



Web i cinema. Anàlisi comparativa de dues bases de dades per a la recerca en línia

Lluís Codina i Maria del Valle Palma



1. La web invisible

A mesura que augmenta el volum d'informació disponible a través de la WWW es fa patent la paradoxa principal de la informació: la seva abundància és un dels obstacles principals per al seu consum.

Com pot ser això? Malgrat que sovint es diu que la informació és immaterial, el cert és que no només no és immaterial sinó que el seu processament està sotmès a les limitacions més materials que sigui possible imaginar i, com a conseqüència, la capacitat de processament de la informació dels éssers humans resulta estar tan sotmesa a restriccions com qualsevol altre procés material, sigui córrer els 100 metres, escriure un poema o talar arbres.

Davant de la caòtica realitat d'Internet, per una banda, i de les restriccions reals de la cognició humana per al processament de la informació, per l'altra, no és estrany que cada cop sigui més important la (mal) anomenada Internet o web invisible, és a dir, aquella part de la web el contingut de la qual és inaccessible als motors de cerca i que, en canvi, proporciona les informacions més contrastades i de més qualitat.

2. Bases de dades i metodologia

En aquest treball presentarem l'esbós d'una anàlisi comparativa de les dues millors bases de dades de cinema que podem trobar a la web. Es tracta de les següents:

- *All Movie Guide* (AMG, a partir d'ara).
- *Internet Movie Database* (IMDB, a partir d'ara).

Tant AMG com IMDB són bases de dades que contenen descripcions de films i de cineastes estructurades en forma de registres. Aquests registres aporten nombroses dades de cadascuna d'aquestes dues entitats (films i registres) que poden resultar de gran valor per als estudiosos.

Aquestes dades van (vegeu les figures 1 a 6) des de les dades més elementals, com el títol i el realitzador de cada film, fins als premis i les recaptacions obtingudes, passant per les fitxes tècniques i artístiques, la ubicació dels rodatges, l'existència o no de segones versions, les frases promocionals que es van fer servir a la seva estrena i, si és el cas, errors de continuïtat, errors factuais o contradiccions dels guions, etc.

A més a més, els registres contenen diverses informacions de valor afegit, com sinopsis dels films, biofilmografies dels cineastes, paraules clau, estimacions del valor històric i artístic dels films, etc. Per últim, tant AMG com IMDB (però especialment aquesta última) presenten moltes facilitats de consulta, de manera que és possible realitzar recerques a partir d'un o més conceptes o paràmetres.

Amb aquest treball pretenem d'una banda, contribuir al millor coneixement de dues eines de treball d'incalculable valor per als estudiosos, els professionals i els investigadors del món del cinema, com són les dues bases de dades mencionades. I d'altra banda, pretenem aportar criteris per estimar el valor i la qualitat relativa de cadascun d'aquests dos recursos digitals i, per tant, aportar elements per respondre a preguntes com aquestes : n'hi ha prou mde consultar una de les dues bases de dades o cal que les consultem totes dues davant una recerca determinada?; aporten les dues les mateixes informacions?; hi ha diferències en el tractament qualitatiu de la informació?; quines són les diferències qualitatives entre elles?, etc.

Finalment, pretenem aportar uns primers elements per a una metodologia d'anàlisi d'aquesta classe de recursos digitals. El que es presenta aquí no és més que una proposta que en cas que es demostrés que és útil encara s'hauria d'ampliar i refinar amb elements addicionals. De moment ens conformem amb mostrar alguns dels paràmetres que es podrien incloure en aquesta metodologia, així com alguns dels resultats que es podrien obtenir.

Per a la nostra anàlisi, per tant, hem comparat el tractament descriptiu dels registres de films i de cineastes, i hem comparat el nombre de camps que tenen tractament documental. També hem examinat les estadístiques bàsiques de cada base de dades i hem efectuat proves de rendiment aplicant quatre tests de recuperació d'informació. Finalment, hem aplicat dos tests qualitius per tal de comprovar la capacitat relativa de proporcionar informació de context o informació estratègica de cadascuna de les dues bases de dades.

3. Anàlisi

AMG (vegeu les figures i taules corresponents) proporcionava, a l'octubre del 2000, informació sobre més de 180.000 films i més de 246.000 cineastes del tot el món. De les dues bases de dades analitzades, no és la que posseeix més quantitat d'informació, però en canvi és la que fa un millor tractament documental de la informació, és a dir, proporciona més informació de valor afegit.

Això últim és especialment palès quan comparem el nombre de camps que tenen tractament documental - com paraules clau, gènere, sinopsis, etc.- en els seus registres, que és de dotze en el cas d'AMG i de només tres en el cas d'IMDB.

És a dir, a AMG hi ha dotze camps que no contenen una mera transcripció de dades extretes de *pressbooks* o d'altres fonts d'informació, sinó que per poder proporcionar aquestes informacions els col·laboradors i redactors d'AMG n'han hagut de realitzar alguna classe d'anàlisi intel·lectual. El cas paradigmàtic són les paraules clau (temàtiques, geogràfiques, cronològiques i onomàstiques), les línies argumentals, els estils artístics, les sinopsis i la crítica. Com a resultat, en teoria hi ha moltes més possibilitats de realitzar cerques de tipus heurística en AMG que a IMDB.

A més a més, AMG destaca per la qualitat i l'eficiència en la presentació d'aquestes informacions. Els dos tests qualitius, que hem anomenat "test Polonsky" i "test Dracula", i que discutirem més endavant, ho deixen ben clar.

Per la seva banda, IMDB destaca, sobretot, per tres coses: (1) el seu gran nombre d'informacions; (2) les seves esplèndides opcions de cerca i obtenció d'informació, i (3) el que podríem dir les seves informacions complementàries, com per exemple, informacions sobre bandes sonores o sobre versions alternatives dels films (censura, muntatge del director, versions per a televisió, etc.).

En el primer àmbit, IMDB presenta informació sobre més de 258.000 títols i sobre més de 864.000 cineastes. Quant a les seves opcions de cerca, només direm que són tantes i tan nombroses que IMDB ha creat una imaginària *IMDB University*, de la qual surt llicenciat o doctorat qui ha superat els diversos tutorials en línia que cal practicar si es volen conèixer totes les possibilitats d'IMDB.

Finalment, IMDB és una autèntica mina d'informacions complementàries al voltant dels films i dels cineastes. Des de detalls tècnics del rodatge fins a les frases utilitzades en la promoció, passant pel pressupost del film, la llista detallada de les composicions de la banda sonora, webs relacionades, etc.

Amb tot, all`en què probablement IMBD destaca sobre AMG és sobre el perfecte funcionament de totes les opcions. mentre que a AMG són freqüents els missatges d'error del sistema, a IMDB totes les funcions van funcionar a la perfecció durant les dues setmanes en les quals es van fer les proves d'aquest article (primeres setmanes d'octubre del 2000).

3.1. Test Polonsky

El test Polonsky parteix del següent: com és sabut, Abraham Polonsky, quan tenia 38 anys, va dirigir un film considerat una de les joies del cinema negre: *Force of Evil* (1948). Ara bé, malgrat l'èxit obtingut, tant de crítica com de públic, Polonsky no va tornar a dirigir un altre film fins 1969 (*Tell Them Willie Boy is Here*). Per quina raó un realitzador jove que és capaç de fer una gran obra en un dels seus primer treballs no torna a dirigir cap film fins a 21 anys després?

El test consisteix a veure si, amb una primera revisió de les dades que proporcionen les dues bases de dades, és possible: *a)* advertir la incògnita indicada; *b)* esbrinar la raó d'aquesta incògnita. Com el lector sabrà o haurà deduït, la raó va ser que Polonsky va passar a formar part de les tristament famoses llistes negres de Hollywood a causa de la seva negativa a testificar contra els seus col·legues en el Comitè d'Activitats Antiamericanes.

Si examinem la forma en la qual presenten els seus registres les dues bases de dades podrem veure fàcilment que només AMG supera de manera òptima aquest test. Només en obrir el registre dedicat a Polonsky, i a partir de la línia número 8 de la seva biografia, pràcticament sense necessitat de fer *scroll*, trobem aquesta frase:

Unfortunately, Polonsky blacklisted in 1950 after refusing to testify before the House UnAmerican Activities Committee. He continued to work as a screenwriter, but didn't get another chance to direct until 1970 (sic) in *Tell Them Willie Boy Is Here*.

No es tracta que IMDB no proporcioni també aquesta informació, sinó que ho fa de forma molt desagregada, per la qual cosa és molt probable que a un lector desprevingut li passi desapercebuda. En concret, només si el lector té la paciència d'arribar a la secció número 11 de les 14 que formen el registre podrà llegir una informació semblant i, tot i així, sota el rètol "Trivia", nom que probablement farà, precisament, que lectura molts estudiosos prescindeixen de llegir-la.

Es pot considerar que aquest cas és anecdòtic, ja que no hi ha, després de tot, tants cineastes amb perfils biogràfics com aquest. Nosaltres, en tot cas, creiem que no és una anècdota, sinó la conseqüència lògica de la forma poc afortunada de presentar la informació que té IMBD, i que contrasta amb la forma esplèndida en què AMG sap agregar i sintetitzar la informació. Un altre test qualitatiu pot reforçar aquesta idea; és tracta del test Dràcula, que exposem a continuació.

3.2. Test Dràcula

Els fets en què es basa aquest test són els següents: com és sabut, durant uns anys el cinema de Hollywood no realitzava doblatges dels seus films destinats a la distribució a l'estranger, sinó que, quan es rodava una producció de cert pressupost, hi havia dos o tres equips d'actors, un per a cada idioma, que rodaven el mateix film en els mateixos platós de forma paral·lela (aquesta és, per cert, l'excusa argumental de *La niña de mis ojos*, de F. Trueba). Després, a cada zona idiomàtica es distribuïa la versió que li corresponia: la versió amb actors castellans a l'Amèrica Llatina i a Espanya; la versió francesa a França i al Canadà, etc.

La qüestió és que hi ha una versió de Dràcula de 1931 que es va rodar als mateixos platós (i amb el mateix equip de tècnics) on Browning va rodar la seva aclamada versió, però amb un equip d'actors en castellà. Aquesta versió va ser dirigida per George Melford, un veterà realitzador que havia fet el seu primer film l'any 1911 i que, quan va es va encarregar de rodar aquesta versió ja anava, més o menys, pel seu film número cent.

Des de la recent recuperació d'una còpia de l'anomenada "versió castellana" (*Spanish version*) d'aquest film, aquest ha esdevingut un film de culte, i entre aquells que han tingut la sort de poder-lo veure hi ha la idea quasi unànime que es tracta d'un film amb personalitat pròpia i de la mateixa qualitat, si no superior, al film de Browning.

El test aquí consistia en el següent: suposem que un estudiós del cinema fantàstic es proposa comprovar quants films s'han rodat sobre el mite del comte Dracula. Podria començar per una opció òbvia, és a dir, per comprovar quants films, en tota la història del cinema, tenen la paraula Dràcula en el títol. Si ho fa així, no obtindrà la mateixa informació a AMG que a IMDB. En la primera (AMG) és impossible que li passés desapercebut aquest film, ja que en la llista de resposta a aquesta qüestió, dels 42 títols llistats, el film que fa el vuit apareix així: <4.5 1931 Dracula (Spanish Version)> (4.5, per cert, és la qualificació sobre un màxim de 5).

En canvi, si fem la mateixa pregunta a IMDB, no hi ha res en la resposta que pugui alertar el nostre estudiós. A canvi, IMDB destaca per la seva capacitat quantitativa: proporciona 94 títols. Però cap d'aquests títols correspon a la versió castellana de Melford. No figura aquest film a la base de dades d'IMDB? Sí, però només se'n pot trobar la fitxa si ja se sap el que es vol. Gràcies a les seves superiors funcions de cerca, és possible preguntar per tots els films que tinguin la paraula *Dràcula* i que, al mateix temps, siguin de llengua castellana. Aleshores, efectivament, apareix el misteriós títol <Dracula/II>. Si obrim la fitxa, de nou cal tenir una certa paciència i avançar per l'estructura del registre per arribar a un comentari que ens aclareix tota la qüestió.

De nou hem vist com, efectivament, es poden trobar les mateixes informacions crítiques o de context en una base que en l'altra, però mentre que a AMG la informació acostuma a saltar a la vista, a IMBD només es pot adonar d'aquestes qüestions qui ja les coneix prèviament o un lector realment pacient.

4. Conclusions

La millor base de dades seria una combinació de les dues: l'eficiència de presentació de la informació d'AMG amb l'eficiència tècnica i funcional i la quantitat d'informació d'IMDB.

Això significa que, si bé per a la major part de les necessitats d'informació n'hi hauria prou amb IMDB, si volem estar segurs d'estar obtenint les informacions rellevants en cada cas és convenient no deixar de consultar AMG.

D'altra banda, si bé es cert que IMDB proporciona una quantitat molt més gran d'informació i moltes més varietats d'informació, no és menys cert que AMG la presenta amb tanta eficiència que, quan es disposa d'un temps limitat per obtenir informació, és preferible AMG per raons d'eficiència.

Malgrat tot, la superioritat global d'IMDB és manifesta. D'acord amb la nostra metodologia, d'un total teòric de 24 punts AMG n'obté 12, mentre que IMDB n'obté 20. En percentatge, mentre que AMG està a un 50% de l'ideal dels 24 punts, IMDB està a un 87%, és a dir, 37 punts per sobre del seu rival.

En aquesta puntuació, AMG ha quedat molt perjudicada per les seves constants falles i missatges d'error, un apartat que hem avaluat sota el nom de "rendiment tècnic" i en el qual hem agut d'atorgar un zero a AMG; les seves escasses funcions de cerca, així com la pobresa relativa d'informacions complementàries - sempre en comparació amb IMDB- també han perjudicat molt AMG..

Naturalment, aquests resultats es deriven de la nostra metodologia i de la nostra mètrica. Una altra forma de mesurar i de decidir què es mesura pot donar altres resultats.

5. Conclusions (II)

No ens podem estar d'afegir unes segones conclusions, en aquesta ocasió relatives a les polítiques d'informació. És molt significatiu, per previsible, però ens hauria de ser insuportable, que no hi hagi cap projecte en cap dels nostres àmbits (Catalunya, Espanya, Europa...) que es pugui comparar a AMG o a IMDB.

Caldria iniciar urgentment un projecte solvent que, en poc temps, ens dotés d'un recurs digital semblant als anteriors, però, amb dues diferències: (1) que estigués fet des de la nostra perspectiva i, (2) que defensés els nostres interessos davant els interessos - legítims, però diferents- de Hollywood.

Som conscients que la nostra reclamació s'assembla sospitosament a la situació del ratolí que proposava de posar un cascavell al gat. És per tant ben cert que és més fàcil dir les coses que fer-les, però, com a mínim, diguem-les.

6. Fonts

AllMovie Guide
<http://allmovie.com>

Internet Movie Database
<http://www.imdb.com>

Internet Invisible
<http://www.internetinvisible.com>

Invisible Web
<http://www.invisibleweb.com>

Annex. Anàlisi comparativa AMG/IMDB: figures i taules

Figura 1. Model de registre per a films d'AMG (s'ofereix una traducció aproximada dels noms dels camps)

- 01. Títol
- 02. Any
- 03. Nacionalitat
- 04. Durada
- 05. Color/BN
- 06. Ficció/Documental
- 07. Altres títols (AKA)
- 08. Qualificació AMG (*AMG Rating*)
- 09. Director (*Director*)
- 10. MPAA Rating
- 11. Gènere (*Genre/Type*)
- 12. Estils de producció artístics (*Artistic Production Styles*)
- 13. Avisos (*Flags*) (Classificació moral)
- 14. Paraules clau temàtiques (*Keywords*)
- 15. Autor de la font (*From book by, From story by, From play by, From idea by*)
- 16. Títol de la font (*From book, From play*)

- 17. Paraules clau geogràfiques, cronològiques i sociològiques (*Set In*)
- 18. Paraules clau onomàstiques (*Key name*)
- 19. Data i país d'emissió (*Release*)
- 20. Taquilla (*Box office*)
- 21. Pressupost (*Movie budget*)
- 22. Tipus de color (*Color type*)
- 23. So (*Sound by*)
- 24. Procés cinemàtic (*Cinematic process*)
- 25. Estils de producció (*Production Styles*)
- 26. Producció (*Produced by*)
- 27. País i data d'estrena o d'emissió (*Release*)
- 28. Autor de l'estrena o d'emissió (*Released by*)
- 29. Vegeu també (*See also*)
- 30. Mitjà d'emissió i horari
- 31. Visualització d'un fragment de vídeo (*Preview*)
- 32. Sinopsi (*Plot Synopsis*)
- 33. Crítica (*Review*)
- 34. Actors i personatges (*Cast*)
- 35. Equip de producció (*Production Credits*)
- 36. Premis (*Awards*)
- 37. Films relacionats
- 38. Films amb les mateixes persones (*Movies with same personnel*)
- 39. Films del mateix tema (*Movies on the same theme*)

Figura 2. Model de registre per a cineastes d'AMG (s'ofereix una traducció aproximada del noms dels camps)

- 01. Nom
- 02. Nom de naixement (*Birth name*)
- 03. Altres noms (*AKA*)
- 04. Rol (*Occupation*)
- 05. Data i lloc de naixement (*Birth*)
- 06. Data de defunció (*Death*)
- 07. Ocupacions
- 08. Anys en actiu (*Years active*)
- 09. Països (*Countries*)
- 10. Gèneres (*Genres*)
- 11. Vegeu també (*See Also*)
- 12. Biografia (*Biography*)
- 13. Persones amb qui ha treballat (*Worked with*)
- 14. Filmografia (*Filmography*)
- 15. Emissió per televisió (*TV Schedule*)
- 16. Premis (*Awards*)
- 17. Nom de naixement (*Birth Name*)

Figura 3. Model de registre per a films d'IMDB (s'ofereix una traducció aproximada dels noms dels camps)

- 01. Títol
- 02. Any
- 03. Director
- 04. Guió
- 05. Gènere
- 06. Frases promocionals

- 07. Sinopsi
- 08. Comentaris d'espectadors
- 09. Qualificació atorgada pels espectadors
- 10. Actors
- 11. Títols alternatius
- 12. Durada a diferents països
- 13. Nacionalitat
- 14. Llengua
- 15. Color/BN
- 16. Tipus de banda de so
- 17. Certificacions a diferents països
- 18. Crítiques (*Reviews*)
- 19. Premis i candidatures
- 20. Paraules clau (*Plot keywords*)
- 21. Citacions memorables (*Memorable quotes*)
- 22. Dades diverses (*Trivia*)
- 23. Errades (*Goofs*)
- 24. Llista de temes de la banda sonora (*Soundtrack listing*)
- 25. Versions alternatives
- 26. Films relacionats
- 27. Recaptacions i dades econòmiques
- 28. Localitzacions de la filmació
- 29. Especificacions tècniques
- 30. Bibliografia sobre el film
- 31. Notícies sobre el film
- 32. Trailers
- 33. Pòsters
- 34. Fotografies

Figura 4. Model de registre per a cineastes d'IMDB (s'ofereix traducció aproximada dels noms dels camps)

- 01. Nom
- 02. Data i lloc de naixement
- 03. Data i motiu de defunció
- 04. Biografia
- 05. Filmografia
- 06. Premis i candidatures
- 07. Altres treballs
- 08. Publicitat
- 09. Representat
- 10. Fotografies
- 11. Articles de premsa
- 12. Web oficial

Figura 5. Vista parcial d'un registre d'AMG del film [Force of Evil](#), d'Abraham Polonsky

Figura 6. Vista parcial d'un registre d'AMG del director [Abraham Polonsky](#)

Figura 7. Vista parcial d'un registre d'IMDB del film [Force of Evil](#), d'Abraham Polonsky

Figura 8. Vista parcial d'un registre d'IMDB del director [Abraham Polonsky](#)

Taula 1. Estadístiques bàsiques

Base de dades	Films	Cineastes
AMG	182.441	246.188
IMDB	258.337	864.419

Font: AMG, IMDB, octubre del 2000

Taula 2. Resultats tests quantitius

Base de dades	Test 1	Test 2	Test 3	Test 4
AMG	-	23	55	8
IMDB	52	27	79	10

Notes:

Test 1: films amb la paraula *Barcelona* en el títol

Test 2: films basats en obres de Hemingway

Test 3: films amb guions de Rafael Azcona

Test 4: films de Luis Buñuel amb la col·laboració de Luis Alcoriza

El guionet indica que ha fallat el test

Taula 3. Resultats tests qualitius

Base de dades	Test Polonsky	Test Dracula
AMG	**	**
IMDB	*	*

Notes:

**= Result de forma òptima. *= Result de forma no òptima

Taula 4. Descripció i tractament documental

Base de dades	NC Films	NC Cineastes	NC TD
AMG	39	16	12

IMDB	34	12	3
-------------	----	----	---

Notes:

NC Films: nombre de camps en Films

NC Cineastes: nombre de camps en Cineastes

NC TD: nombre de camps amb tractament documental (entre els dos models de registre)

Taula 5. Valoració global

<i>Base de dades</i>	<i>qLi</i>	<i>qTi</i>	<i>FR</i>	<i>PI</i>	<i>FC</i>	<i>RT</i>	<i>Valoració global</i>
AMG	4	2	1	4	1	0	12 (50%)
IMDB	2	4	4	2	4	4	20 (87%)

Notes:

i) Cada apartat es pot valorar entre 0 i 4

ii) Significat de les columnes:

qLt: qualitat de la informació

qTi: quantitat de la informació

FR: funcions de recerca

PI: presentació de la informació

FC: funcions complementàries

RT: rendiment tècnic

VG: valoració global, sobre 24, absoluta i en tant per cent

