



Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals

Ferran Adell

Professor del màster d'Edició digital
UOC

Data de presentació: abril de 2015

Data d'acceptació: maig de 2015

Data de publicació: juny de 2015

Resum

En aquest article es planteja la problemàtica associada a la creació de nous objectes culturals en l'ecosistema digital. Es defineix què és i com ha de ser un llibre digital i quines són les característiques indispensables perquè aquest segueixi essent un contenidor vàlid per al coneixement. El marcatge semàntic, indispensable per a la identificació dels elements que formen un llibre, i la seva catalogació, indexació i recuperació en forma de recerques avançades, així com la definició del procés d'enriquiment dels llibres digitals completen una revisió teòrica de com ha de ser el procés de digitalització i enriquiment dels llibres. A més a més, es parla de la necessitat de diferenciar el tractament dels continguts textuals extensos de la creació de nous mitjans i nous objectes culturals.

Paraules clau

edició digital, llibres digitals, EPUB, marcatge semàntic, llibres enriquits

Hybridisation, Semantic Markup And Enrichment Of Digital Books

Abstract

This paper discusses the issues associated with the creation of new cultural objects in the digital ecosystem. It defines what is and what should be a digital book and what are its essential characteristics so that it continues to be a valid container for knowledge. Semantic markup, indispensable for identifying elements in a book, and its cataloguing, indexing and retrieval as advanced searches, as well as defining the enrichment process of digital books complete a theoretical review of how the digitisation and enrichment of books should be done. Furthermore, it also discusses the need to differentiate the treatment of extensive textual content from the creation of new media and new cultural objects.

Keywords

digital publishing, e-books, EPUB, semantic markup, enriched books



<http://digithum.uoc.edu>

Els nous mitjans són, *per se*, modulars i reproduïbles; Manovich (2001) ho explicà molt bé a *The language of new media*. La codificació numèrica fa que sigui molt senzill multiplicar les còpies d'un contingut, les versions digitals existents i els llocs d'emmagatzematge. Els antics suports físics, com ara la cinta magnètica, el paper o els discos compactes han deixat pas a un nou sistema informacional que ja no depèn –necessàriament– d'un suport físic en concret i que permet una gran diversitat de formats de codificació diferents.

Fins al salt cap a l'emmagatzematge digital, els formats en els quals es guardaven els continguts depenien de la tipologia d'aquests i no de la voluntat del creador. És a dir, el procés de creació d'un disc en vinil, independentment de si l'origen del contingut és digital o en cinta magnètica, és un procés mecànic amb un únic format final. Avui, en canvi, quan creem contingut musical amb eines digitals, hem d'escollir el format final de sortida; d'aquesta decisió en dependrà no només la qualitat final de reproducció, sinó el tipus i el nombre de dispositius que el podran reproduir. Amb el vídeo, les imatges i els llibres digitals succeeix exactament el mateix: les opcions a l'hora d'escollir el format, la codificació o la compressió són múltiples.

El suport analògic condiciona el format d'enregistrament dels continguts, no només perquè obliga a procedir d'una manera determinada per a la creació sinó perquè així assegura, de retruc, una universalització del format, limitat i exclusiu, però universal.

La popularització de les tecnologies digitals i l'entrada massiva d'usuaris amb intencions creatives a internet han comportat una obertura en l'accés a les eines i un canvi radical en tots els nivells de producció, difusió i consum d'objectes culturals. Per primera vegada, amb l'aparició de la web 2.0, el consumidor esdevé creador, participant actiu (prosumidor): deixa enrere un estat passiu de consum per endinsar-se en un escenari en el qual les noves formes d'interacció amb els mitjans digitals, l'evolució de la interfície i l'abaratiment i millora de l'accés a la xarxa li permeten participar en els processos creatius des de la comoditat del dispositiu personal.

L'aparició de noves eines ha generat una cursa per a la imposició de nous formats, codificacions d'arxiu i plataformes per a la creació de continguts, amb la consegüent multiplicació dels formats en els quals s'emmagatzema i es distribueix la producció cultural. Així, es canvia el paradigma dels formats únics associats a la tipologia del contingut per un entorn en el qual cada empresa, actor o agent implicat desenvolupa els seus propis formats privats per a potenciar noves oportunitats de negoci vinculades a aquests.

Noves narratives i hibridació

La publicació de l'enciclopèdia Encarta¹ per part de Microsoft va suposar per als usuaris no experts una forma diferent de veure la

Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals

digitalització dels continguts. No és que no existissin alternatives que oferissin propostes similars, sinó que el seu impacte segurament era molt més local. Les grans editorials oferiren productes semblants més o menys enfocats a temàtiques concretes i, ràpidament, el fenomen multimèdia aplicat a la difusió de la informació va arribar a l'usuari, que aleshores encara no era conscient de les potencialitats de la xarxa. Aquells productes multimèdia, que combinaven text, àudio i vídeo, van ser les primeres propostes de *fusió* en un mateix vehicle comunicatiu de diferents tipus de mitjans fins aleshores independents. I no va ser fins a la popularització de l'accés a internet, amb la tarifa plana, que es va desplaçar aquest tipus de continguts a la xarxa a favor d'un nou espai obert i en constant revisió, que oferia moltes més possibilitats comunicatives i un camí clar cap a la hibridació dels mitjans.

Aquest procés de convergència dels diferents tipus de mitjans en els mateixos espais de reproducció, així com el desenvolupament de les tècniques de programari independents del mitjà, permeten una millora de l'experiència narrativa i comunicativa i, a més a més, possibiliten una optimització del consum de recursos de la màquina, molt important per als dispositius mòbils.

Què és un llibre digital?

Els llibres són contenidors d'informació que tenen una forma determinada, un objectiu i unes característiques bàsiques. Constitueixen un ecosistema propi del qual depèn la conservació i la transmissió de la cultura i la vertebració del corpus del coneixement. Per bé que la dita sosté que una imatge val més que mil paraules, el fet és que no podem emmagatzemar en imatges l'avenç científic, la reflexió filosòfica o l'estudi de la cultura contemporània.

Certament, de llibres n'hi ha de tot tipus; però, tanmateix, tots comparteixen unes característiques comunes que és important reivindicar i conservar. Els llibres tenen una narrativa establerta; fins i tot en els populars títols encapçalats pel «tria la teva pròpia aventura», que entrecreen trames, les diferents històries són tancades, amb un discórrer fixat: tenen una introducció, un nus i un desenllaç predefinit per l'autor. Cal posar de relleu que la naturalesa narrativa d'aquests continguts no obeeix a les limitacions de l'entorn analògic o de les condicions imposades pel format imprès. Si bé el paper no permet la interacció que ens ofereixen els nous mitjans digitals ni enllaçar a continguts externs de forma directa, l'extensió d'una història, la seva estructura bàsica i la seva formalització depenen de l'autor. Així doncs, la capacitat per al tractament profund de temàtiques diverses i la construcció d'històries en les quals la imaginació té un paper protagonista no depenen del suport.

De la mateixa manera que ningú que faci una anàlisi acurada podrà concloure que una pàgina web és la versió digital d'un diari,

1. http://es.wikipedia.org/wiki/Microsoft_Encarta.



<http://digithum.uoc.edu>

o un videojoc d'aventures un objecte cultural equiparable a una pel·lícula del mateix gènere, no és sensat afirmar que qualsevol objecte digital en el qual el text tingui un paper protagonista és un llibre digital. Una cosa és que els nous mitjans siguin reformulacions dels antics i una altra, ben diferent, és que l'objecte fruit de la nova formulació sigui el mateix que l'objecte de nou format digital.

La definició curosa dels conceptes és de vital importància per a la construcció d'un marc teòric sostenible i una anàlisi tecnològica que permeti la teorització i la construcció de sistemes de conceptes, encara que les relacions que s'hi estableixin siguin relatives i amb una alta tendència a la fugacitat. No fer-ho significa seguir potenciant una filosofia de la tecnologia caòtica i que queda ràpidament obsoleta.

En aquest ecosistema inestable en el qual els conceptes encara estan en formació i discussió acadèmica, la definició del què ha de ser, com ha de ser i quina ha de ser la finalitat d'un llibre digital requereix una atenció especial. El que hi ha en joc en la digitalització del llibre, com a objecte cultural, és la conceptualització d'un format universalment accessible i indexable capaç de mantenir la relació que en el món analògic s'establí entre el llibre i el corpus del coneixement. El curs del llibre digital ha passat desapercbut durant molts anys. La facilitat dels ordinadors per a gestionar el text en format nu ha endarrerit la digitalització del llibre, mentre que altres mitjans culturals ja fa temps que s'han adaptat i expandit en l'ecosistema digital.

Per tant, la migració digital del llibre es complica perquè l'entorn en el qual ha d'instal·lar-se està saturat de nous objectes culturals: videojocs, narratives interactives, aplicacions que simulen els llibres, nous espais per a la difusió de continguts textuals o l'herència dels antics formats. La fusió dels mitjans, la capacitat de processar música, text, imatge i vídeo en un mateix entorn reproductiu, així com l'expansió de les possibilitats d'interacció amb els dispositius, dibuixen un paradigma que tal vegada podria qüestionar el sentit del llibre digital com a objecte, el protagonista principal del qual sempre és el text.

No obstant això, més enllà de la validesa de la història narrativa textual que treballa en la nostra imaginació, que per part de l'autor no és qüestionable, el llibre digital ha de garantir que les generacions que més informació estem generant siguem capaces de transmetre-la, de forma ordenada i sostenible, a la societat del futur. No es tracta només de garantir que el coneixement es distribueixi àmpliament gràcies al potencial tecnològic, sinó d'assegurar que el gran volum d'informació que generem contínuament, i de forma caòtica, esdevingui un corpus accessible i estructurat en funció del nostre *logos*, i no de la lògica binària.

El valor semàntic d'un contingut i de les parts que el constitueixen –malauradament, els grans desconeguts per al sistema de gestió de la informació que hem construït a l'entorn digital– necessiten una potenciació de la web semàntica i dels seus requisits per a començar a construir un llarg i difícil pont que ens permeti arribar fins a un sistema de gestió de la informació molt més proper

Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals

al llenguatge natural. Convé que sigui un procés que no només atengui a les necessitats i als requeriments de la gestió dels grans volums de dades, sinó que s'erigeixi com a base per a l'ordenació del coneixement.

I perquè aquest procés sigui possible, hem d'allunyar els llibres de les aplicacions informàtiques i hem d'associar el text a un format universalment accessible, multisistema i multidispositiu. En el cas dels continguts textuals aquest format és l'EPUB.

A hores d'ara ja no cal fer presentacions de l'EPUB: el format ha guanyat pes i protagonisme en el sector editorial i, a poc a poc, amb molt desconeixement tècnic i encara amb poca atenció a les seves potencialitats semàntiques, s'ha fet popular. Aquesta popularització, però, ha anat acompanyada d'un elevat grau de confusió en l'ús dels conceptes i en la definició de l'edició digital. La gestió de metadades que incorpora el format crea un sistema per a l'afegit dels paràmetres associats a cada exemplar (títol, autor, editorial, etc.), però hi suma els avantatges de l'etiquetatge semàntic amb XHTML adaptat a l'edició digital. Aquest marcatge semàntic avançat, més detallat, que no atén només a les característiques bàsiques d'una obra escrita, és de vital importància per a la construcció d'un sistema digital de gestió del coneixement proper al llenguatge natural, i que permeti respondre a les necessitats socials de documentació, investigació i revisió enfront d'un creixement exponencial de la informació.

Els llibres, a més, no només són text: poden incorporar imatges, diagrames, informació tècnica o audiovisuals que també han de ser classificats. Per tant, no únicament és important l'etiquetatge semàntic quant al contingut textual, sinó que tots els elements que formen un llibre han de ser identificats per a determinar-ne el paper en el conjunt narratiu. Els llenguatges de marcatge semàntic permeten identificar les diferents tipologies de contingut aplicant-hi un tractament particular si així ho requereixen.

Ara bé, perquè això sigui possible, és a dir, perquè un algoritme informàtic pugui accedir al contingut d'un llibre i el marcatge semàntic tingui sentit, la codificació del llibre ha de ser transparent i accessible, ha de permetre una interacció avançada amb els sistemes complexos de gestió de la informació. El que invalida les aplicacions informàtiques, *appbooks*, videojocs i la majoria dels objectes digitals programables com a contenidors del coneixement, és aquesta falta de transparència, aquesta dificultat d'accés als continguts textuals que en formen part sense descompilar, processar o transcodificar el codi informàtic. Però no sols hem de parlar de transparència per a justificar que el contingut textual tingui un format propi que respongui correctament a les exigències multisistema i multidispositiu; també és indispensable perquè la seva construcció sigui escalable i progressiva. Tot plegat fa que l'EPUB sigui ara com ara l'únic format vàlid per a recollir tot el bagatge dels llibres –enriquets o no– en l'entorn digital.

La naturalesa tancada de la narrativa dels llibres no és una característica suplementària de la qual puguem prescindir. Si definim entorns en els quals la interacció i la presa de decisions de l'usuari



<http://digithum.uoc.edu>

són protagonistes, transcendents i canvien el curs del relat definit pels creadors, aleshores és molt difícil que puguem definir-los com a llibres. De la mateixa manera que no és el mateix gravar una pel·lícula per al cinema que pintar un quadre a l'oli, no és el mateix programar un videojoc, o crear una aplicació qualsevol, que crear un llibre digital.

Un llibre digital és un arxiu; normalment es tracta d'un arxiu comprimit que integra altres arxius creats amb llenguatges de marcatge i estil. Per contra, una aplicació és un programa. Un videojoc és una aplicació; una *appbook* és una aplicació; el Libreoffice Writer amb el qual escrivim aquest text és una aplicació; i fins i tot el sistema operatiu que fa funcionar qualsevol dispositiu conté un gran nombre d'aplicacions preinstal·lades. Una aplicació, tot i que amb la popularització dels telèfons intel·ligents i les tauletes tàctils s'ha associat només a aquest nou mitjà, és qualsevol programari informàtic que fa una funció concreta. Vegeu-ne la definició que recull la Wikipedia:

Application software (an application) is a set of computer programs designed to permit the user to perform a group of coordinated functions, tasks, or activities. Application software cannot run on itself but is dependent on system software to execute. Examples of an application include a word processor, a spreadsheet design and management system, an aeronautical flight simulator, a console game, a drawing, painting, and illustrating system, or a library management system.

Wikipedia – Application Software

Un creador que prengui de referència un llibre i, basant-se en la seva narrativa, creï una història interactiva, estarà programant una aplicació. Si versiona el conte de la Caputxeta Vermella i el transforma en un videojoc en el qual és possible que sigui l'àvia la que es menja el llop, estarà creant una aplicació. Si adapta textos de les obres clàssiques en forma d'aventura gràfica, estarà creant un videojoc, és a dir, una aplicació. El primer pas per a construir un ecosistema sostenible del llibre a l'entorn digital és diferenciar, de forma clara, els llibres digitals de les aplicacions, siguin del tipus que siguin.

Llibres digitals enriquits

Finalment, cal subratllar la importància de diferenciar què és un llibre digital enriquit –en una traducció no exacta del terme *enhanced* anglès, que probablement encaixa millor amb el concepte de millora en la nostra llengua– del que no ho és. Els llibres enriquits són aquells que incorporen recursos suplementaris per a ampliar l'experiència lectora. Així, aquests recursos amplien el contingut del llibre sense que, tanmateix, aquest procés d'enriquiment alteri

Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals

la narrativa central del llibre. És a dir, un usuari que llegeix un llibre enriquit, si no pot (perquè un determinat dispositiu de lectura no li ho permet) o no vol visualitzar els recursos suplementaris, ha de poder seguir tenint accés complet al significat del contingut. Per tant, el procés d'enriquiment ha de ser creat tenint en compte el que s'anomena *fallback*, una de les propietats que EPUB demana per a complir els estàndards del format quan afegim interactivitat (*scripts*) a un llibre digital:

EPUB 3 adopts a progressive enhancement approach for scripted content, whereby scripting must not interfere with the integrity of the document (i.e., must not result in information loss when scripting is not available). Consequently, although documents that do employ scripting may provide fallbacks to further facilitate access to their contents, the documents must be accessible without them.

EPUB3, Recommended Specification²

Aquest principi que defineix l'especificació del format estàndard per als llibres digitals, basat en una recerca de la màxima accessibilitat possible al contingut d'un EPUB, és extrapolable a qualsevol recurs que s'afegeixi a un llibre digital. Quan afegim imatges a un contingut textual, hem de garantir que informem el sistema del títol i el valor de la imatge per a aquells lectors que no puguin visualitzar-los. Si afegim una peça musical hem d'indicar al sistema quina peça és i, si és possible, donar informació sobre el seu sentit i la seva funció. Si afegim recursos audiovisuals ho hem de fer de manera que el lector sigui capaç de seguir entenent el contingut del llibre si no pot visualitzar-los; convé, doncs, aportar informació alternativa o, fins i tot, facilitar-ne la subtitulació en format textual. Només seguint aquests procediments podem enriquir veritablement els llibres digitals, millorar-los, sense provocar una fractura en l'accessibilitat als continguts.

Els llenguatges que hi ha implicats en la creació d'una aplicació per a un dispositiu mòbil no tenen res a veure amb els que s'usen per a crear un llibre digital. En el primer cas, hem de programar-los tenint en compte el sistema operatiu de la màquina a fi d'obtenir accés a determinats recursos i permetre una interacció activa i continuada de l'usuari amb l'entorn, no només a un nivell bàsic de funcionament, sinó en un entramat complex que determina el comportament d'un programari concret. Quant als llibres, el que instal·lem és el lector, l'aplicació que permet accedir als continguts d'un EPUB. Ara bé, en aquest cas els llenguatges no han d'interactuar amb els recursos de la màquina, atès que aquesta tasca ja l'executa el programari que s'encarrega d'interpretar i de donar forma al contingut de l'EPUB.

Amb la desaparició de les limitacions de la segona especificació del format, a favor de l'ús de HTML5 i els seus avantatges pel

2. <http://www.idpf.org/epub/30/spec/epub30-overview.html#sec-access-scripting>.



<http://digithum.uoc.edu>

Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals

que fa al marcatge semàntic i a la seva capacitat nativa de reproducció audiovisual, l'EPUB ens ofereix un entorn de creació de llibres digitals enriquits que, alhora que garanteix una conservació eficient del coneixement i una indexació semàntica avançada, permet experimentar amb l'enriquiment dels llibres i l'ampliació de l'experiència lectora a la llum de les noves possibilitats que ens brinden els entorns digitals.

Referències bibliogràfiques

- LENNON, J. (2015). "To Markup or to Markdown?". *Digital-bookworld.com* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://www.digitalbookworld.com/2015/to-markup-or-to-markdown/>>
- MYER, T. (2005). "A Really, Really, Really god introduction to XML". *Sitepoint.com* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://www.sitepoint.com/really-good-introduction-xml/>>
- ELOFF, K.; BORGSTROM, L.; CASSELLS, L. (s.d.). "An introduction to mark-up languages". *PUB310. Publishing in the*

- Digital Environment* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://pub310.pressbooks.com/chapter/3-1-markup-languages-proving-why-you-should-know-html/>>
- BLATNER, D. (2014). "What's the Difference Between EPUB, DPS, and PDF?". *InDesign Secrets* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://indesignsecrets.com/difference-epub-dps-pdf.php>>
- MANOVICH, Lev (2001). *The language of New Media*. Cambridge, Mass: MIT Press.
- WIKERT, J. (2012). "HTML5, EPUB 3, and ebooks vs. web apps". *Tools of Change for Publishing* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://toc.oreilly.com/2012/09/html5-epub-3-and-ebooks-vs-web-apps.html>>
- WIKERT, J. (2012). "EPUB 3 facts and forecasts". *Tools of Change for Publishing* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://toc.oreilly.com/2012/10/epub-3-facts-and-forecasts.html>>
- WIKERT, J. (2012). "Ebooks as native apps vs. web apps". *Tools of Change for Publishing* [artículo en línea]. [Fecha de consulta: 1 de julio 2015]. <<http://toc.oreilly.com/2012/10/ebooks-as-native-apps-vs-web-apps.html>>

CITACIÓ RECOMANADA

ADELL, Ferran (2015). «Hibridació, marcatge semàntic i l'enriquiment dels llibres digitals». *Digithum*, núm. 17, pàgs. 3-8. UOC. [Data de consulta: dd/mm/aa] <<http://journals.uoc.edu/index.php/digithum/article/view/n17-adell/n17-adell-pdf-ca>> <<http://dx.doi.org/10.7238/d.v0i17.2608>>



Els textos publicats en aquesta revista estan subjectes –llevat que s'indiqui el contrari– a una llicència Reconeixement 3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los, comunicar-los públicament i fer-ne obres derivades sempre que reconegueu els crèdits de les obres (autoria, nom de la revista, institució editora) de la manera especificada pels autors o per la revista. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by/3.0/es/deed.ca>.



Ferran Adell

Professor del màster d'Edició digital (UOC)
ferran.adell@gmail.com

Ferran Adell és llicenciat en filosofia i expert en Tecnologia i Societat. És professor del Màster d'Edició digital i del Grau de Multimèdia de la Universitat Oberta de Catalunya, i de diversos programes acadèmics d'Elisava, Escola Superior de Disseny i Enginyeria de Barcelona.

Les seves especialitats en l'àmbit de la informàtica són la gestió de sistemes i la creació de nous mitjans digitals. L'ús de les tecnologies en l'educació i en la creació cultural i, especialment, la incorporació de videojocs en entorns d'aprenentatge i les noves narratives digitals són els temes centrals de la seva recerca.

Més enllà d'aquesta activitat acadèmica, desenvolupa projectes de *digital manager* per a empreses. Realitza tasques de programació i gestió web, dissenya estratègies de màrqueting digital, edita llibres digitals en format EPUB, i materialitza creacions de noves narratives augmentades. També ofereix suport tècnic per gestionar els canvis en la forma d'accedir, classificar i conservar el coneixement en l'entorn digital.

Estudis d'Arts i Humanitats
Av. Tibidabo, 39-43
08035 Barcelona

