor cuvinte (frecvent întâlnite în domeniul juridic).

Utilizatorii unui sistem de regăsire a informației

Deși în principiu, oricine poate utiliza un sistem de regăsire a informației, există 3 categorii de persoane care se implică în procesul regăsirii: specialiștii în informare, utilizatorii finali și utilizatorii începători.

a) Specialiștii în informare

Un specialist în informare în această ipostază, cea de utilizator, este o persoană care oferă un serviciu de informare, respectiv îi asistă pe alții să găsească informația de care au nevoie. Multă vreme această profesie a fost identificată cu cea de intermediar obligatoriu între utilizatori și sistemele de regăsire. În această categorie intră și brokerii de informații, denumire care derivă din faptul că acești specialiști nu creează informații pentru clienții proprii, ci doar le găsesc și eventual le interpretează.

Indiferent de rolul pe care îl joacă în procesul regăsirii informației acest tip de utilizator are, în general, o cunoaștere asupra:

- tipurilor de informații accesibile în disciplina sau specialitatea în care lucrează (drept, medicină, finanțe etc) și a organismelor de informare care le oferă;

- terminologiei specifice domeniului pentru care oferă asistență;
- metodelor de comunicare/dialog cu cel căruia îi oferă serviciul pentru a evidenția nevoia de informare și a o exprima;
- procedurilor de utilizare a sistemelor de regăsire relevante pentru satisfacerea acestor nevoi.

b) Utilizatorii finali.

Inițial sistemele de regăsire a informației au oferit servicii utilizatorilor doar printr-un intermediar.

O dată cu răspândirea largă a calculatoarelor personale, utilizatorii finali au început să devină importanți pe această piață din două motive:

- PC-urile permit utilizatorilor finali să utilizeze direct serviciile on-line de regăsire a informației;

- Centrele de informare oferă servicii de regăsire on-line din ce în ce mai diversificate, piața fiind practic saturată la sfârșitul anilor 90 (evident ne referim la țările dezvoltate).

O primă caracteristică a acestor noi veniți pe piața serviciilor de informare on-line este aceea că fac parte din profesii, discipline sau industrii foarte dependente de informație (arhitecți, medici, avocați, oameni de afaceri, oameni de știință sau tehnologi, jurnaliști).

Membrii unor astfel de grupuri profesionale au abilități în interpretarea informației din propriul domeniu, de care mulți intermediari sunt lipsiți și anume recunoașterea imediată sau evaluarea unei informații noi, precum și capacitatea de a face asociații surprinzătoare între nevoile lor de informare și informații aparent fără nici o legătură.

O altă caracteristică este aceea că nu toți utilizatorii finali au experiența lucrului cu calculatorul și cunoașterea mecanismelor de căutare și regăsire a informației.

c) Utilizatorii – Începători.

Acest grup constă în principal din elevi sau studenți în primii ani de studii care nu au grad ridicat de cunoaștere a unui domeniu, anume, a mecanismelor de căutare și probabil nici competențe pentru a face evaluări ale informației regăsite.

Componentele unui sistem de regăsire a informației

Sistemele de informare reprezintă un complex de metode și echipamente și programe pentru prelucrarea și înmagazinarea informațiilor în scopul regăsirii lor ulterioare. Obiectivul unui sistem este deci regăsirea informațiilor solicitate de utilizatori în legătură cu diferite subiecte, mai

specifice sau mai generice, dintr-un anumit domeniu, sau, cu alte cuvinte, de a efectua cercetări ale literaturii. În urma efectuării unei asemenea cercetări, sistemul pune la dispoziția utilizatorilor care au adresat cereri de informare, o listă de adrese (referințe bibliografice) ale documentelor de interes probabil pentru tema solicitată. Există și sisteme care furnizează chiar documentele ca atare sau diverse forme prescurtate ale acestora (referate) sau informațiile extrase din textul documentelor.

În majoritatea acestor sisteme, informațiile sunt analizate și prelucrate pe principiul indexării coordonate, în care baza intelectuală o reprezintă bagajul termenilor de specialitate ce constituie terminologia de indexare și regăsire. Dat fiind faptul că folosirea limbajului de specialitate așa cum este el utilizat de autori în documente (limbaj natural) ridică multe probleme sistemele folosesc un limbaj normalizat (controlat) pentru a indexa documentele și a formula ecuațiile de căutare.

Principalele componente ale unui sistem de regăsire a informațiilor sunt:

- subsistemul de selecție a documentelor;
- subsistemul de indexare;
- subsistemul vocabularului;
- subsistemul de căutare;
- subsistemul de interacțiune între utilizator și sistem (interfața utilizator-sistem);
- subsistemul de comparare, adică subsistemul care compară reprezentările documentelor stocate cu reprezentările cererilor de informare.

În anumite privințe subsistemul de comparare poate fi considerat cel mai puțin important dintre cele șase subsisteme deoarece nu are o influență directă asupra eficacității sistemului respectiv, asupra capacității de regăsire a documentelor care satisfac nevoile de informare ale utilizatorilor. Cu toate acestea, eficiența subsistemului de comparare exercită o influență relativ ridicată asupra aspectelor economice și asupra eficienței generale ale sistemului, măsurate spre exemplu, în timpul de răspuns. În majoritatea sistemelor, calculatorul nu contribuie în mod direct la selectarea documentelor, la controlul selecției, la indexarea și căutarea documentelor, la pregătirea strategiilor de căutare sau la interacțiunea cu utilizatorii sistemului – cum ar fi spre exemplu, în cazul „negocierii cererii”. Acestea sunt activități intelectuale, care intră în responsabilitatea specialiștilor și care determină eficacitatea sistemului.

Factorii cei mai importanți care influențează eficiența unui sistem de regăsire a informațiilor pot fi împărțiți în două grupe principale:

- factori legați de crearea și actualizarea bazei de date;
- factori legați de exploatarea bazei de date;

Principalii factori legați de crearea bazei de date, care pot fi considerați și „*factori de intrare*” sunt:

- Politicile și criteriile de selecție a documentelor ce urmează a fi incluse, care implică o bună cunoaștere a nevoilor de informare ale utilizatorilor;
- Reprezentarea completă și precisă, în operația de indexare, a conținutului acestor documente;
- Vocabularul sistemului, adecvat pentru a reprezenta subiectele tratate de aceste documente.

Există, de asemenea, „*factori de ieșire*” sau de exploatare a bazei de date:

- Personalul sistemului de regăsire care trebuie să înțeleagă nevoile de informare ale utilizatorilor (interacțiune utilizator-sistem).
- Strategiile de căutare care trebuie să reflecte cât mai bine aceste nevoi;
- Vocabularul sistemului adecvat, pentru a reprezenta interesele utilizatorilor sistemului.

Toți acești factori contribuie direct la eficiența sistemelor de regăsire a informațiilor. De asemenea, ei sunt definatorii într-o anumită măsură și pentru sistemele de informare propriu-zise. De multe ori, în literatura de specialitate sistemele de regăsire sunt considerate ca „motorul” sistemelor de informare.

BIBLIOGRAFIE

- BANCIU, D.; MANDEAL, R.** Sisteme de informare – Sisteme de regăsire a informației. În: *Studii de bibliologie și știința informării*, vol. 3 (1997). București, Editura Universității din București, 2001, p.18-20.
- CHAUMIER, J.** *Les techniques documentaires*. Paris: PUF, 1992.
- GUINCHAT, C.; MENO, M.** *Introduction generale aux sciences et techniques de l'information et de la documentation*. Paris: UNESCO, 1990.
- LANCASTER, F. W.; WARNER, A. J.** *Information Retrieval Today*. Arlington: Information Resources Press, 1993.
- Le COADIC, Y. F.** *La science de l'information*. Paris: PUF, 1994.
- Le COADIC, Y. F.** *Le besoin d'information: formulation, négociation, diagnostique*. Paris: ADBS Editions, 1998.
- MANDEAL, R.** Studiul utilizatorilor de informații. În: *Studii de Bibliologie și Știința Informării*, vol. 3 (1997), p. 25-33.
- MANDEAL, R.** Utilizatorul și căutarea informației. În: *Buletinul ABIR*, vol. 13, nr. 2 (2002), p. 10-14.
- MOLES, A. A.** *Théorie structurale de la communication et société*. Paris: Masson, 1998.