

Arhivarea electronica, componenta de baza a unui sistem de gestiune automata a documentelor

Romeo Mihai PAVELESCU

Ministerul Apararii Nationale, RRomica@yahoo.com

Evolutia spectaculoasa a tehnologiei informatiei din ultimul deceniu ridica stacheta în privinta accesului rapid si facil la o informatie corecta si de actualitate. Pentru realizarea acestui deziderat este necesara implementarea în cadrul firmei sau institutiei respective a unui sistem de gestiune si arhivare electronica a documentelor. Arhivarea electronica a documentelor poate fi vazuta ca parte componenta a unui sistem de gestiune automata a documentelor dar poate constitui si o activitate de sine statatoare, mai ales în cadrul institutiilor publice detinatoare de arhive impresionante ca volum si valoare a continutului.

Datele si informatiile existente în arhive se afla sub cele mai diverse forme, cum ar fi hârtie, fisiere electronice, înregistrari audio, clipuri video, microfilme. Preluarea acestora într-un sistem de prelucrare automata a datelor presupune, în principal, scanarea, indexarea, stocarea lor pe suporti de memorie de mare capacitate, caracterizati de o fiabilitate foarte ridicata precum si regasirea lor ulterioara dupa diferite criterii.

Cuvinte cheie: arhivare electronica, documente electronice, sistem de gestiune automata a documentelor.

Aspecte privind arhivarea clasica a documentelor

Orice arhiva manuala este formata pe de-o parte din magazia de documente organizata în asa fel încât sa se poata ajunge la fel de usor la orice document iar pe de alta parte din cataloage sau cartoteci necesare pentru cautarea dupa diferite criterii (titlu/denumire, subiect, autor, domeniu) în arhiva. Deci, cartotecile contin referinte catre documentele arhivate.

Operatiile care se efectueaza la nivelul unei arhive manuale sunt:

- **intrarile** în arhiva, presupun, pe lângă gasirea locului potrivit pentru noile documente, si actualizarea tuturor cartotecilor existente.
- **iesirile** din arhiva necesita ca si intrarile actualizarea cartotecilor existente;
- **microfilmarea** documentelor de hârtie în vederea micsorarii spatiului pe care acestea îl ocupa;
- **întretinerea/reconditionarea**

documentelor este o operatie foarte importanta deoarece, pe masura trecerii timpului, documentele aflate pe hârtie sau film pierd din calitate, ajungând sa fie de neutilizat;

Analiza unui sistem de arhivare electronica a documentelor trebuie sa înceapa din perspectiva existentei unui sistem de gestiune a documentelor în care primul trebuie sa se integreze. Astfel, arhiva electronica trebuie sa fie capabila sa preia documentele finalizate (aprobate si exploatate) împreuna cu datele si informatiile adiacente (despre circulatia documentului, rezolutii, etc.) si sa le organizeze de asa fel încât accesul la document sa fie rapid si facil.

Totodata, S.A.E.D. trebuie sa permita arhivarea documentelor aflate pe hârtie sau microfilm si indexarea acestora în vederea interogarilor ulterioare.

Structura unui sistem de arhivare electronica a documentelor (S.A.E.D.)

Datorita faptului ca cea mai mare parte a documentelor apar în contextul rezolvarii anumitor probleme sau în scopul atingerii anumitor obiective, este recomandabila gruparea documentelor în functie de problema rezolvata sau obiectivul urmarit. Din aceste considerente se recomanda organizarea documentelor în dosare electronice.

Din punct de vedere al organizarii datelor si informatiilor, arhiva trebuie structurata astfel:

- **baza de date**, care contine referinte la documente, clasificari ale documentelor, dosare electronice, circulatia dosarelor etc.
- **magazia de documente**, care reprezinta locul unde efectiv se afla documentele arhivate.

Relatia dintre un dosar si un document este supusa urmatoarelor reguli:

- un document poate fi arhivat doar la un singur dosar;
- în cadrul unui dosar un document nu poate aparea decât o singura data.

Drept urmare, din punct de vedere al arhivarii, un document poate fi în una din urmatoarele stari:

- **în lucru** (creat, în curs de aprobare, aprobat, exploatat), specific sistemului de rutare a documentelor;
- **arhivat temporar** – documentul a fost inclus într-un dosar însa dosarul este de actualitate;
- **arhivat definitiv** – problematica dosarului nu mai este de actualitate si dosarul este arhivat împreuna cu documentele pe care le contine. De obicei, arhivarea definitiva presupune intrarea dosarului în arhiva centrala a organizatiei.

Baza de date a arhivei contine informatii referitoare la dosare electronice existente în evidenta filialei/sediului central al organizatiei, documentele pe care acestea le contin si date despre discurile pe care se afla dosarele arhivate definitive.

Magazia temporara reprezinta o structura de subdirectoare în care se afla documentele electronice.

Arhivarea poate fi efectuata pentru orice document electronic care a fost aprobat si exploatat. Dosarele electronice sunt constituite din documente electronice al caror continut trateaza un subiect comun sau o tematica comuna. Aceasta structurare a unei arhive, chiar si electronice, permite o mai buna interogare ulterioara a bazei de documente arhivate.

Arhiva temporara este constituita din multimea dosarelor (organizate pe prob-

leme, activitati, obiective, etc.) care au caracter de actualitate. Operatia de arhivare temporara se refera la documente si are ca efect înregistrarea documentului în cadrul dosarului, alocându-i un numar curent care nu va mai putea fi alocat în cadrul dosarului. Prin intermediul aplicatiei de arhivare, orice lucrator își poate deschide, transfera, anula si arhiva definitiv dosare electronice.

Deschiderea unui dosar are ca efect adaugarea unei înregistrari în nomenclatorul de dosare electronice si totodata evidentiarea lui în contul unui utilizator. Astfel, numai acest utilizator poate arhiva documente în cadrul dosarului. La deschiderea dosarului trebuie sa se ataseze acestuia un numar unic în întreaga organizatie, pentru ca acesta sa poata circula fara a produce confuzii. Acces la acest dosar vor avea proprietarul dosarului si sefi acestuia.

Anularea dosarelor electronice are ca efect stergerea documentelor pe care le contine, ramânând totusi evidentiat faptul ca la o anumita data si o anumita ora s-a procedat la anularea unui anumit dosar, care va ramâne în nomenclatorul dosarelor electronice la ANULAT.

Arhivarea definitiva a dosarelor electronice reprezinta operatia prin care un dosar nu mai poate fi modificat si poate fi migrat pe suporti de memorie care au rezistenta la intemperii si o fiabilitate sporita. Operatia inversa (dezarhivarea) poate fi efectuata, însa aceasta trebuie utilizata numai în cazuri exceptionale.

Transferul unui dosar poate fi efectuat :

- în cadrul aceleiasi filiale (de fapt se schimba proprietarul);
- între filiale (este o operatie care implica scoaterea dosarului din evidenta filialei, transmiterea sub forma de e-mail la filiala destinatie/sediu central si înregistrarea dosarului în evidenta filialei destinatie si în contul lucratorului destinatie care devine noul proprietar).

Baza de date a arhivei trebuie sa permita evidentiarea circulatiei unui dosar si a operatiilor efectuate asupra acestuia de-a lungul timpului.

Deși nu apare ca o parte componentă distinctă a SAED, sistemul de indecsuri al documentelor constituie un element deosebit de important al arhivei.

Deoarece documentele electronice care constituie dosarele pot exista și în format imagine (TIF, JPG, BMP etc.) este necesară aplicarea tehnicilor de recunoaștere a caracterelor din imagini și constituirea cuvintelor obținute în pseudo-documente necesare în procesul de indexare. Astfel, sistemul de căutare a documentelor va fi format dintr-o căutare structurată, bazată pe datele și informațiile din baza de date a arhivei și dintr-o căutare după cuvinte cheie, un rol important având-ul serverul de indexare utilizat.

Modalități de arhivare electronică a documentelor de tip hârtie sau microfilm

Documentele electronice rezultate în urma scanării hârtiei sau microfilmului reprezintă "calcâiul lui Ahile" în cadrul unui sistem de arhivare electronică, din următoarele motive :

- documentele de tip imagine ocupă foarte mult spațiu pe disc;
- acestea nu pot constitui o bază pentru crearea de alte versiuni (nu sunt modificabile);
- nu se poate căuta după cuvinte cheie decât după recunoașterea caracterelor, care este un proces consumator de timp și de resurse și cu un randament destul de scăzut în cazul documentelor vechi;
- scanerile, mai ales cele de film, sunt foarte scumpe.

Totodată, conversia documentelor în variantă electronică prelungește viața acestora datorită fiabilității și rezistenței ridicate la intemperii a discurilor magneto-optice.

Din măsurătorile efectuate asupra procesului de scanare a documentelor aflate pe suport film am constatat o pierdere considerabilă de calitate, de la documentul scanat până la imaginea preluată în calculator. Dacă microfilmul nu are o calitate foarte bună pot apărea zone din

în cadrul imaginii care nu sunt inteligibile chiar dacă pe un cititor de microfilme aceste sunt descifrabile.

Ca urmare a fotografierii de slabă calitate, imaginile scanate de pe microfilme conțin mult "fum" (zone înnegrite datorate supraexpunerii la lumină) care afectează, pe de-o parte, procesul de recunoaștere a caracterelor, în vederea indexării documentelor, iar pe de altă parte mărește considerabil mărimea fișierelor. Astfel, un document ce a trecut mai întâi prin fotografiere și apoi prin scanarea microfilmului poate ocupa un spațiu pe disc de până la 10-15 ori mai mare decât fișierul rezultat în urma scanării directe a documentului de hârtie.

O soluție care poate rezolva parțial problema documentelor scanate de pe microfilm constă în utilizarea de instrumente de procesare a imaginii scanate (eliminarea "fum"-ului, ajustarea contrastului și luminizării pe porțiuni de document), retusarea imaginilor făcându-se în dauna timpului total de conversie a documentului de hârtie în document electronic.

Imaginile obținute din scanarea hârtiei au o calitate foarte bună, iar spațiul ocupat pe disc este acceptabil datorită formatelor grafice (TIF, PDF) ce utilizează algoritmi de compresie puternici.

Integrarea documentelor electronice obținute din scanare într-un sistem de arhivare electronică este deficitară la capitolul de indexare a documentelor datorită faptului că acestea nu conțin cuvinte cheie, ci matrice de pixeli. Rezolvarea acestei probleme constă în construirea unui fișier de cuvinte cheie referitoare la conținutul documentului și atasarea acestuia la documentul scanat într-un mod care să nu permită separarea lor. Fișierul atasat poate fi construit prin editare de către un operator care va stabili cuvintele cheie sau printr-un proces, mai special, de recunoaștere optică a caracterelor (O.C.R.). Specializarea O.C.R.-ului constă în includerea în fișierul atasat doar a cuvintelor cheie care se regăsesc într-un anumit dicționar de cuvinte. Astfel, O.C.R.-ul trebuie să inte-

greze câte un dicționar pentru fiecare limba pe care este destinat să o recunoască.

Dacă regulile enunțate mai sus vor fi respectate, orice motor de indexare a documentelor va putea permite regăsirea documentelor grafice cu un randament rezonabil.

Datorită costurilor ridicate pe care le implică achiziționarea unor stații grafice dotate cu scanere este indicată planificarea necesarului de echipament și distributia acestuia în cadrul sistemului de arhivare electronică a documentelor.

Includerea unui modul de preluare în sistem automat a documentelor de hârtie sau microfilm în cadrul sistemului de arhivare electronică proiectat este necesară doar în cazurile în care:

- se preia o arhivă manuală în sistem automat în vederea căutării după cuvinte cheie;
- există documente importante (de hârtie în principal) care provin de la parteneri (firme, instituții) și acestea trebuie incluse în baza de documente a firmei.

Motoare de indexare/cautare a documentelor

Motoarele de indexare/cautare a documentelor sunt specifice, în principal, aplicațiilor WEB datorită volumului mare de informație existent în Internet. Ele au ca scop regăsirea de documente electronice prin contținutarea de baze de date în care se păstrează referințe la documente și rezolvarea cererilor de interogare bazate pe expresii logice, formate din cuvinte cheie și operatori logici.

Motoarele de căutare pot fi configurate să indexeze documentele existente în anumite subdirectoare sau aflate la anumite adrese WEB. Răspunsul pe care îl dă motorul de căutare este format dintr-o listă de caracteristici ale documentelor și adresele unde acestea pot fi accesate.

Având în vedere că cercetarea pe care am realizat-o are la bază sistemul de operare Windows 2000 și alte produse ale firmei

Microsoft (SQL 2000 și Exchange 2000) voi prezenta în cele ce urmează serverul de indexare a documentelor, creat de aceeași firmă, denumit Microsoft Indexing Server. Microsoft Indexing Server nu este o noutate pentru cunoscătorii sistemului de operare Windows NT 4.0. Noua versiune nu mai este distribuită într-un pachet adiacent tip Back Office ci a fost integrat în sistemul de operare Windows 2000 împreună cu alte servere precum DNS (Domain Name System), IIS (Internet Information Server), etc.

Noutatea în cadrul acestui server de indexare constă în posibilitatea de creare de scripturi (VBScript, JavaScript) de interogare a bazei de indecsuri prin comenzi SQL. Astfel, pot fi create atât ASP-uri pentru aplicații WEB dar și alte aplicații în arhitectura client-server.

Indexing Server-ul poate fi configurat cu ușurință la nivelul serverului de Windows 2000 putând include și exclude anumite subdirectoare din tinta de indexare. Procesul de indexare are loc în mod transparent, server-ul actualizând automat indecsuri când documentele au fost modificate, șterse sau create.

Totodată, răspunsul pe care îl dă serverul de indexare ține cont de drepturile de acces ale utilizatorului care generează căutarea.

Concluzii

Un sistem de arhivare electronică a documentelor este o necesitate în cadrul unei organizații datorită posibilităților uriașe de interogare pe care le oferă și nu în ultimul rând datorită micșorării spațiilor de depozitare necesare unei arhive manuale.

De remarcat este faptul că preluarea în sistem automat a unei arhive manuale este pasul cel mai greu care trebuie depășit (deoarece necesită eforturi financiare și umane mari), însă odată trecut acest moment orice altă trecere la un sistem mai nou de arhivare se va putea face cu ușurință