

Gestiunea automata a documentelor

Romeo Mihai PAVELESCU

Ministerul Apararii Nationale, RRomica@yahoo.com

Gestiunea automata a documentelor are un rol hotarâtor în activitatea organizatiilor în care cea mai mare parte a sistemului informational (peste 80%) are la baza documentul, indiferent de suportul sau fizic. Nevoia de a manevra mai rapid informatiile si, implicit, documentele care le contin a dus la dezvoltarea Sistemelor de Gestiune Automata a Documentelor (SGAD). Managementul documentelor tinde sa înlocuiasca complet managementul datelor – centrul atentiei în tehnologia informatiei în ultimii 20 de ani.

Cuvinte cheie: gestiune automata, documente, versiuni, instrumente.

Managementul documentelor reprezinta controlul automat al documentelor electronice (imagini, spreadsheet-uri, documente word, fisiere HTML, SGML, PDF etc.) de-a lungul întregului ciclu de viata al acestora în cadrul organizatiei, de la creare pâna la arhivare. Managementul documentelor permite exercitarea unui control mai mare asupra productiei, stocarii si distributiei de documente, obtinând astfel o reducere semnificativa a timpilor morti în fluxul documentelor. Un SGAD trebuie sa permita printre altele:

- identificarea unica a documentelor;
- stocarea, regasirea, monitorizarea documentelor;
- controlul versiunilor fiecarui document;
- gestiunea traseelor pe care un document îl poate urma;
- prezentarea/vizualizarea documentelor.

În general, pot fi identificate doua clase de documente electronice:

- de tip imagine;
- de tip word sau spreadsheet (care pot fi editate).

Astfel, gestiunea automata a documentelor difera major functie de clasificarea de mai sus, deoarece imaginile sunt statice, în timp ce documentele care pot fi editate sunt dinamice (într-o continua schimbare). De asemenea, difera si operatiile care pot fi efectuate cu ajutorul acestor documente. Sistemele care gestioneaza imagini se concentreaza pe creare (prin scanare, de exemplu), indexare si regasire ca operatii principale în timp ce sistemele care gestio-

neaza documente ce pot fi editate se focalizeaza pe creare (cu concursul mai multor persoane – grup de autori), gestionarea traseului de urmat (workflow) si controlul modificarilor la nivel central. Oricum, barierele dintre aceste doua tipuri de SGAD-uri au fost darâmate, ajungându-se ca firme specializate pe productia de software pentru un anumit tip de SGAD sa integreze din ce în ce mai multe functii ale celui alt tip de SGAD. Pentru usurarea gestionarii automate a documentelor, firmele implicate coopereaza si creeaza standarde care sa puna bazele unor sisteme deschise de gestiune automata a datelor.

Pe termen lung, expertii în domeniu prognozeaza disparitia separarii functiilor specifice managementului de documente si a aplicatiilor dedicate. Locul lor urmeaza a fi luat de instrumente puse la dispozitie de sistemele de operare, asemanator cu mecanismul de gestiune a fisierelor.

Elementele componente obligatorii ale unui SGAD sunt:

- **Infrastructura** este un set de statii de lucru, dispozitive periferice si servere interconectate în retele locale (LAN) si/sau întinse (WAN). Aceasta trebuie sa permita utilizarea oricarui sistem de operare de retea, oricarui format de fisier, numelor lungi de fisiere etc;

- **Instrumentele specializate în creare de documente** sunt instrumente complexe (software) care permit structurarea informatiei introduse si ghidarea utilizatorului

pentru a introduce datele în anumite moduri anterior proiectate;

- **Workflow-ul** este definit ca fiind coordonatorul task-urilor, datelor și utilizatorilor cu scopul de a transforma procesul de munca într-unul mai eficient și mai ușor adaptabil la schimbări. El reprezintă controlul informației în toate fazele procesului de munca. Calea către un anumit document este dată de tipul documentului (ultimele stiri, manuale, rapoarte, adrese etc), procesele care acționează asupra documentului și de poziția ierarhică în cadrul organizației a utilizatorilor (drepturile de acces la document). Aceasta componentă a unui sistem de gestiune automată a documentelor trebuie să permită efectuarea de operații asupra documentelor, precum scrierea de rezoluții, modificarea, rutarea, comentarea, aprobarea, repartizarea și stabilirea de termene. Workflow-ul reprezintă sistemul nervos al unui SGAD, deoarece permite organizațiilor să controleze și să mărească eficiența fluxului de documente care stă la baza activității lor.

- **Indexarea** reprezintă procesul prin care se creează o listă de cuvinte cheie pe baza cărora se poate accesa rapid informația din baza de documente. Controlul grupurilor de documente este o caracteristică cheie a unui SGAD deoarece de foarte multe ori este necesar să se prelucreză mai multe fișiere (documente) și datele asociate ca un singur obiect în timpul operațiilor de arhivare și regăsire.

- **Stocarea** în cadrul unui SGAD se realizează în principal în baze de date cu motoare de căutare rapide care permit regăsirea de documente. Se observă mutarea masivă a SGAD-urilor dinspre baze de date relaționale spre baze de date orientate obiect. În prezent, sunt utilizate bazele de date mixte numite Object-Relational Database Management Systems.

- **Serviciile de protecție** se referă la mecanismele de control a documentelor precum înregistrarea la intrare și la ieșire, audit-ul, securitatea și controlul versiunilor.

- **Serviciile de prezentare/distribuire** se referă la formele și metodele prin care uti-

lizatorii sunt alimentați cu informație. SGAD-ul trebuie să permită distribuirea în diferite formate a informației: web (intranet), pe CD-ROM sau pe hârtie.

Tendințe în structura documentului electronic

Tendința principală în gestiunea automată a datelor pleacă de la managementul documentelor statice spre documentele compuse și complexe.

Documentele statice, precum imaginile digitale, sunt cele mai puțin flexibile deoarece nu pot fi editate sau citite de către mașina fără o prelucrare preliminară (adică recunoașterea optică a caracterelor). În imediată apropiere a acestora sunt documentele statice, precum cele de tip Word sau spreadsheet-urile, care pot fi editate. Slabiciunile acestora din urmă constau în faptul că depind în mod explicit de o anumită aplicație (de exemplu Excel), sunt "mute" – adică nu conțin informații referitoare la ele sau conțin prea puține informații și sunt "plate" – adică pentru a accesa o informație dorită trebuie să citească și alte informații adiacente.

Documentele complexe sunt la capatul celalalt al evoluției, structura lor eliminând toate slabiciunile mai sus prezentate. Acestea nu sunt legate de o anumită aplicație sau platformă, sunt dinamice și sunt "inteligente", transportând în ele și informația și structura lor. Astfel, documentele reflectă tendința către arhitecturi orientate obiect unde informația rezidă în obiecte (unități de informație cu o granularitate mai mică decât documentele tradiționale) care, de asemenea, conțin informații despre ele și despre aplicația originară. Documentele în această nouă concepție sunt considerate containere cu o gamă largă de informații. Ele sunt colecții de pointeri către elemente externe care sunt asamblate dinamic atunci când sunt regăsite. De exemplu, ele pot conține pointeri către:

- **obiecte cu informație** – conțin corpul propriu-zis al documentului care se assemblează temporar cu un anumit scop și pot fi reutilizate în alte documente;

- **informatii despre comportamentul documentului** – informatii referitoare la cine poate vedea documentul, cine trebuie sa-l aprobe, care este secventa de aprobare si informatii despre procesul de workflow al documentului;

- **metadata** - autorii, istoricul modificarilor si starea curenta;

- **legaturi catre alte elemente externe** - grafice, imagini, fonturi, seturi de date etc. Un document compus mai poate fi privit ca o “lista de materiale” utilizata pentru a releva o anumita reprezentare a informatiei. Documentele compuse pot fi impartite în doua categorii:

- documente prestabilite, care se construiesc de fiecare data dupa aceeasi regula;
- documente virtuale, care sunt asamblate pentru un anumit scop, exista o perioada de timp si se pot schimba de la o vizualizare la alta.

Stadiul actual

În ultimii ani au aparut atât de multe tehnologii care suporta managementul de documente încât principala problema care retine atentia este integrarea acestora.

În general, trei componente mari pot fi regasite la orice SGAD existent pe piata: magazia de documente, motorul de workflow si tehnologia de indexare-cautare.

Pentru a implementa sistemul cel mai bun pentru o firma, trebuie stiut cum lucreaza SGAD-urile si care componenta a managementului de documente se potriveste cel mai bine proceselor si nevoilor din activitatea desfasurata de aceasta.

Magazia de documente reprezinta “sufletul” unui SGAD. Functiile principale ale unei magazii de documente sunt: controlul accesului la un anumit document, catalogarea documentelor, evidenta intrarii-iesirii documentelor, cautarea si regasirea documentelor, controlul versiunilor unui document (gestiunea tuturor instantelor la un anumit document si a modificarilor acestuia de-a lungul timpului). Totodata, sunt gestionate relatiile între documente si partile lor componente (capitole, paragrafe).

Fiecare firma producatoare de SGAD-uri își proiecteaza sistemele în arhitectura 2-tier sau 3-tier. În arhitectura 2-tier clientul realizeaza greul procesarii, pe când într-un mediu 3-tier serverul este cel ce preia greul procesarii.

De obicei, baza de date înmagazineaza toate informatiile referitoare la documente însă nu si documentele în sine. Baza de date contine pointeri la fisiere.

Multe dintre aplicatiile client utilizeaza standardul OLE, ceea ce permite utilizarea continutului obiectelor dintr-o aplicatie în interiorul altei aplicatii fara a mai realiza alte operatii de “cut”, “copy” sau “paste”. SGAD-urile recunosc, în general, documentele OLE si sunt destul de inteligente pentru a mentine aceste fisiere si legaturile catre acestea în magazia de documente.

Magaziile de documente permit crearea de obiecte “container” si utilizarea SGAD-ului pentru legarea obiectelor “document”. De asemenea, foarte important este nivelul de integrare între aceste instrumente OLE si magazia de documente.

Workflow-ul poate elimina timpii morti ai documentului pierduti în tranzitul de la un lucrator la altul. Acesta permite vizualizarea unui document în paralel de catre mai multi utilizatori, ceea ce micsoreaza timpul pierdut în procesul de aprobare a documentelor. Cuplat cu magazia de documente, workflow-ul permite un audit complet, inclusiv a adnotarilor. Sistemul de workflow permite anuntarea lucratorilor implicati atunci când o noua versiune a unui document a devenit disponibila si poate coordona procesul de conversie a documentelor dintr-un format în altul. Motorul de workflow integreaza, de fapt, magazia de documente a SGAD-ului si sistemul de e-mail utilizat în transportul documentelor.

Cea mai mare parte a rutarii mesajelor dintr-un sistem de workflow se finalizeaza într-o casuta postala privata.

Acest fapt este benefic pentru lucratorii care sunt implicati în producerea documentelor, însă pentru aceia care lucreaza în echipa nu este chiar la îndemâna. Astfel,

pentru acestia sistemul nu trebuie sa ruteze prin e-mail documentul propriu-zis, ci mesaje ce fac referire la acest document. Pentru a mentine integritatea magaziei de documente nu este recomandat transferul de documente catre un sistem prea putin controlabil precum cel de e-mail.

Tehnologia de indexare-cautare a unui SGAD poate fi mult mai eficienta decât tehnologia standard de cautare "full text", datorita posibilitatilor de cautare a documentelor dupa anumite atribute. Magazia de documente poate adauga documente atât într-un index "full text", cât si într-o baza de date relationala.

Motorul de cautare îmbina cele doua instrumente obținându-se astfel rezultate mult mai apropiate de necesitati.

Considerații tehnice

Trebuie luate în calcul trei probleme importante în proiectarea unui SGAD:

- alegerea unei platforme de calcul adecvate;
- proiectarea rețelei prin care se va face transferul documentelor si mesajelor;
- proiectarea bazei de documente.

Platforma de calcul face referire la mediul de lucru la nivelul clientului si a serverului si, în general, acestea trebuie sa fie compatibile perfect pentru a elimina necesitatea de a scrie cod de program pentru fiecare dintre acestea.

Până de curând, cele mai multe servere destinate magaziei de documente rulau sistemul de operare UNIX deoarece acesta oferea puterea necesara procesarii si garanta securitatea documentelor. Acum exista un bun competitor, (Windows 2000 Server), care permite implementarea cu succes a componentelor unui SGAD. Oricum înainte de a alege un sistem Windows 2000 trebuie analizat daca acesta se potrivește cu dezvoltarea ulterioara preconizata

a organizatiei în cauza si daca aplicatia server utilizeaza avantajul multiprocesarii simetrice (SMP).

Bazele de date utilizate în cadrul magaziei de documente pot fi relationale, orientate-obiect sau mixte. Cea mai mare parte a producatorilor de SGAD-uri utilizeaza unul din urmatoarele SGBD-uri: Oracle, Sybase, Informix si SQL Server.

Concluzii

Concurenta acerba existenta astazi în orice segment de piata îi determina pe participantii sa caute mijloace noi de obtinere a informatiilor, indiferent de natura lor: economice, de personal, tehnice, etc.

Evolutia spectaculoasa a tehnologiei informatiei 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:

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

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 firme sau institutii 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.

Proiectarea si implementarea unui sistem de gestiune automata a documentelor în cadrul unei organizatii 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.