

Arhivarea electronica a documentelor

Romeo Mihai PAVELESCU
Doctorand, Informatica Economica

Articolul de fata încearca sa scoata în evidenta importanta exploataarii la maxim a surselor interne de date si informatii în adoptarea rapida a deciziilor. Pentru aceasta este necesara implementarea în cadrul organizatiei respective a unui sistem de gestiune si arhivare electronica a documentelor. Proiectarea si implementarea unui sistem de gestiune si arhivare electronica a documentelor în cadrul unei firmei sau institutii este o activitate complexa si consumatoare de resurse financiare, în special pe linia achizitiei de tehnica de calcul, însa eforturile vor fi rasplatite din plin de eficienta accesului la informatie si de eliminarea timpilor morti în circuitul documentelor.
Cuvinte cheie: arhivare, proiectare, resurse, gestiune eficienta.

Evolutia spectaculoasa a tehnologiei informatiei (IT) din ultimul deceniu ridica stacheta în privinta accesului rapid si facil la o informatie corecta si de actualitate.

Managerii zilei de azi pot obtine informatii, în domeniul lor de activitate, din doua tipuri de surse:

- Surse externe organizatiei pe care o conduc;
- Surse interne organizatiei pe care o conduc.

Sursele externe sunt variate si INTERNETUL ar trebui sa ocupe un loc important printre acestea.

Sursele interne se refera la acele instrumente care ar trebui sa existe în cadrul unei organizatii astfel încât, periodic sau ori de câte ori este nevoie, sa ofere date despre activitatea interna: flux informational, calitatea serviciilor si produselor, probleme de personal, etc.

Arhivarea electronica a documentelor poate fi vazuta ca parte componenta a unui sistem de gestiune a documentelor, dar poate constitui si o activitatea 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 da-

telor presupune, în principal, scanarea/importarea, indexarea, stocarea lor pe suporti de memorare de mare capacitate si caracterizati de o fiabilitate foarte ridicata si regasirea lor dupa diferite criterii.

Înainte de a intra în amanuntele procesului de arhivare electronica este necesara definirea exacta a notiunii de arhiva.

O **arhiva** este formata din date si informatii, indiferent de suportul pe care se afla, care trebuiesc protejate si conservate în vederea consultarii ulterioare pe criterii istorice, culturale sau juridice. Chiar si în ziua de astazi cea mai mare parte a informatiei se afla înregistrata pe hârtie ceea ce necesita un spatiu mare de stocare si este foarte vulnerabila la conditiile de mediu ca umiditate, foc, aciditate, etc. Concluzionând, poate deveni foarte dificil de organizat si conservat în bune conditii informatia existenta într-o arhiva.

Solutia la aceste probleme este transferul informatiilor existente pe suporti de memorare mai putin vulnerabili si care sa ocupe spatiu fizic de stocare cât mai mic.

Scanarea presupune "fotografierea" documentelor (hârtie sau microfilm) si constituirea documentelor electronice pe baza imaginilor obtinute.

Viteza de scanare a hârtiei poate atinge valori mari (40 pag./minut) în cazurile în care calitatea acesteia este buna, altfel depinde de viteza de manevrare a operatorului.

În cazul scanării microfilmelor, viteza poate fi mare doar dacă documentele se afla pe role de film.

În proiectarea unui sistem de arhivare este foarte importanta alegerea formatului în care se vor memora imaginile scanate. Acesta trebuie să asigure pe lângă o ușoară manevră și un algoritm bun de compresie. În mod evident, unele imagini au o rată de comprimare mai bună decât altele. Un document cu un procentaj mare de spațiu alb va putea fi comprimat cu o rată foarte ridicată spre deosebire de o fotografie care nu poate fi comprimată chiar deloc. În acest scop Consultative Committee on International Telephones and Telegraphy (CCITT), o agenție a Națiunilor Unite, a elaborat câteva standarde internaționale pentru transmisia de imagini digitizate sau FAX. Aceste metode, cunoscute ca Group III și Group IV FAX, utilizează algoritmi Huffman și run-length. Diferența între cele două metode de compresie constă în faptul că Group IV FAX se bazează pe o analiză bidimensională a imaginii spre deosebire de o analiză unidimensională în cazul Group III.

Este recomandabilă utilizarea fișierelor multipagina în format TIFF datorită ușurinței în manipulare a acestora.

În etapa de scanare a documentelor este necesară uneori prelucrarea și reconditionarea imaginilor obținute datorită calității scăzute a documentelor originale.

Indexarea documentelor electronice, obținute în urma scanării sau în urma importării, în cazul email-urilor sau fișierelor electronice existente pe diferiți suporti magnetici, constă în atasarea unui număr unic de identificare și a mai multor cuvinte cheie după care acesta va putea fi regăsit. O altă metodă de indexare se bazează pe recunoașterea optică a caracterelor (OCR) din cadrul imaginilor și introducerea textului obținut în baza de date având ca referință numărul unic de identificare a documentului. Aceasta ultimă variantă devine eficientă atunci când documentul original a fost batut la o mașină de scris sau scos la imprimantă și în plus soft-

ware-ul de recunoaștere optică a caracterelor utilizează un dicționar de cuvinte specific limbii în care a fost scris documentul.

Totodată prin indexare se convertește informația nestructurată în informație ce poate fi accesată cu ușurință, din orice punct al intranet-ului.

O alternativă la structurarea documentelor electronice (non imagine) o constituie utilizarea motoarelor dedicate de indexare/cautare documente precum: dtSearch și ISIS. Pe scurt, acestea pot fi configurate să indexeze automat orice document (în diverse formate) la adăugarea sa într-un director pe disk. Cu ajutorul unei aplicații client se pot genera cereri de căutare bazate pe operatori logici și subsiruri obținându-se ca rezultat documente și pozițiile acestora în structura de directoare indexată.

Toate informațiile introduse sau obținute în etapa de indexare vor constitui o bază de date care va permite regăsirea rapidă a documentelor și efectuarea de statistici cu privire la vechimea fondului arhivistic, categoriile de documente existente, calitatea documentelor originale la momentul introducerii în prelucrare automată, etc.

Imaginile digitizate necesită capacități foarte mari de memorare. În paralel cu alte calități stocarea imaginilor trebuie să se facă la un preț scăzut per bit pentru a realiza un sistem viabil de arhivare. Tehnologia discurilor magneto-optice oferă o soluție bună pentru stocare pe termen lung a imaginilor digitizate.

Un disc magneto-optic este un disc pe care se pot memora date prin schimbarea polarității locale cu ajutorul unei raze laser folosită pentru încălzirea locală la temperaturi mari și a unui mic câmp magnetic extern aplicat după atingerea punctului curie. Citirea discului se bazează pe reflexia diferită a razei laser atunci când aceasta întâlnește un spot polarizat al discului.

Un disc magneto-optic poate memora până la 5 GB. Dacă sunt necesare capacități mari de stocare se pot utiliza jukebox-uri ce pot

incorpora de la 20 de discuri magneto-optice pâna la 258.

La alegerea sistemului de **stocare** trebuie ținut cont de faptul ca nu exista o standardizare a discurilor magneto-optice si deci schimbarea ulterioara a producatorului de jukebox-uri sau de drive-uri magneto-optice poate ridica probleme serioase.

Software-ul de arhivare va trebui sa permita o anumita ritmicitate a operatiilor mai sus mentionate, astfel încât sa nu apara supraîncărcari ale rețelei de arhivare si totodata sa permita **consultarea** facila a documentelor electronice introduse.

Datorita faptului ca o arhiva presupune existenta unui volum foarte mare de date, în etapa de proiectare a aplicatiei de arhivare electronica trebuie sa se aleaga un sistem de gestiune a bazelor de date relationale (SGBDR) care poate prelucra asemenea cantitati de informatie (ORACLE, SQLSERVER, INFORMIX, etc.)

Totodata, integritatea bazei de date, create în etapa de indexare, este la fel de importanta ca documentele existente pe discuri magneto-optice. De aceea, trebuiesc prevazute sisteme de siguranta corespunzatoare precum: backup-uri periodice, utilizarea sistemelor redundante de discuri ca RAID 1 (mirroring) si RAID 5 (disk striping).

Problematica arhivarii electronice a documentelor este veche însa odata cu evolutia tehnicii de calcul pot fi identificate solutii din ce în ce mai fiabile. La ora actuala exista pe piata româneasca câteva solutii software integrate însa pentru ca acestea sa poata fi implementate cu succes trebuiesc particularizate functie de specificul si istoricul documentelor ce vor face obiectul arhivarii electronice.

Arhivarea electronica a documentelor prin scanare este o activitate foarte mare consumatoare de resurse si de aceea în paralel cu aceasta ar trebui sa functioneze un sistem de gestiune electronica a documentelor care va minimiza crearea de documente sub forma de hârtie sau film.

Piata aplicatiilor de management a documentelor este foarte dinamica, mai ales în ultima vreme, datorita faptului ca în acest moment exista din punct de vedere hardware solutii viabile însa din punct de vedere software mai sunt multe de pus la punct.

Astfel, Xerox Corporation si Compaq Computer Corporation au anuntat, pe data de 18 mai, stabilirea unei înțelegeri prin care sa se realizeze conectarea informatiilor preluate de pe hârtie, scanate cu ajutorul sistemului Xerox Document Centre, cu Compaq Work Expeditor, cel mai important produs software de management al activitatilor, care extinde facilitatile oferite de Microsoft Exchange prin adaugarea unor procese care automatizeaza si ofera siguranta managementului activitatilor.

Totodata, pe piata româneasca si-a facut aparitia firma FILENET, o firma cu experienta foarte mare în domeniul managementului documentelor, care a realizat pâna în acest moment mai multe implementari la firme, institutii si corporatii cu renume în lume.

Bibliografie

- Archival Information System (AIS), 1992, by Standard Technology Incorporated
- Optical Digital Image Storage System, 1991, by National Archives and Records Administration
- Study on the Automation of the Archiving System, 1995, by Ampaco

Schema conceptuala a unui sistem electronic de arhivare a documentelor

