

La documentació de premsa sota la llum de la tecnologia: el futur avui

Begoña Aguilera
Delphi Information Solutions
delphi@ptv.es

Richard Lehner
richard.lehner@t-online.es

R E S U M

La documentació de premsa és una eina indispensable per als centres de documentació. Es presenten i valoren les diferents alternatives possibles (premsa impresa *versus* premsa electrònica, arxiu en paper *versus* arxiu digital, serveis interns *versus* sub contractació, eines electròniques de digitalització, etc.) a considerar quan es vol treballar amb aquest tipus de documentació i quines són les tendències de futur.

Paraules clau: Documentació, Premsa, Mitjans de comunicació, Digitalització, Retalls de premsa.

R E S U M E N

La documentación de prensa constituye un instrumento indispensable para los centros de documentación. Se presentan y valoran las diferentes alternativas posibles (prensa escrita o prensa electrónica, archivo en papel o archivo digital, servicios internos o subcontratación, instrumentos electrónicos de digitalización, etc.) a considerar cuando se quiera trabajar con este tipo de documentación, así como las tendencias del futuro.

Palabras clave: Documentación, Prensa, Medios de comunicación, Digitalización, Recortes de prensa.

A B S T R A C T

Press documentation is a valuable tool for documentation centers. Different possible options such as Printed Press vs. Electronic Press, Paper Archives vs. Digital Archives, In-house Services vs. Outsourcing, Digitalization Electronic Tools, etc., are presented and evaluated for consideration when working with this kind of documentation and what are the future trends.

Keywords: Documentation, Press, Broadcasting, Digitization, Press clippings.

La documentació de premsa és una font indispensable per a qualsevol centre de documentació. Parlem voluntàriament de documentació de premsa i no de premsa en general; cal una selecció, un filtrat, una organització i estructuració de la premsa per aconseguir que aquesta es converteixi en una eina d'informació. Una col·lecció de premsa per sí mateixa, per molt exhaustiva que sigui, no és més que una col·lecció d'informació «atrapada» en uns do-

cuments. Cal extreure aquesta informació i donar-li una nova vida a la premsa. Fem servir així el concepte de «documentació de premsa» d'una forma restrictiva voluntàriament.

La documentació de premsa ofereix un ampli ventall d'avantatges i possibilitats als seus usuaris: presenta una visió completa sobre el desenvolupament cronològic de qualsevol tema, mostra quines són les influències i diferències entre temes concrets, és el mirall dels fets diaris i també dels retrospectius. És per aquestes raons que no només els centres de documentació dels mitjans informatius i editorials, sinó també la majoria d'empreses i organismes fan servir la documentació de premsa com informació molt valuosa per a la seva activitat.

És significatiu com en els últims anys han sorgit diferents respostes a aquesta necessitat que presenta el mercat de la documentació i la informació en relació a la premsa. L'oferta és extensa i variada, tant en la vessant de continguts (bases de dades, portals, cercadors especialitzats, serveis d'alerta), com en la vessant de serveis (empreses de seguiment de premsa, de serveis de digitalització retrospectiva, etc.) i també en la de les eines (programaris de captura especialitzats en premsa, OCRs millorats per reduir el soroll característic del material de premsa, etc.)

En un moment econòmic en què cal justificar adequadament les inversions i la despesa corrent, hi ha una necessitat de rendibilitzar-les oferint nous serveis i noves possibilitats a les col·leccions que ja tenen la majoria de centres de documentació i biblioteques. L'èmfasi des del nostre punt de vista no estaria només en les col·leccions sinó en els nous formats, els nous mitjans i els nous serveis que es poden extreure d'aquestes col·leccions.

Amb el desenvolupament de la tecnologia i els serveis, un centre de documentació té un ventall de possibilitats entre les quals escollir quan es planteja com treballar amb documentació de premsa. La intenció d'aquest article és analitzar quines són les noves possibilitats que es presenten relacionades amb la documentació de premsa (premsa electrònica *versus* digitalització de premsa, serveis interns *versus* subcontractació, premsa impresa *versus* premsa electrònica, etc.) i plantejar propostes de futur.¹

Les arrels de la documentació de premsa

La història de la documentació de premsa està fortament lligada a la història de la premsa mateixa, de manera que les seves arrels es situen en el periodisme més que en la informació i la documentació. Amb els desenvolupaments tècnics del processament de dades en els últims vint anys, els documentalistes de premsa (en la seva majoria periodistes) van ser conscients de la possibilitat de fer servir els mètodes de processament de dades de les tècniques documentals. D'aquesta manera, les tècniques documentals que s'havien aplicat només a disciplines científiques i tècniques es van aplicar també a la documentació de premsa, amb resultats molt satisfactoris.

Els primers arxius de premsa van consistir en col·leccions de diaris, però ja en el segle XVIII es va veure la necessitat de donar una estructura a la informació de premsa per poder extreure el contingut i verificar dades i fets.

Originàriament els editors de premsa eren capaços de assimilar les dades i els coneixement que utilitzaven per a la seva professió. Amb l'explosió informativa i el desenvolupament de la natura de l'edició, cada vegada més global, es va fer necessari trobar una metodologia que els permetés arribar a la informació que necessitaven. Calia, ja en el segle XVIII emma-

1. El tema és tan ric i divers que permetria escriure una monografia analitzant en profunditat totes les opcions i experiències relacionades amb la documentació de premsa. Serveixi aquest article com una primera aproximació al tema.

gatzemar el coneixement. Les cases editorials van detectar que podien millorar la seva activitat amb col·leccions ben seleccionades i organitzades. És el naixement dels primers arxius de premsa a les cases editorials europees.

En aquesta primera fase la metodologia era descriptiva, de manera que es referenciava la localització de la notícia, la qual cosa requeria recuperar posteriorment la notícia de la col·lecció. Aquestes grans col·leccions de diaris, amb el temps van passar a les biblioteques nacionals, com la Biblioteca Nacional de Noruega (NBR) que conserva una àmplia col·lecció de diaris des del segle XVI. Una altra col·lecció interessant i molt completa es conserva a la biblioteca de la Universitat de Uppsala a Suècia. I ja en un entorn més proper a nosaltres podem mencionar les col·leccions de la Biblioteca Nacional de Madrid i les de la Biblioteca de Catalunya.

Aviat en entorns periodístics els sistemes de fitxes es van considerar poc flexibles i es va optar per retallar els articles interessants i arxivar-los separatament de les col·leccions. Es van formar els arxius de retalls de premsa, que ordenats temàticament i emmagatzemats cronològicament en dossiers permetien respondre ràpidament a les necessitats dels periodistes.

A la vegada que es desenvolupaven aquests arxius a les cases editorials, una nova indústria va néixer ja en el segle XIX, descobrint un nou mercat: empreses que subministraven a cada client els retalls de premsa que necessitava. Al 1896 es va crear a Viena l'empresa Observer que ja al segle XIX treballava amb 2500 diaris i 7500 revistes de tot el món.²

Un cop generalitzada aquesta nova manera de treballar amb la premsa, els arxius de retalls de premsa es van implantar per tota Europa. El primer arxiu, independent de qualsevol editorial data de 1908. En aquest any a l'Institut d'Investigació Econòmica de la Universitat d'Hamburg (HWWA) es va crear un arxiu de retalls de premsa amb el propòsit de col·leccionar informació científica sobre el desenvolupament econòmic i polític en general.³

El 1913 el periodista i editor Ludwig Munzinger va veure la necessitat de construir el que posteriorment seria el Munzinger Archiv, conceptualitzat com l'arxiu per als editors. Avui aquest treball s'ha plasmat en productes editorials (CD-ROM, etc.). I encara que els arxius es van concebre com un servei de notícies per a la premsa, avui són utilitzats per mitjans de comunicació, biblioteques, empreses, partits polítics, etc.⁴

El 1914 es va posar en marxa el *Pressearchiv der Reichsregierung* o l'arxiu de premsa del govern alemany, en el que treballaven al voltant de 18 lectors, amb 720 títols de premsa.

En el període de reconstrucció després de la II Guerra Mundial, Europa va experimentar un període de creixement pel que fa a la documentació de premsa, que es va estendre a diaris regionals i locals, així com setmanaris, diaris tècnics i corporacions de medis de comunicació. Una bona mostra és la col·lecció de premsa de la BBC. Moltes altres institucions, a més de les cases editorials i medis de comunicació van començar a construir els seus arxius de premsa. Els diaris s'analitzaven, retallaven i arxivaven als ministeris i parlaments, als partits polítics, a les empreses, a les universitats ...

Un moment molt important va ser a rel de la introducció de les bases de dades en la documentació de premsa. El primer pas es va fer amb la microfilmació, que permetia iniciar l'automatització a preus raonables. El 1970 més del 90 % dels diaris americans es conser-

2. Hans Bohrmann, Marianne Englert, eds. *Handbuch der Pressearchive*. München: Saur, 1984, p. 7-19

3. <<http://www.hwwa.de>>

4. <<http://www.munzinger.de>>

vaven reproduïts en microfilm. El desenvolupament europeu va ser més lent que l'americà però constant. El 1965 a Alemanya es van posar en marxa els arxius de films de la premsa alemanya, que van registrar en microfilm valuoses col·leccions de diaris per a la recerca.

El salt qualitatiu i quantitatiu es produeix als anys setanta, quan el 1972 el *New York Times* ofereix la primera base de dades de premsa en línia del món. Aquestes primeres bases de dades combinaven la reproducció del microfilm amb el processament de dades electrònic, de manera que es retornava d'alguna manera al sistema de fitxes: el text complet es conservava en el film i la descripció bibliogràfica així com un resum de l'article es conservava a la base de dades.

Amb la introducció del microxip les tècniques de processament electrònic van canviar de nou. Inicialment es va tendir, per limitacions dels mateixos ordenadors (els costos eren molt elevats i difícilment justificables) per construir bases de dades bibliogràfiques, però amb l'evolució dels últims anys de la tecnologia és avui possible construir sense gaire dificultats tècniques o pressupostàries les bases de dades de premsa en text complet

Avui, les necessitats bàsiques continuen sent les mateixes que a principis de segle: conservar les col·leccions de premsa, i donar accés al contingut de les col·leccions. Però els requeriments són cada vegada més grans: els usuaris dels centres de documentació de premsa no es contenen amb buscar manualment en els dossiers de premsa la informació que necessiten, sinó que demanen un accés directe, ràpid i eficaç. Demanen noves eines i serveis que els permetin accedir a la informació des del seu lloc de treball, sense haver de buscar en incòmodes dossiers en paper. Volen tenir ràpidament tota la informació publicada sobre un tema, i demanen també serveis d'alerta que els mantinguin al dia dels temes que els interessa (d'aquí per exemple, l'expectació que està provocant les noves possibilitats associades al WAP).

I els mateixos documentalistes, conscients d'aquestes necessitats busquen les solucions que els permeten donar als seus usuaris aquella informació que busquen, de manera ràpida, àgil, exportable a altres formats, etc. És realment sorprenent, fins i tot per a un documentalista, reconèixer com la verdadera ànima dels arxius de premsa han estat els documentalistes, coneixedors en profunditat dels continguts de la seva col·lecció, capaços de recuperar amb ben poques i limitades eines bibliogràfiques els articles demanats. I aquests documentalistes veuen amb expectació les possibilitats que les noves eines electròniques poden portar a la seva feina, per rendibilitzar i optimitzar el seu esforç d'una manera considerable.

Noves eines per als mateixos problemes

Una de les primeres preguntes que es fa un documentalista a l'hora d'abordar el tema de premsa és: quina premsa és la que interessa als meus usuaris? En què ha de basar-se el meu servei? M'interessa la premsa «impresa» o puc treballar amb altres fonts? Què m'interessa de la premsa, el contingut o la imatge? O les dues coses?

Habitualment quan parlem de premsa ens referim a la premsa impresa. És aquesta la premsa amb una tirada més gran, a la qual accedeix qualsevol ciutadà que només necessita apropar-se a un quiosc i comprar un exemplar. Aquesta és la premsa que fins ara tenen en compte les empreses d'estudis de mercat per exemple, o els polítics quan valoren el seu impacte en la societat. Encara que la societat està sotmesa a un canvi profund amb el ràpid desenvolupament d'Internet, de moment són els diaris impresos els que més genèricament reflecteixen els fets socials, econòmics, culturals i polítics. Els centres de documentació han treballat i treballen principalment encara amb aquestes edicions. Les raons són molt simples: els seus usuaris o clients no estan només interessats en el text de l'article de diari sinó que

demanen més informació. Volen veure la tipografia de l'article imprès, el disseny de pàgina en la que l'article va aparèixer, les fotografies, la impressió de l'article dins del diari, etc. En aquests centres de documentació i arxius no es contempla la possibilitat de fer servir una altra font que no sigui la impresa, per la qual cosa quan es dissenyen sistemes que permetin accedir a aquesta informació, la resposta passa sempre –excepte en el cas de *MyNews*, que comentarem més tard– per la reproducció o digitalització.

Però abans de veure quines són les possibilitats amb les que compta el documentalista quan necessita accés a premsa impresa, val la pena observar què està passant amb els nous medis. En aquest cas els diaris més importants –tant a nivell nacional com a nivell internacional– tenen ja una edició en versió electrònica distribuïda per internet. Són ben conegudes les versions electròniques de *El País Digital*,⁵ *La Vanguardia*,⁶ *El Mundo*,⁷ etc., per mencionar algunes. En aquests casos es presenta en format electrònic (habitualment HTML) una edició diferent –encara que aproximada– a l'edició en paper, amb una presentació, selecció, contingut i ordres diferents. Aquestes edicions acostumen a ser gratuïtes, raó per la qual són utilitzades per centres que sense disposar d'un pressupost elevat poden estar al dia de les notícies més rellevants i oferir aquest servei als seus usuaris.

Seguint aquest desenvolupament dels mitjans escrits, de seguida van aparèixer a Internet repertoris d'aquestes edicions de l'estil de *Kiosco Internet*⁸ on es relacionen totes les edicions electròniques de diaris de l'àmbit espanyol i llatinoamericà, i on es recopilen els enllaços als principals diaris, agrupats per temes.

Un cop es va donar accés als diaris recopilats, i agrupades les seves notícies per temes es va tendir a oferir serveis de subscripció gratuïta, on donant-se d'alta com a usuari registrat (per exemple de *Telepolis*)⁹ es rep a la bústia els principals titulars dels diaris més importants d'Espanya. Aquesta tècnica *push* permet definir perfils més o menys personalitzats,¹⁰ i rebre diàriament els temes del nostre interès a la bústia de correu electrònic. Un producte semblant és el *CDETracker*, que realitza un servei de seguiment automatitzat i a la mida de diverses fonts d'informació d'Internet, entre elles la informació de premsa.

Del cuc a la papallona: digitalització de la documentació de premsa

Malgrat el naixement de les edicions electròniques dels diaris, la majoria de centres de documentació continuen treballant amb la premsa impresa. El salt qualitatiu i quantitatiu que s'ha fet és treballar amb mitjans electrònics que permeten donar una nova vida a la documentació de premsa en paper: parlem de la digitalització.

Abans d'analitzar com es treballa en el món de la documentació de premsa amb eines digitals cal fer una primera reflexió: encara que la tecnologia avança molt ràpidament, els usos que se'n fan estan molt lluny de ser racionals. Observem què succeeix en un dia qualsevol d'un documentalista que treballa amb premsa amb les eines més actuals:

El nostre documentalista «escanneja» i retalla en pantalla (no amb tisores i cola; més endavant repassarem els requeriments per a un bon software) els articles que l'interessen de la premsa del dia. A la vegada que fa el retall, el software produeix automàticament tota la des-

5. <<http://www.elpais.es>>

6. <<http://www.lavanguardia.es>>

7. <<http://www.elmundo.es>>

Altres títols: ABC <<http://www.abc.es>>, El Periódico de Catalunya <<http://www.elperiodico.es>>, L'Avui <<http://www.avui.es>>, El Diario vasco <<http://www.diario-vasco.com>> etc.

8. <<http://www.kiosco.net>>

9. <<http://www.telepolis.com>>

10. Veure <<http://www.paperball.de>> i <<http://www.paperboy.de>>

cripció bibliogràfica de l'article (títol, autor, pàgines, etc.) i ell afegeix la indexació o classificació intel·lectual que considera necessària. Un cop ja té els fitxers de la imatge, dels índexs i del text complet, el sistema els distribueix simultàniament a la base de dades, que s'actualitza amb la nova informació, i elabora també automàticament el dossier o dossiers de premsa diaris per subministrar als seus usuaris.

Fins aquí tot correcte, però ens hem parat a pensar què és el que està fent el documentalista? El nostre documentalista està tornant a posar en format electrònic (text, imatge i índexs) una informació que es va generar precisament d'aquesta manera, a les redaccions dels diaris. El problema està precisament en el pas entre la informació electrònica que produeix el diari i la informació electrònica que genera el documentalista: no hi ha un pont generalitzat que permeti obtenir de les editorials els fitxers que necessiten els documentalistes.

En aquest sentit val la pena destacar una arquitectura, que encara que limitada, va encaminada a construir aquest pont entre els diaris i els documentalistes. Es tracta del producte *MyNews*,¹¹ que obté directament dels diaris els fitxers que els seus editors fan servir per imprimir-los. Aquests fitxers alimenten automàticament la base de dades de *MyNews* desenvolupada sobre el programa anomenat *Autonomy*.¹² La limitació de *MyNews* és precisament que la majoria de diaris, per diferents raons, no poden subministrar els seus fitxers. Així els clients de *MyNews* en molts casos acaben completant la informació que adquireixen a través d'aquest servei amb informació que bé obtenen d'altres fonts o que digitalitzen ells mateixos.

La documentació de premsa es pot trobar avui en diferents aplicacions i en quasi totes les indústries de l'economia així com institucions i organitzacions governamentals. Probablement les més habituals són les aplicacions dels mitjans de comunicacions i les cases editorials; aquests són els dos grups que més ràpidament han transformat els seus arxius en paper en arxius digitalitzats. La seva necessitat és clarament l'accés ràpid i precís a dades i fets actuals. Amb les possibilitats de la digitalització els documentalistes poden dedicar-se a les feines que els són més pròpies: la selecció experta dels articles rellevants així com la indexació intel·lectual. Desgraciadament el que havia de ser una racionalització de les tasques a realitzar pels documentalistes ha derivat en alguns casos en la reducció de llocs de treball, i en uns altres en la confiança absoluta –excessiva– en les possibilitats de la tecnologia, infravalorant la feina del documentalista en favor dels motors de cerca, etc. Sense una adequada indexació intel·lectual que recolzi la documentació de premsa, els usuaris de les bases de dades es troben absolutament ofegats pel número de respostes a la seva cerca, i el temps que abans dedicaven a remenar papers als dossiers de premsa el dediquen ara a netejar les cerques per paraula clau al text complet.

Cada vegada que es planteja un nou projecte d'arxiu digital de premsa a una institució cal començar per una planificació acurada. És recomanable buscar un sistema documental ben dissenyat conceptualment que permeti no només realitzar la tasca per a la que ha estat concebut (accés a la documentació de premsa) sinó que també permeti el desenvolupament de nous productes així com la reconversió de material ja publicat o l'anàlisi qualitatiu. Un exemple d'una bona conceptualització a l'hora de dissenyar un arxiu de premsa digital, i una bona implementació és l'exemple de la televisió privada alemanya *SAT1*. Aquesta cadena no només utilitza l'arxiu de premsa per a la cerca retrospectiva de fets i temes. Des del començament de la definició del projecte es va veure que un arxiu digital podia ser utilitzat també com un sistema de *datamining*. Aquest és l'ús addicional que li han donat a l'arxiu: els documentalistes

11. <<http://www.mynewsonline.com>>

12. Tot i que no és una base de dades molt coneguda a Espanya val la pena estudiar les seves característiques i prestacions, que en alguns casos poden considerar-se com revolucionàries si es comparen amb altres bases de dades del mercat. <<http://www.autonomy.com>> Realment l'equip de *MyNews* va fer un molt bon treball seleccionant i dissenyant l'arquitectura del seu producte.

porten a termini anàlisis de ressonància amb l'ajuda del programa. El sistema de captura dels retalls de premsa de SAT1, *newsWORKS* analitza automàticament els articles seleccionats, i extreu informació de situació de la notícia, dimensió i continguts. Aquestes informacions són processades de manera que la cadena fa un seguiment acurat i precís de la ressonància dels programes de televisió de la cadena en la premsa de cara a desenvolupar futures accions que millorin la percepció dels periodistes que escriuen sobre els programes de SAT1.

Una altra tendència que s'observa en les empreses de comunicació quan treballen amb digitalització de premsa és organitzar la captura i selecció de la informació en xarxa. Generalment es tracta d'empreses situades a diferents llocs del país per la qual cosa els retalls de premsa es preparen localment i es comparteixen en xarxa. D'aquesta manera es solucionen dos problemes: la necessitat d'emmagatzematges de gran volum –perquè es construeixen bases de dades distribuïdes– i la reducció del retard en la producció de la documentació de premsa, que ara es distribueix entre els diferents centres.

A Espanya la situació és que –excepte mitjans de comunicació i algun grup editorial– molts centres continuen treballant amb paper, amb totes les implicacions que això suposa: limitació de l'accés a la informació, limitació de la recuperació de la informació pertinent, multiplicació del número de fotocòpies, limitació en els serveis als usuaris, etc. Les dades del mercat espanyol serien molt semblants¹³ a les presentades a l'informe d'explotació del projecte Laurin,¹⁴ on un 50 % dels centres entrevistats preveuen instal·lar durant el 2001 un software de retall de premsa i una base de dades.

En el pas de la pràctica manual en paper a la pràctica automàtica en format digital hi ha molt sovint un pas intermedi no desitjat que ve justificat moltes vegades pel desconeixement de sistemes que resolen aquests problemes. En alguns centres es continua retallant de forma manual la fotocòpia dels diaris, i es digitalitza la còpia d'aquesta feina manual. Digitalitzar d'aquesta manera no és de cap manera rentable perquè introdueix un pas més en la digitalització (fotocòpia i retall manual) i produeix fitxers de baixa qualitat, punt decisiu a l'hora de produir el text complet de la notícia amb l'ajut d'un OCR.

Hi ha però centres que ja treballen retallant en pantalla els articles escollits. La manera més habitual de treballar en aquestes digitalitzacions¹⁵ és fer servir un software d'edició gràfica que emula el treball fet manualment. Això és, s'escanneja la imatge de l'article i s'emmagatzema, afegint els índexs necessaris per a la seva identificació. La majoria de programes que existeixen al mercat espanyol treballen d'aquesta manera, i així substitueixen les tisores i la cola per obtenir una imatge digital, però no acaben d'automatitzar la feina, amb la qual cosa es perden en part els avantatges de la digitalització. La inversió que suposa fer servir un programa d'aquestes característiques (inversió en diners però també inversió en temps de formació del personal, etc.) no és totalment rentabilitzada perquè es perden moltes possibilitats i funcionalitats fàcilment aplicables als retalls de premsa digitalitzats.

Han aparegut en el mercat nous programes especialment dissenyats per als retalls de premsa, que permeten automatitzar tota la feina manual, que no té un valor afegit per sí mateixa, de manera que el documentalista pot dedicar més esforços a la indexació intel·lectual, aspecte en el que l'experiència ha demostrat que no pot ser substituït per cap sistema artificial (al menys de moment). Però fins i tot en aquest punt els nous programes ajuden al documentalista amb sistemes intel·ligents que suggereixen, per exemple, els articles que *podrien* coincidir amb el perfil del centre.

13. Degut a la nostra activitat com a consultors en digitalització, específicament en premsa, hem pogut constatar directament aquestes dades en els nostres contactes amb molts professionals que realitzen aquesta tasca.

14. Günter Mühlberger, Kurt Habitzel. *Deliverable D11: Exploitation plans. Version 1, august 2000.* <<http://laurin.uibk.ac.at>>

15. Segons hem pogut comprovar en la nostra recerca a l'estat espanyol i Portugal

Nous productes, nous serveis: el dia després de la digitalització

Fins ara hem comentat les possibilitats que presenta la digitalització per als centres de premsa, basant-nos en reproduir la mateixa activitat que fins ara es feia en paper, o sigui confeccionar un arxiu de documentació de premsa. Però en el moment en què ja es disposa dels articles digitalitzats, com a valor afegit es presenten noves possibilitats i nous serveis que poden ser contemplats pels centres de documentació.

El típic fluxe de treball d'un arxiu de premsa digital seria el següent:



Alguns centres de documentació, especialment per justificar una inversió en digitalització que moltes vegades és difícil d'entendre per a les institucions en les que treballen estan tendint a variar aquest flux de treball per poder introduir nous productes o serveis com és el dossier de premsa. Així el seu flux de treball quedaria modificat a:



Com podem observar només una petita variació en la manera d'organitzar les tasques diàriament permet, sense més recursos en personal o en temps, obtenir un segon producte.

Aquest canvi suposa un canvi de cultura organitzativa per a moltes institucions, però aporta en contrapartida una racionalització de la despesa i dels recursos. En bancs i en organitzacions governamentals, per exemple, aquestes dues funcions, arxiu de premsa i dossier de premsa es troben separades en departaments diferents. El requeriment més important si es pretén donar un servei global d'aquestes característiques des d'un únic departament és que el dossier de premsa per tenir sentit ha d'estar confeccionat a primera hora del matí. És per això que cal un programa suficientment àgil que permeti confeccionar els dossiers amb el mínim esforç, automatitzant fins i tot la generació de cobertes, sumaris, i la conversió de les dades en els formats que permetran la distribució electrònica (PDF per exemple) o la publicació en intranet (HTML).

Per apreciar el valor d'aquesta documentació de premsa, bé com a arxiu de premsa, bé com a dossier de premsa en format digital podem veure un últim exemple en el qual la documentació de premsa es converteix en un producte amb valor comercial. Un dels més grans arxius de premsa especialitzat en la indústria cinematogràfica és el del grup Kirch, que s'encarrega de distribuir pel·lícules a les televisions i als cinemes. L'any 1994 es va posar en marxa un projecte d'elaboració electrònica de retalls de premsa i de confecció de dossiers de premsa a mida. El resultat és que actualment els comercials de Kirch fan servir els retalls de premsa en les seves negociacions amb possibles clients com a prova comercial.

Estat de la matèria o Què li hem de demanar a un programa per digitalitzar premsa?

Val la pena dedicar una mica de temps a observar quins programes estan disponibles al mercat, comparar les seves funcionalitats, la seva flexibilitat, la seva adaptabilitat a la tasca que volem fer. Aquesta és la nostra tasca professional, i encara que per raons òbvies no compararem els programes disponibles, sí que val la pena remarcar quines serien les característiques bàsiques que hauria de complir un programa per digitalitzar premsa:

- a) el programa ha de ser estàndar i obert. És molt més interessant que el programa permeti triar lliurement tots els components de l'arquitectura. D'aquesta manera, el software de captura (digitalització) escollit ens ha de deixar triar el maquinari que ens interressi (escàners, sistemes d'emmagatzematge) sense problemes de compatibilitat, i la base de dades que més s'adapti a les nostres necessitats.
- b) Tant els formats dels fitxers que generi l'aplicació com la comunicació que es farà amb la base de dades o altres sistemes ha de ser estàndard. En el cas dels fitxers estariem parlant de formats extensos com TIFF, PDF, ASCII, ANSI, HTML, etc., i en general hem de poder triar segons les necessitats en quin format volem produir la digitalització de la notícia. La comunicació amb la base de dades haurà de respectar l'estàndard industrial per ser compatible amb les bases de dades de mercat: l'ODBC.
- c) No hi ha cap arxiu de premsa que sigui igual a un altre. Els fluxos de treball són diferents, els productes també, per tant hem de demanar que el programa sigui fàcilment adaptable. És per això recomanable que es tracti d'un programa modular i que no necessiti gaire programació (si no hem de tocar el codi millor; avui en dia això es soluciona fàcilment amb *scripts*).
- d) El programa escollit hauria de poder aprofitar els desenvolupaments d'altres empreses especialitzades en temes de OCR, visualitzadors d'imatges, controladors per als escàners, interfícies amb les bases de dades, etc.

- e) El programa ha de treure el màxim profit de la tecnologia per reduir al mínim les tasques manuals i rutinàries.

Serveis de premsa: un vestit a mida o confecció industrial?

En la nova economia, després de la defunció de l'estat del benestar es tendeix a reduir les inversions en favor de contractació de serveis que derivin a la despesa corrent. D'aquesta manera s'espera reduir les despeses de personal i la d'infraestructura, pagant en contrapartida per un servei. Aquesta és la justificació que explica que en Espanya existeixin més de trenta empreses de seguiment de premsa, tant a nivell local com a nivell estatal, un número incomptable de serveis de marketing i relacions públiques que ofereixen addicionalment aquest servei. És una dada a valorar que de totes aquestes petites i mitjanes empreses espanyoles només quatre formen part de la FIBEP (Federation Internationale des Bureaux d'Extraits de Presse) FIBEP, amb seu a Zurich, es va fundar el 1953 i està formada per empreses que es dediquen a l'observació de medis i el processament de continguts.

Les empreses de monitorització de premsa ofereixen als seus clients diferents tipus de servei, relacionats amb el tipus de seguiment demanat i al pressupost. La seva rendibilitat es basa en què poden servir a diferents clients una mateixa notícia. És un servei prou útil i còmode per a empreses que no volen fer el seguiment de premsa elles mateixes sinó que prefereixen rebre la informació. Una de les principals queixes que tenen els usuaris d'aquests serveis és la manca d'exhaustivitat –no sempre seleccionen totes les notícies que els interessa– i a la vegada la manca d'especificitat.

El mercat europeu dels serveis de premsa està actualment en ebullició, precisament per les noves possibilitats de negoci que presenten les eines electròniques. En el mercat centroeuropeu, però també a nivell mundial, i comença tímidament a Espanya on algunes empreses de serveis ja formen part de grans grups de comunicació europeus, s'està tendint a una concentració d'empreses per abraçar el màxim de mercat. Caldrà veure quina serà la repercussió d'aquests moviments de mercat en els seus clients.

Un cop d'ull al futur de la documentació de premsa

Estem vivint a la societat de la informació, concepte que tenen clar tant documentalistes com usuaris. El quart poder, com se li anomena a la premsa, avui es desenvolupa fortament en tots els entorns: escrits, audiovisuals, etc.

Si en la documentació de premsa els documentalistes s'enfrontaven a la selecció i documentació de la premsa impresa, avui han de tractar també els continguts multimèdia encara en quantitats més elevades. En particular el desenvolupament del WWW ha contribuït al fet que la publicació de qualsevol tipus de contribució, informe, missatge de qualsevol és possible, en una immensa varietat i qualitat.

Les empreses editorials de premsa impresa compten en la seva totalitat amb una versió digital, el que suposa una duplictat de la capacitat d'informació. A més existeixen nous productes digitals sense versió impresa, els *Ezines* publicats només a la web. Els editors clàssics de mitjans impresos completen les seves publicacions amb continguts multimèdies. El creixement de documentació de premsa en tots els formats és exponencial. Aquest punt representa un problema per als centres de documentació per dos motius; a) encara que la indexació intel·lectual és necessària, els documentalistes no compten amb eines reals que augmentin la seva capacitat de producció, i b) els arxius de premsa, especialment a institu-

cions sense ànim comercial com les universitats, pateixen d'una malaltia crònica: pressupostos molt reduïts.

Aquesta és una solució molt delicada perquè encara no s'ha trobat una solució satisfactòria que permeti introduir el flux d'informació d'una manera eficient, i efectiva. La tecnologia no ha sigut capaç encara de donar les eines adequades per poder suportar al documentalista en la seva feina d'indexació, i no parlem de substituir-lo. En aquest sentit la indústria tendeix actualment a oferir eines com *Autonomy*, de l'empresa anglesa del mateix nom i *SERbrainware* de l'empresa alemanya *SER AG*.¹⁶ Aquestes són eines que pretenen anar un pas més enllà en l'anàlisi intel·ligent dels textos i els sistemes de recuperació. Els dos sistemes mantenen els anomenats «documents sense estructura» i són capaços de processar-los i classificar-los automàticament. *Autonomy* encara va una mica més enllà i detecta i extrau conceptes claus dels articles, i és capaç de produir¹⁷ resum de manera interactiva (*on the fly*). L'aplicació d'aquests sistemes, des del nostre punt de vista, no ha de ser la substitució del documentalista sinó el seu suport. Eines que facin la seva tasca més àgil.

Com hem vist per a la documentació de premsa, encara que no moltes, ja existeixen eines que permeten automatitzar fins a un cert grau la digitalització de la documentació i la seva recuperació dins d'un sistema electrònic. Però si observem els altres mitjans informatius com vídeo i audio veurem que tota la feina és absolutament artesanal. L'extracció d'informació –l'equivalent del retall de premsa– de vídeo i d'àudio és fa avui en un 100 % de forma manual, fins i tot en les grans empreses de seguiment de mitjans. I aquesta és una situació asfixiant, especialment per als centres de documentació de mitjans de comunicació. Pensem que la quantitat d'informació generada en mitjans audiovisuals ha crescut en els últims anys gràcies al naixement de les cadenes privades, i amb l'aparició de la televisió digital aquesta situació encara s'intensificarà més. És per aquesta raó que simultàniament en diferents universitats europees s'estan posant en marxa projectes –d'alguns ja amb resultats prou positius– per aconseguir una extracció d'informació automàtica des de mitjans audiovisuals.

Hi ha un últim punt que no hem considerat encara: existeix tecnologia disponible al mercat per treballar amb la documentació de premsa d'una forma electrònica i alguns arxius de premsa estan ja introduint aquestes eines, però què farem amb les velles col·leccions en paper, microfilm, en vídeo i en cintes d'àudio? Haurem de viure amb sistemes híbrids de paper i bases de dades documentals?

La digitalització d'aquestes col·leccions retrospectives no hauria de ser un problema, encara que s'han de considerar tots els inconvenients que comporta. Cal que els OCR millorin encara per permetre l'entrada automàtica –sense intervenció manual– del text complet i que l'emmagatzematge deixi de ser un problema. L'inconvenient més gran amb el que es presenta un sistema de retrodigitalització és la indexació d'enormes quantitats d'informació, així com la seva edició per a un accés òptim. Molts dels projectes que s'estan portant a terme actualment no consideren, i no garanteixen ni un aspecte ni l'altre.

És per això interessant mencionar els projectes d'investigació que estan treballant en aquesta línia –fer possible la indexació de grans volums, de qualitat, i produir objectes de qualitat– tant en els Estats Units com a la Unió Europea. Val la pena mencionar el projecte *Metae*, subvencionat per la CE¹⁸ que pretén desenvolupar un programa capaç no només d'analitzar automàticament l'estructura i la gramàtica de diferents tipus de publicació com lli-

16. <<http://www.ser-brainware.com>>

17. Tal i com hem pogut comprovar en l'aplicació de l'empresa MyNews

18. <<http://metae.uibk.ac.at>>

bres i diaris, sinó també d'extreure les característiques d'indexació rellevants com a metadada.

Però aquesta és la nostra visió avui. Potser demà ja no necessitarem res de tot això. Potser l'ocupació del documentalista s'està quedant obsoleta perquè tothom creu que pot trobar la informació. Potser els diaris ja no existiran com ara, sinó que rebrem de forma personalitzada les notícies que ens interessin. I els sistemes definiran els nostres interessos segons les nostres pautes de comportament. Potser no necessitarem un pont entre la informació impresa i les bases de dades sinó que els productors de notícies ens faran arribar la informació rellevant directament al cervell. Potser? Però segurament per a tots aquests desenvolupaments encara falta una mica, i ens calen els documentalistes, ànimes de la documentació de premsa. *Màtrix* us saluda.