
UN PROGRAMA

LECTOR INTEL.LIGENT DE TEXT

Agustí Romero Gavilán

Un grup de tècnics joves de Barcelona guanya un premi a la creativitat amb el programa TexPert i aconseguix comercialitzar aquest programa als Estats Units.

El paper ha estat el suport tradicional de la informació i cal ser conscient del que suposa aconseguir fusta per fabricar pasta de paper. Només per donar una idea del que diem, s'ha de tenir en compte que la confecció de l'exemplar dominical del New York Times requereix el sacrifici de 66.000 arbres per produir la pasta de paper necessària.

Els imperatius ecològics del nostre temps imposen unes altres vies d'emmagatzemar informació, perquè ja estem arribant a uns límits insostenibles pel que fa al "cost" del paper.

La micrografia, amb uns setanta anys d'existència, ha significat una opció interessant, si bé s'ha de reconèixer que no ha assolit una extensió i una aplicació massives. Els suports informàtics, tant els òptics com els magnètics, tenen grans avantatges.

Representen un cost més baix i no requereixen tant espai d'emmagatzematge. A més a més, són molt més accessibles, es poden comunicar des de punts molt distants i permeten un accés ràpid i segur. El futur, en aquest camp, ens presenta la necessitat de poder transvasar grans quantitats d'informació des d'un suport de paper

a un d'informàtic, i això s'ha d'aconseguir de manera ràpida i fiable. Fins ara, aquest procés requeria la feina manual de teclistes que introduïssin la informació en el suport.

Les possibilitats que ofereix la tecnologia del làser han permès que la impressió, fins fa poc limitada a sistemes matricials i a papers continus especials, es pogués estendre a altres formats i mides i tingués uns nivells de qualitat estimables.

Ara bé, les noves eines de pro-

gramació per a llenguatges articulats permeten desenvolupar programes adaptats a les més diverses funcions. Les consecucions de les anomenades tècniques d'intel·ligència artificial permeten aplicacions tan complexes com és ara el reconeixement de formes o bé de sons per mitjà de màquines cada vegada més senzilles.

Un equip de joves tècnics de Barcelona ha guanyat un accèssit en el VII Premi de Creativitat 1989 perquè havien aconseguit un programa de gran eficàcia en una de les aplicacions de més recent aparició.

TexPert és l'obra de Joan Ignasi Martínez Roldán, llicenciat en Medicina, encara que relacionat amb el món de la informàtica, i Alfons Sort Vidal, enginyer industrial. Aquests joves investigadors van crear fa alguns anys l'empresa Ciència i Tecnologia Aplicada, que actualment sembla estar entre les quatre primeres del món en el camp de l'OCR (Reconeixement Òptic de Caràcters) o lectura òptica de textos mitjançant les tècniques de la intel·ligència artificial.

El projecte a què ens referim se situa entre les tècniques que permeten passar el text del paper al disc sense necessitat de teclejar.

■ **Tècniques
que permeten
passar el text
del paper al
disc sense
necessitat de
teclejar**

Preferencias de lectura

Modo de lectura	Preferencias en área de lectura
<input checked="" type="checkbox"/> Usar Biblioteca Interna <input checked="" type="checkbox"/> Autodefinición de áreas <input checked="" type="checkbox"/> Incluir titulares <input checked="" type="checkbox"/> Incluir texto invertido <input type="checkbox"/> Páginas múltiples Alfabeto: <input type="text" value="Romano"/> Contexto: <input type="text" value="Activo"/>	Carácter al final del área: <input type="text"/> Carácter al final de la línea: <input type="text"/> Carácter al principio del área: <input type="text"/> <input type="checkbox"/> Retorno al final de cada área <input type="checkbox"/> Tabuladores en vez de espacios <input type="checkbox"/> Retorno al final de cada línea <input type="checkbox"/> Tabla financiera
Conjunto de caracteres: <input type="text" value="Todos"/> <input type="button" value="Personalizar"/>	<input type="button" value="Cancelar"/> <input type="button" value="OK"/>



Un dels menús del programa TexPert.

UN ESDEVENIMENT ENCORATJADOR

És un motiu de satisfacció molt particular veure com uns joves investigadors catalans porten a bon terme innovacions valuoses en el camp de l'alta tecnologia. També és encoratjador veure com gosen obrir una filial de la seva empresa, Ciència i Tecnologia Aplicada (CTA), als Estats Units per comercialitzar-hi el programa TexPert i preveuen iniciar una campanya enèrgica de venda dels seus productes en un país d'alta competència pel que fa a les novetats de tipus tecnològic.

Sens dubte, aquesta iniciativa respon al fet que cada cop s'imposa amb més força la necessitat d'introduir la informació en suports de tipus digital, amb la possibilitat de transvasar-la després a paper. Aquest és un procés de doble sentit, adaptable a les més diverses conveniències.

Les possibilitats que ofereixen les tècniques de lectura electrònica representen un gran avantatge en un món on la informació es duplica cada quatre anys. És perfectament previsible la necessitat d'assegurar suports adequats per absorbir, reconèixer i transvasar al més ràpidament possible aquesta informació. ■

Agustí Romero Gavilán

és vice-president de l'Associació Nacional d'Usuaris de Cad i Autoedició d'Espanya (ANCAE)

CREACIÓ DE TEXPERT

Les primeres versions del programa TexPert s'han fet per a ordinadors Macintosh, desenvolupades en llenguatge Pascal i adaptades a la majoria d'escàners disponibles en el mercat. Posteriorment s'ha traduït al llenguatge C, versió 3.0, amb una velocitat de lectura de 3.500 cpm, amb una màxima facilitat d'ús i amb una versió per a PC.

Bàsicament, el programa permet que l'ordinador reconegui les pàgines i literalment "llegeixi", és a dir, "entengui" que una determinada taca negra és un caràcter determinat i el pugui captar i reproduir.

El sistema de treball del programa TexPert consta d'un ordinador de la família Macintosh, amb una memòria mínima d'1 Mb i amb disc dur; d'un escàner controlat pel programa per fer, en una sola operació, tant el procés de digitalitzar la pàgina com el de reconèixer la informació gràfica o tipogràfica i el programa, que es presenta en forma de disc de 3.5". Tot això, juntament amb un manual d'utilització, guia de l'usuari i exemples per a l'aprenentatge.

FUNCIONS DEL PROGRAMA

Les funcions més importants que duu a terme el programa en l'ac-

tual versió 3.0 són les següents:

- Identificar i aïllar automàticament els blocs de text i gràfics de la pàgina, que s'analitzen per comparació amb els models matemàtics de les formes.

- Autoaprenentatge del tipus de lletra del text.

- Possibilitat d'aprendre manualment nous tipus de lletra per a documents especials (en cas de baixa qualitat d'impressió, tipografies especials, etc.)

- Possibilitat de llegir quatre alfabet: el llatí, el ciríl·lic, l'hebreu i el grec. Adaptació a vint-i-cinc idiomes que utilitzen l'alfabet llatí, incloent-hi el català.

- Fiabilitat d'un 99,9% en condicions adequades de treball.

- Arxiu de característiques de format, distribució de blocs de text a la pàgina i característiques tipogràfiques del text per editar-lo posteriorment.

- Inserció de caràcters especials per exportar dades llegides a bases de dades. Per exemple, llegir dades de directoris telefònics o bé de les pàgines grogues i arxivar-les en bases de dades.

- Adaptació per llegir dades financeres quan s'hi col·loquen tabuladors i selecció exclusiva de caràcters numèrics.

- Editar, modificar el text llegit i corregir-ne els errors des del mateix programa, dotat de funcions específiques de correcció.