

QUI TÉ POR DEL CONEIXEMENT? APLICACIÓ DE LA MINERIA DE DADES A UN FONS FOTOGRÀFIC DE LA FILMOTECA DE CATALUNYA

*Miquel Àngel Pintanel Bassets
Filmoteca de Catalunya*

1. Introducció

En aquest projecte hem aplicat el Descobriment de Coneixement en Bases de dades, que Fayyad *et alii* (1996b) defineixen com «(...) el procés no trivial d'identificar patrons en les dades que siguin vàlids, nous, potencialment útils i, en definitiva, comprensibles», per a millorar la gestió del patrimoni cultural moble. Per a aquesta aplicació en la gestió del patrimoni cultural, sense oblidar que el procés ha de complir totes les condicions, ens hem de centrar en com extreure patrons vàlids i potencialment útils de les dades dels objectes per aplicar al dia a dia de la millora d'aquesta gestió. Per a millorar la difusió del seu coneixement i aportar-lo a la societat potenciarem que aquests patrons siguin nous i comprensibles.

Aquest procés té varies capes i múltiples possibilitats, però se n'han volgut acotar els límits en aquest cas. La primera capa, imprescindible, és el tractament de les dades, la seva transformació i la seva tria, el que alguns autors han considerat la part més gran de la feina (Sangüesa *et alii*, 2010). La segona capa aplica algorismes de mineria de dades i a partir dels patrons creats per aquesta poder generar una tercera capa de visualització interactiva, que és la que permet mostrar conclusions útils de tot el procés. Totes aquestes capes depenen de dos factors inicials: el coneixement expert de les dades i, a partir d'aquest, saber realitzar les preguntes indicades i veure si les hipòtesis que es plantegen són correctes.

L'aplicació de la mineria de dades s'ha fet amb les dades d'una catalogació existent dels fons fotogràfics de la Filmoteca de Catalunya. Des de l'any 1999 fins al 2011 a la Filmoteca de Catalunya es va dur a terme un projecte de col·laboració amb la Universitat de Barcelona en el marc del qual van participar alumnes d'Història de l'Art en la catalogació de fotografies fixes del seu arxiu. Com a resultat es van catalogar prop de 20.000 fotografies, que corresponien a 472 pel·lícules produïdes a Espanya entre 1932 i 1984, i hi van intervenir prop de 400 alumnes. El procés tenia dues fases: descripció prèvia en una fitxa en paper pautada i introducció de les dades de forma tradicional en la base de dades relacional utilitzada per la Filmoteca que s'anomena Mobydoc.

2. Anàlisi de les propietats de les dades i la seva qualitat

Per a poder fer el procés de mineria de dades s'ha de comprovar abans que les dades de la catalogació són suficients i que tenen la qualitat necessària per a fer el procés. Com no és viable, per la complexitat que suposa, comprovar tots els camps de la base de dades, és necessari triar-ne una part que compleixi aquestes condicions: (a) que la informació entrada sigui molt completa, preferentment que fos obligatori introduir-la; (b) que tinguem prou coneixement previ com per treure conclusions de la seva anàlisi; (c) que hi hagi una relació directa entre els camps i no depenguin d'un camp que no s'analitzi ja que podria distorsionar-ne les relacions.

Per a comprovar la qualitat de les dades una de les millors formes és agrupar-les d'alguna forma i cercar els punts atípics (outliers), que normalment són els elements erronis o excepcionals. També és molt útil aquest procés per a realitzar la correcció de les dades. En la **figura 1** visualitzem el resultat de comparar els valors dels camps Mides i Suport del total de 20.000 registres de la base de dades amb el programa de mineria de dades. La forma de representació és un diagrama de dispersió de les comparacions dels valors per parells de camps. Aquesta forma de veure els valors comparats ens permet comprovar com aquestes s'agrupen i quines són les dades atípiques, que corresponen a les que no s'agrupen amb la resta. També ens permet veure si les agrupacions corresponen a les dades esperades pel

coneixement previ que tenim. La gràfica que crea el programa utilitzat és interactiva i això ens permet saber amb un clic a quines dades es refereix cada punt i fer-ne la comprovació.

Els valors de configuració que hem triat per a representar les dades en les següents figures fa que destaquin més les dades atípiques, però es pot comprovar que la major part de dades estan agrupades si es canvien aquests ajustaments. S'ha deixat com a exemple l'opció que fa ressaltar més les dades atípiques ja que és la que es fa servir per a distingir entre els punts erronis i excepcionals en la pràctica de control de la informació. Després de les comprovacions es va arribar a la conclusió que el fons escollit era vàlid per a poder fer un procés de Descobrimament del Coneixement en Bases de dades.

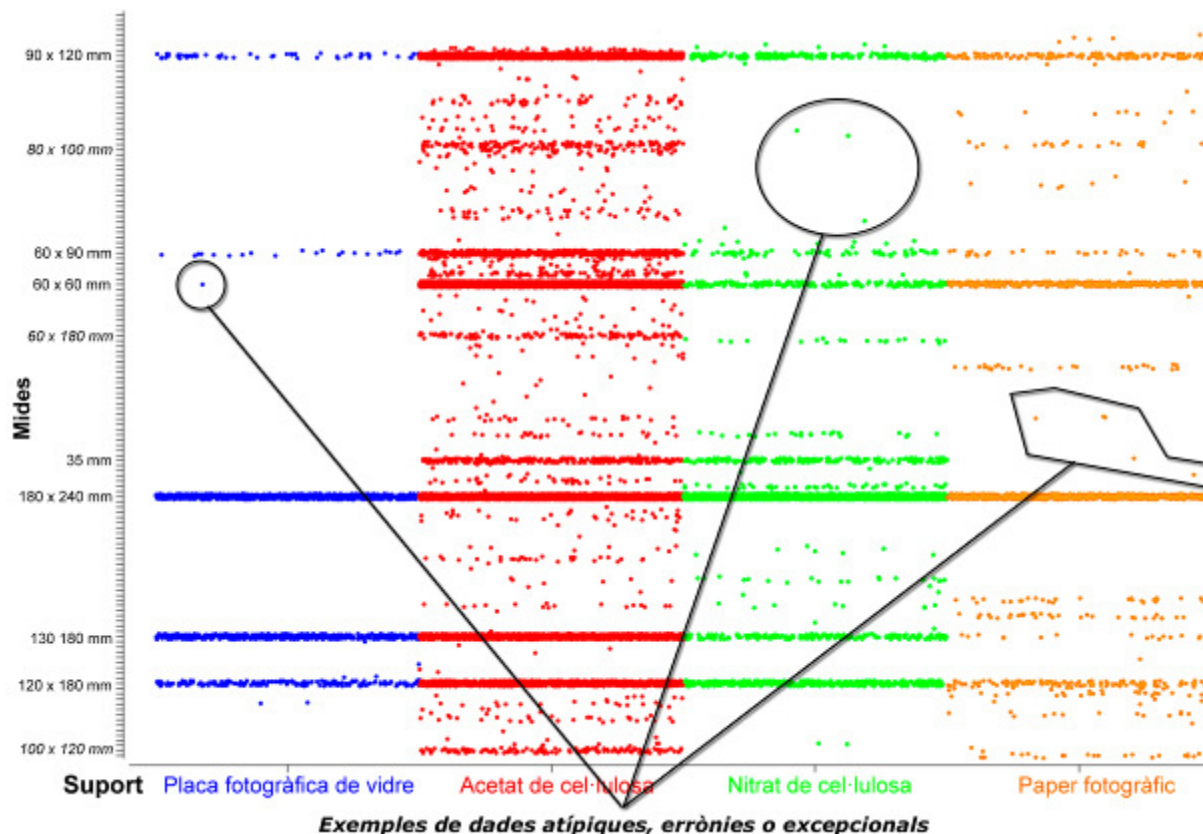


Figura 1. Diagrama de dispersió amb la indicació de les dades atípiques.

3. Coneixement adquirit

3.1. Amenaces, febleses i forteses del fons fotogràfic

a. Tipus de materials conservats

Destaca la quantitat de materials diferents que formen el fons, com es pot veure a la **figura 2**, tant per la diversitat de mides estàndard com per la presència de tots els suports possibles per l'època. Això suposa un esforç extra en les condicions d'emmagatzematge, tant a nivell de condicions climàtiques com per les necessitats de conservació o la diversitat de contenidors necessaris.

També aquesta diversitat fa destacar el valor de la col·lecció, tant des del punt de vista d'aportar coneixement gràcies a la seva riquesa tipològica, com des del punt de vista de la seva composició. És un fons majoritàriament format per negatius i per tant serveix molt bé per explicar la forma de treballar dels fotògrafs i els laboratoris de fotografia industrial de 1932 a 1984.

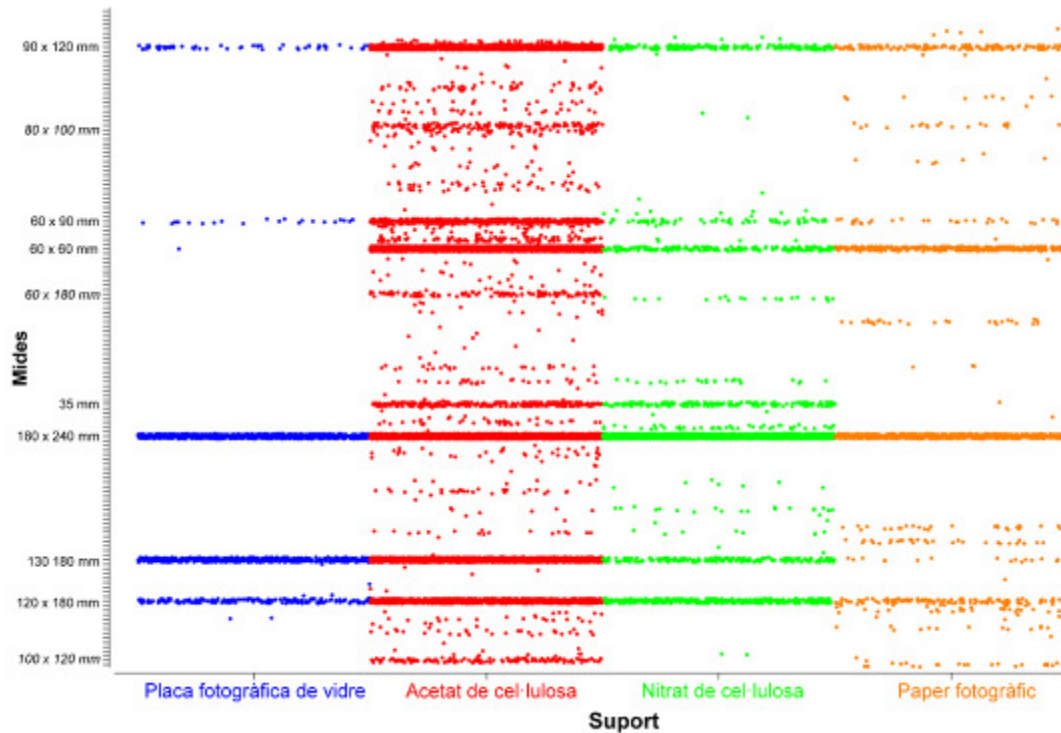


Figura 2. Diagrama de dispersió entre la mida i el suport del material fotogràfic.

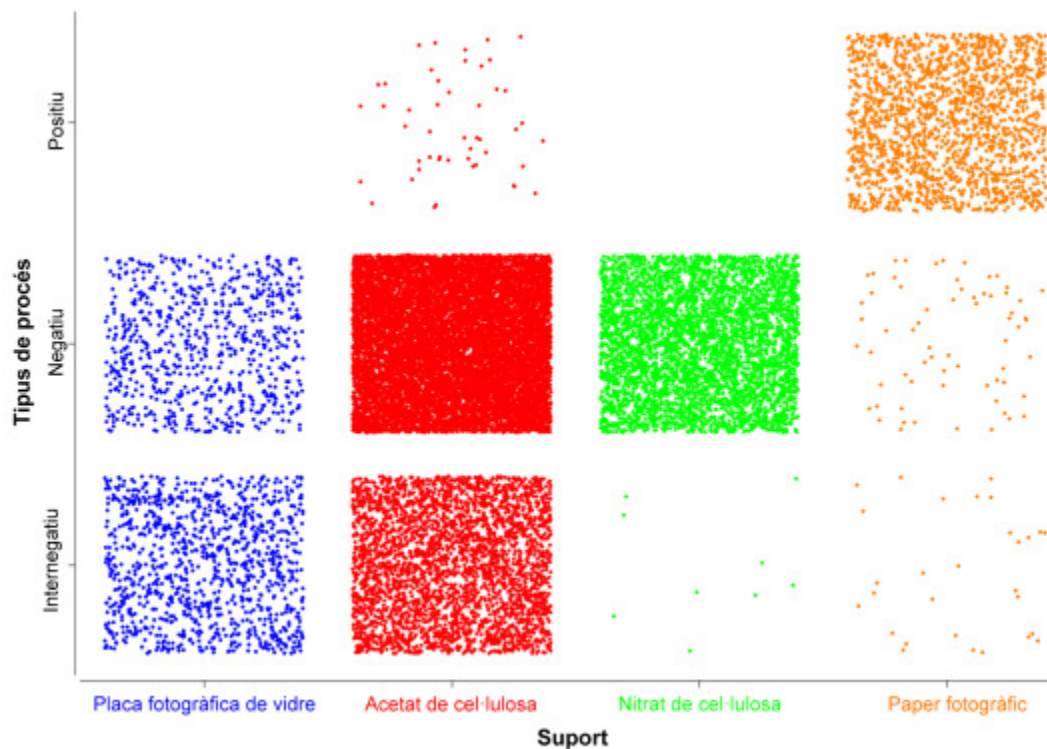


Figura 3. Diagrama de dispersió entre el tipus de procés i el suport del material fotogràfic.

La gràfica de la **figura 3**, que compara el tipus de procés amb el suport, és una de les que dóna un coneixement més evident de la forma de treballar dels fotògrafs i els laboratoris. Com ja hem dit anteriorment, les dades tendeixen a ser uniformes i aquesta uniformització ens aporta coneixement sobre la realitat que reflecteixen. El primer que s’observa en aquest gràfic és el gran valor que tenia el nitrat per als fotògrafs, ja que dóna una qualitat molt superior d’imatge que els altres suports. Aquest valor es veu en el fet que és utilitzat gairebé exclusivament per a

realitzar negatius originals i no per a tasques de reproducció com internegatius¹. Aquesta és una de les informacions que es fa molt evident en una gràfica i que no és fàcil de veure en una llista de dades. També es pot observar a la gràfica els diferents usos que pot tenir l'acetat, ja que és present a tots els tipus de procés, i això correspon tant a les tasques del laboratori purament fotogràfic com a les que fan referència a procediments fotomecànics, que correspondrien a feines d'impremta.

Si aquesta anàlisi s'hagués d'aplicar a la gestió de l'equipament cultural s'hauria d'ampliar amb eines d'aprenentatge automatitzat. Això serviria per a poder controlar de manera informada la compra de contenidors, ja que el sistema podria estimar quins serien més necessaris en un futur.

b. Estat de conservació de la col·lecció

Un punt molt important per a la gestió d'un fons patrimonial és conèixer l'estat de conservació dels objectes custodiats i en quines parts del fons es necessita actuar amb més urgència. Per poder fer això és molt útil tenir una estimació de quines parts tenen més possibilitats de necessitar una actuació en forma de preservació o restauració.

En la introducció de dades a Mobydoc es podien triar d'una llista 6 nivells d'estat de conservació: **Bo**, **Primer grau**, **Segon grau**, **Tercer grau**, **Inservible** i **Baixa**. Aquests nivells estan ordenats de millor a pitjor estat de conservació i tenien algunes particularitats: (a) no es recomanava l'ús del terme **Bo** ja que està reservat a reproduccions noves (b) el terme **Baixa** és un terme «històric», indica que el material ha estat destruït o enviat a un gestor de residus, i no afecta a la gestió de l'equipament. Així, de forma efectiva, la tria es feia en 3 nivells de conservació, ja que **Inservible** era un nivell d'últim recurs.

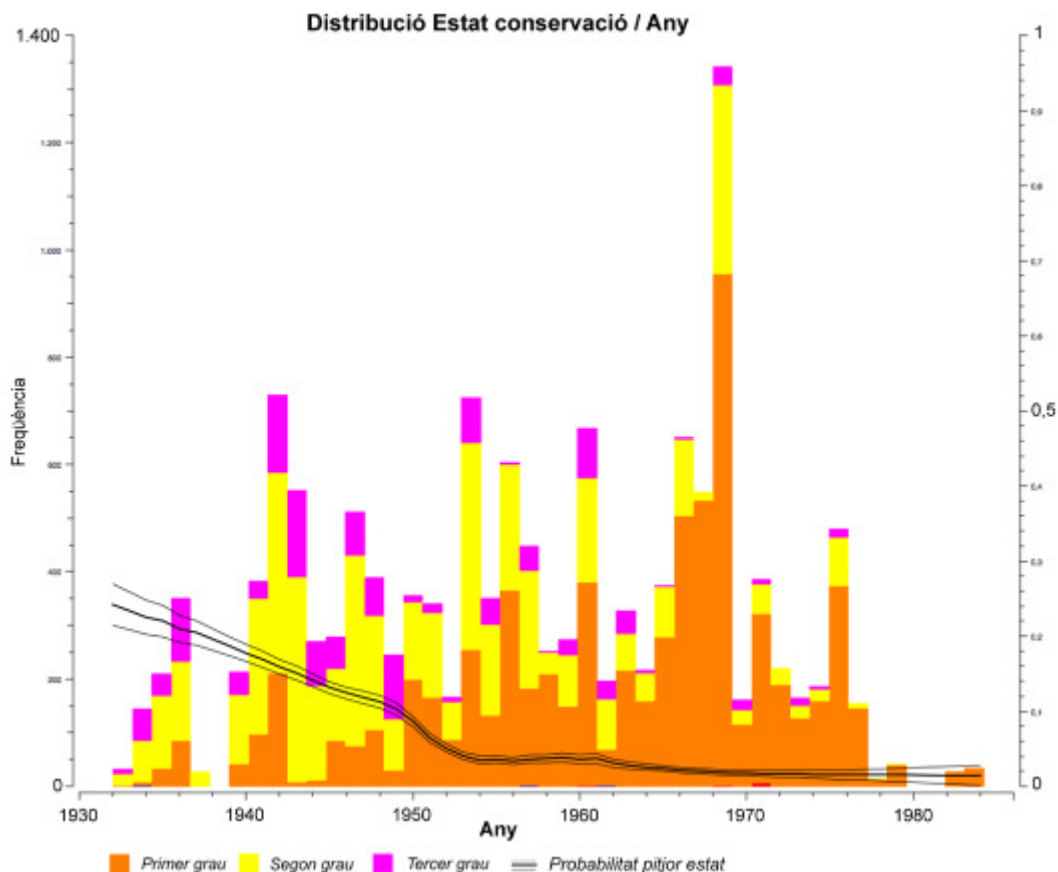


Figura 4. Distribució dels estats de conservació amb la probabilitat que estiguin en el pitjor estat possible / any.

Aquesta reducció a 3 nivells efectius va fer que aquests quedessin molt justos per a descriure l'estat de conservació i va fer que no quedés clara la frontera entre els nivells. Apart d'això, en fer la mineria de dades, ha donat problemes una decisió de catalogació: es va decidir canviar la

forma de realitzar la valoració de l'estat de conservació uns anys després de començada la catalogació dividint-la en **Estat de conservació químic** i **Estat de conservació físic**. Quan es va fer aquest canvi no es va tornar a revisar el material per introduir aquest valor, ni es va posar un valor per defecte al valor que quedava buit, que va ser **Estat químic**. El fet que quedin valors buits distorsiona les gràfiques de distribució i per això en aquest estudi hem decidit descartar els valors de l'Estat químic, que són la minoria. El motiu de dividir els dos tipus d'estats va ser afavorir la tasca dels catalogadors i donar dades més precises sobre l'estat de conservació. Normalment s'ha de tenir més cura amb l'estat químic ja que aquest és reflex que el material té algun tipus de degradació activa i pot estar amenaçat de destrucció. Dels problemes físics normalment no es dedueix que un material pugui estar en perill de canviar de grau a curt o mitjà termini, ja que la majoria dels defectes físics són ratllades, parts trencades i parts absents. Quan els defectes físics són reflex d'una degradació activa, suport balcat o arrugat, van acompanyats de defectes químics i normalment puja de la mateixa manera el nivell de degradació química.

A la **figura 4** es pot veure la distribució d'estats de conservació, limitats en realitat a l'estat de conservació físic pels problemes esmentats anteriorment, ja que és el que té més dades representades. Es pot veure com a mida que el material és més modern es fa més preeminent la presència d'objectes en **Primer grau** de conservació.

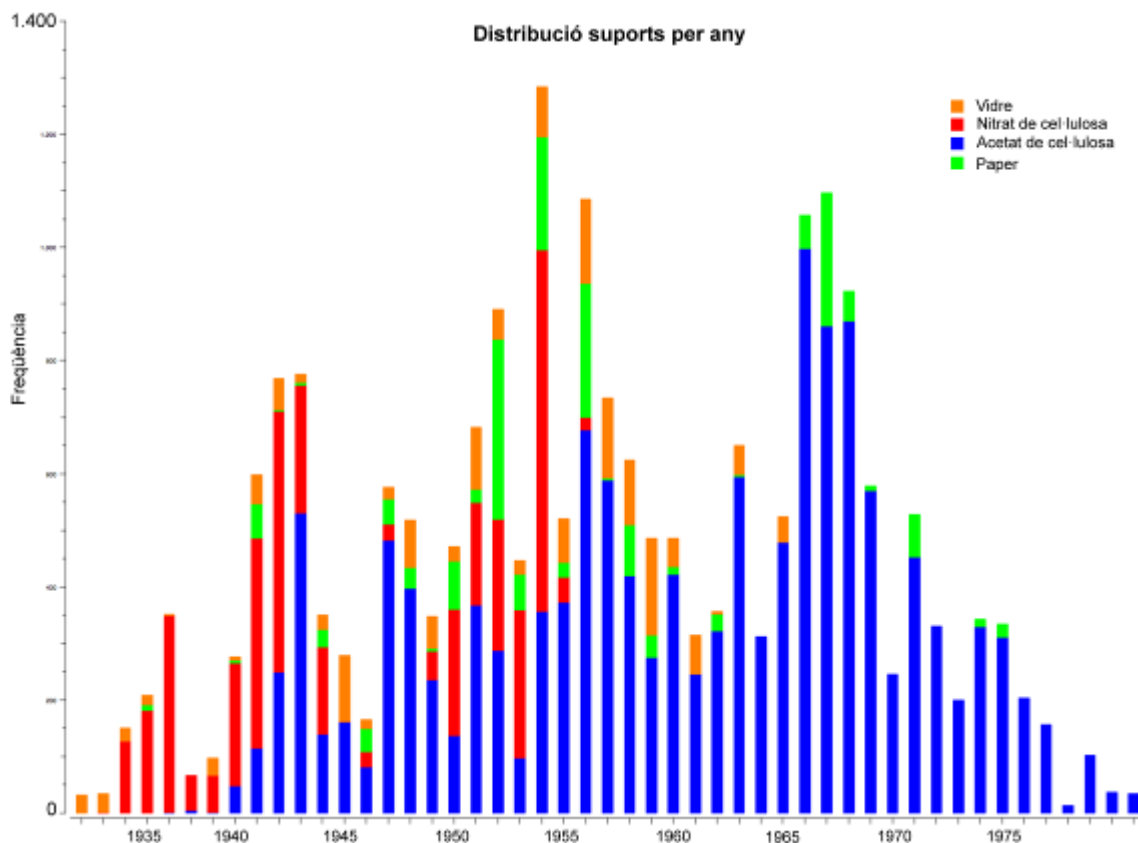
La línia més gruixuda indica una estimació de la probabilitat que un material estigui en el pitjor estat possible; en aquest cas s'ha triat el nivell **Tercer grau**. Les línies fines corresponen a la confidencialitat d'aquesta estimació, això és, el marge d'error que té l'estimació. Es comprova el que la lògica indica, això és, que com més antic és un material més possibilitats té d'estar malmès. Aquesta gràfica també reflecteix un problema que tenen una part dels materials de l'any 1958 a 1962 que s'han degradat més que la resta. Aquest objectes tenen les mateixes característiques físiques (osques, mides, color i la presència del número de clixé fotografiat), la qual cosa fa pensar que són una remesa, o un model concret, d'un fabricant de suport fotogràfic que tenia un defecte, o que va ser mal processat en el revelat.

b. Anàlisi cronològica del fons

Per conèixer el valor dels fons patrimonials un dels punts importants és saber si ocupa tota la franja cronològica de la qual s'ocupa l'equipament i en quin volum. Saber quin és l'estat de la qüestió cronològica en els fons ens servirà per a portar una política d'adquisicions que pal·liï els punts febles de la nostra col·lecció o en reforci els punts forts. Reforçar punts febles o forts participen de polítiques diferents i normalment s'ha de triar per un dels dos quan les possibilitats pressupostàries no són il·limitades. Si el que es vol és que l'equipament sigui especialitzat en un període mot concret reforçarem aquest període i si volem que en el nostre equipament es trobi un ampli espectre de materials reforçarem els punts febles.

A la gràfica de la **figura 5** veiem la distribució de materials per anys del fons estudiat, en ella podem veure l'evolució tant dels suports utilitzats com de la seva quantitat. És útil per veure com evoluciona l'ús dels materials durant els anys i per valorar l'activitat del laboratori. El laboratori Reproducciones Sabaté va funcionar després de l'any 1984, de fet encara és actiu. Que l'últim material sigui de l'any 1984 semblaria que obeeix a un canvi en la forma de tractar el material, que ja no devia ser dipositat pels clients a les instal·lacions del laboratori i no tant a una baixada tant gran de la producció.

És un fons que està format majoritàriament de material que correspon a negatius i internegatius en tota la seva cronologia. Com és un fons tancat, ja que es va comprar tot l'inventari del laboratori i té unes característiques molt especials no es poden prendre decisions per ampliar-lo. El que sí que es pot dir és que donada la magnitud del fons i la presència majoritària de material original, molt difícil d'aconseguir en quantitat, és probable que no es pogués augmentar la fortalesa en aquest aspecte de l'equipament.



5. Distribució de nombre de material per anys del fons.

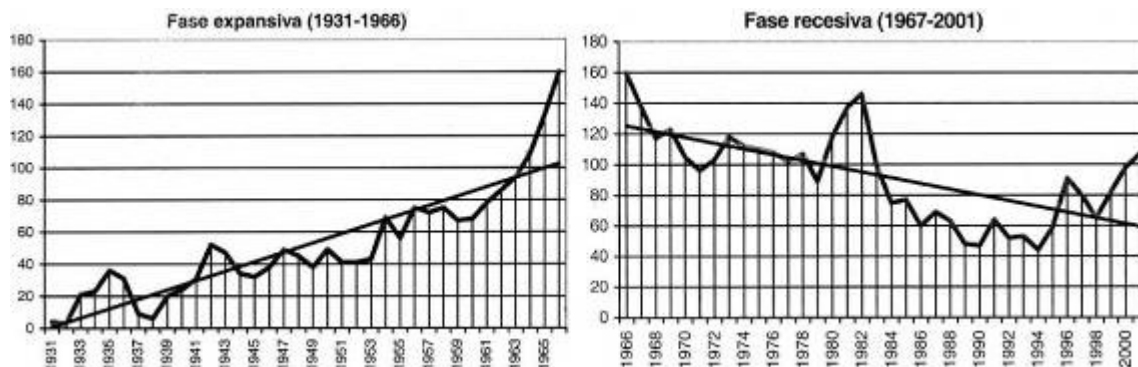


Figura 6. Gràfiques amb la producció cinematogràfica a Espanya de 1932 a 2000. Font: Diez Puertas (2003).

La gràfica de la **figura 5**, però, tal com està queda òrfena per a ajudar a prendre decisions o consideracions sobre el fons, sobretot per conèixer-ne les febleses o forteses, ja que no té relació amb la producció de films que corresponen a cada any. A la **figura 6** podem veure dues gràfiques amb els títols produïts per any a Espanya. Per exemple, la dada que pot ser més evident de comparar amb la **figura 5** és la de l'any 1932, ja que en realitat els films que conserva el fons corresponen a un tant per cent més elevat del que podria semblar. La transició del mut al sonor va ser molt traumàtica a Espanya i gairebé no es van produir pel·lícules, amb la qual cosa seria molt difícil incrementar molt més la quantitat de materials d'aquests anys. També hi ha un efecte distorsionador a la gran pujada del anys 60 provocada per l'impacte de les coproduccions, moltes d'elles amb una part de producció espanyola molt petita o inexistent.

Una altra de les dades que es pot veure a la gràfica de la **figura 5** és la pervivència de material amb suport en nitrat de cel·lulosa fins a l'any 1956, quan aquest material teòricament deixa de fabricar-se i distribuir-se a 1952. Aquest és un dels casos en què la informació adquirida amb el descobriment de coneixement pot aportar un avís als historiadors, ja que s'hauria de buscar la

raó d'aquesta presència, que a més, coincideix amb que durant l'any 1948 deixa d'utilitzar-se i el seu ús es reprèn amb força l'any 1950.

3.2. Connexions entre les dades

Per a poder posar un exemple de connexió entre les dades hem utilitzat els camps **Autor**, **Lloc de creació** i **Data de producció**. Hem triat aquest grup de camps perquè podien portar a conclusions sobre la forma de treballar dels fotògrafs i perquè hi ha relacions clares entre els valors dels camps. Amb això poden ser útils per a treure conclusions sobre el coneixement de la relació entre les dades, però també poden aportar exemples de com poden ajudar en l'entrada de dades.

Mostrarem diferents formes de visualitzar aquestes connexions, una d'elles interactiva, i que corresponen a diferents formes de mineria de dades. Les dues primeres parteixen del concepte de la classificació, així que ens mostren els resultats de les connexions, i la tercera aplica l'anàlisi de xarxes, així que és la que ens mostrarà més clarament les connexions entre les dades.

a. Nomograma

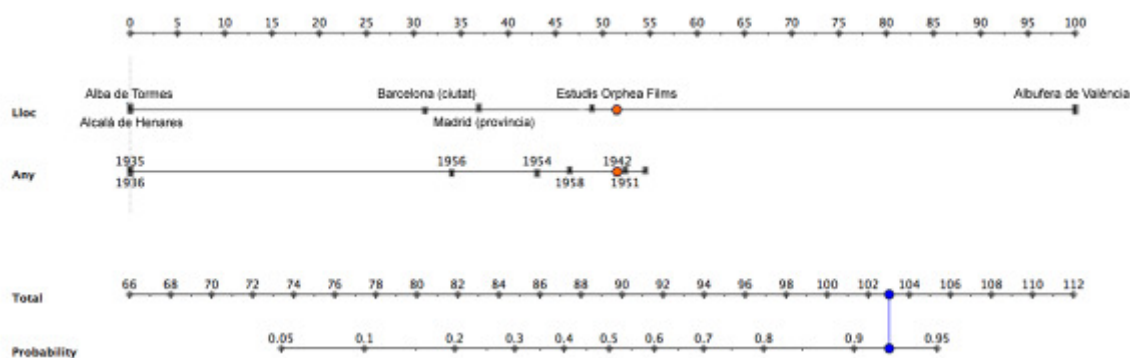


Figura 7. Nomograma amb la millor de les possibilitats que un film sigui d'Emili Godes segons l'any i el lloc on està rodat. En color taronja els punts que es poden moure.

La primera forma utilitzada per demostrar de forma visual la força de les relacions entre el lloc, l'any i l'autor és el nomograma. El nomograma és una eina interactiva que ens permet moure uns punts de forma semblant al funcionament d'un àbac. En aquest cas els punts mòbils corresponen al lloc i l'any, i com a resultat del canvi de posició mouen un altre parell de punts que ens donen la probabilitat que si es donen les dues condicions la combinació de partida sigui certa a partir de les connexions conegudes. En aquest cas que la fotografia sigui de Godes és la hipòtesi proposada.

Abans de poder utilitzar el nomograma s'ha hagut d'aplicar a les dades una classificació bayesiana ingènua (*Naive Bayes Classifier*) per a poder agrupar les dades i per a poder saber de forma numèrica quina influència té cada valor en el total en els exemples coneguts. A la representació visual la predicció de la probabilitat està entre els valors 0.05, improbable, i 0.95, molt probable.

Ja que el programa és interactiu presentem dues figures amb diferents estats per a poder copsar-ne el funcionament de forma més clara. La **figura 7** mostra la posició més favorable a que un film sigui d'Emili Godes que permet el nomograma generat pel programa².

De fet, la posició en la línia superior de percentatge del lloc dels estudis Orphea Films ja ens deixa veure la influència en l'autor, doncs correspon a més d'un 50% de coincidències dels autors que hi treballen. Si es limita a períodes concrets com a la **figura 7** dona una probabilitat gairebé segura. A la **figura 8**, però, es pot veure la importància de combinar les dades del lloc amb les de l'any per a poder donar dades més fiables.

Aquesta eina constata que es podria fer servir els camps **lloc**, **autor** i **any** per a ajudar a la introducció de noves dades: (a) per revisar si les fitxes amb autor desconegut realitzades a un

lloc i període concrets són de l'autor més probable i (b) per afavorir l'entrada de noves fitxes si el gestor de bases de dades integrés un sistema d'aprenentatge automatitzat que presentés els autors més probables en primer lloc a l'introduïdor de les dades noves o avisés d'alguna forma

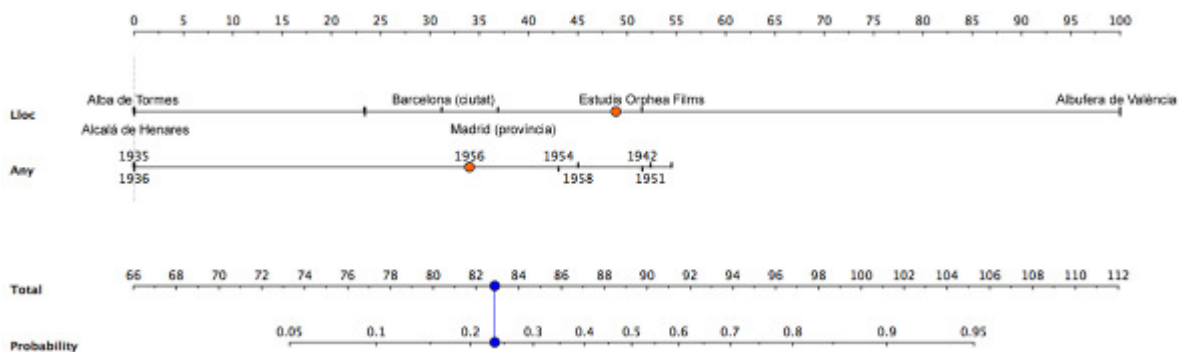


Figura 8. Nomograma amb menors possibilitats que un film sigui d'Emili Godes segons l'any i el lloc on està rodat.

si la dada que ha introduït és altament improbable.

b. Arbre de classificació

A la **figura 9** podem veure una part d'un arbre de classificació amb la distribució dels autors que treballen en un estudi cinematogràfic. S'ha limitat als estudis i no a totes les localitzacions per a fer més llegible el resultat i perquè les conclusions podien ser útils de cara a conèixer el funcionament d'aquests equipaments.

Apart de les relacions en aquesta visualització podem veure quin és l'autor més probable per un estudi i inclou un gràfic amb els percentatges d'altres autors. Si es col·loca el punter a sobre d'un dels gràfics de percentatges apareix un quadre amb el nom de la resta de fotògrafs i les probabilitats respectives. Amb aquesta eina veiem molt gràficament la distribució dels autors en la seva dependència amb l'estudi cinematogràfic. És una forma de tenir al dia la classificació del nostre fons en cadascuna de les seves àrees de coneixement amb l'afegit que ens pot ajudar a predir noves dades en mostrar la més probable. Un dels problemes d'aquesta visualització és que s'ha de limitar a alguna part del coneixement, ja que si no les dimensions de l'arbre fan molt difícil la seva lectura.

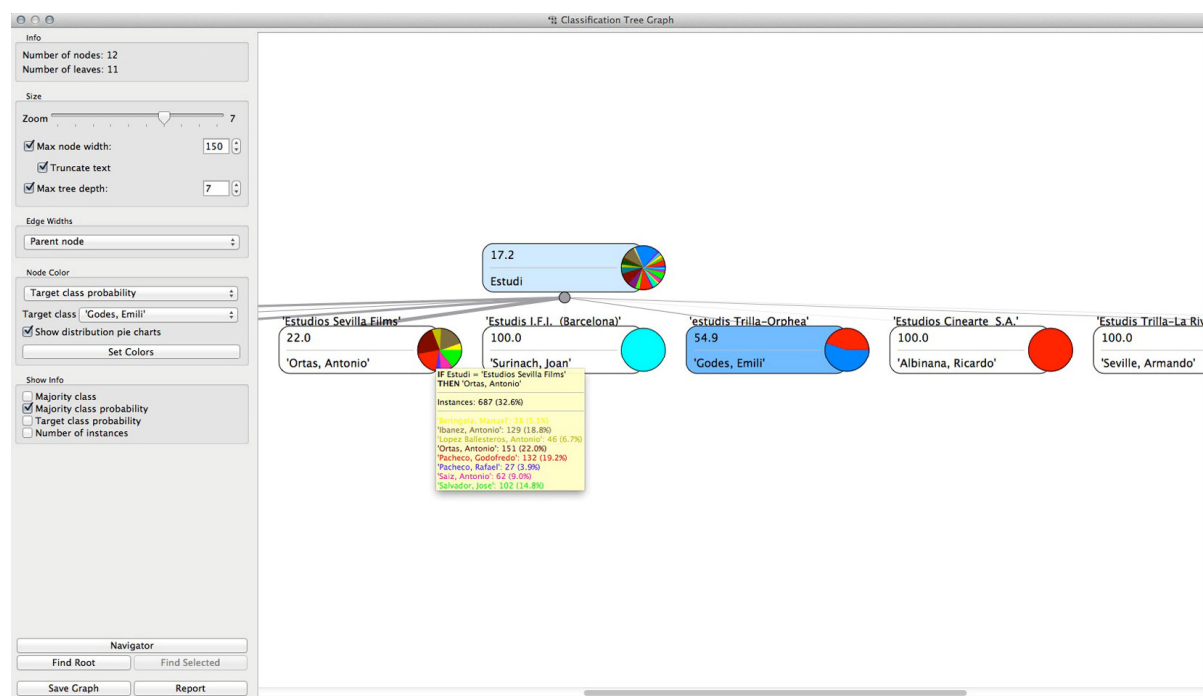


Figura 9. Captura de pantalla on es pot veure l'arbre de classificació amb els autors i els estudis.

En aquesta visualització hem utilitzat una opció que serveix per a estudiar un cas concret com és que marqui la possibilitat que una part de l'arbre sigui d'un autor. Hem triat Emili Godes com a la classe que s'ha de demostrar; així, a la part superior ens marca amb blau, color assignat a Godes, que si una fotografia és d'estudi hi ha el 17,2 de possibilitats que sigui d'ell. En el nivell inferior de l'arbre podem veure en blanc els casos que no corresponen a aquest autor i marcats en blau els que sí ho fan.

L'anàlisi d'aquestes dades no permet concloure que en tots els casos hi hagi una relació tan forta com amb Godes i l'estudi cinematogràfic en el qual treballava més sovint. La conclusió és que és molt probable que en algun cas sigui així, però que no és una norma que es pugui aplicar a tots els casos. El cas més apartat d'aquesta possible norma serien els estudis Sevilla Films, que tenen 8 foto fixes diferents, com mostra el quadre d'informació ressaltat en groc a la **figura 9**.

c. Anàlisi de xarxes

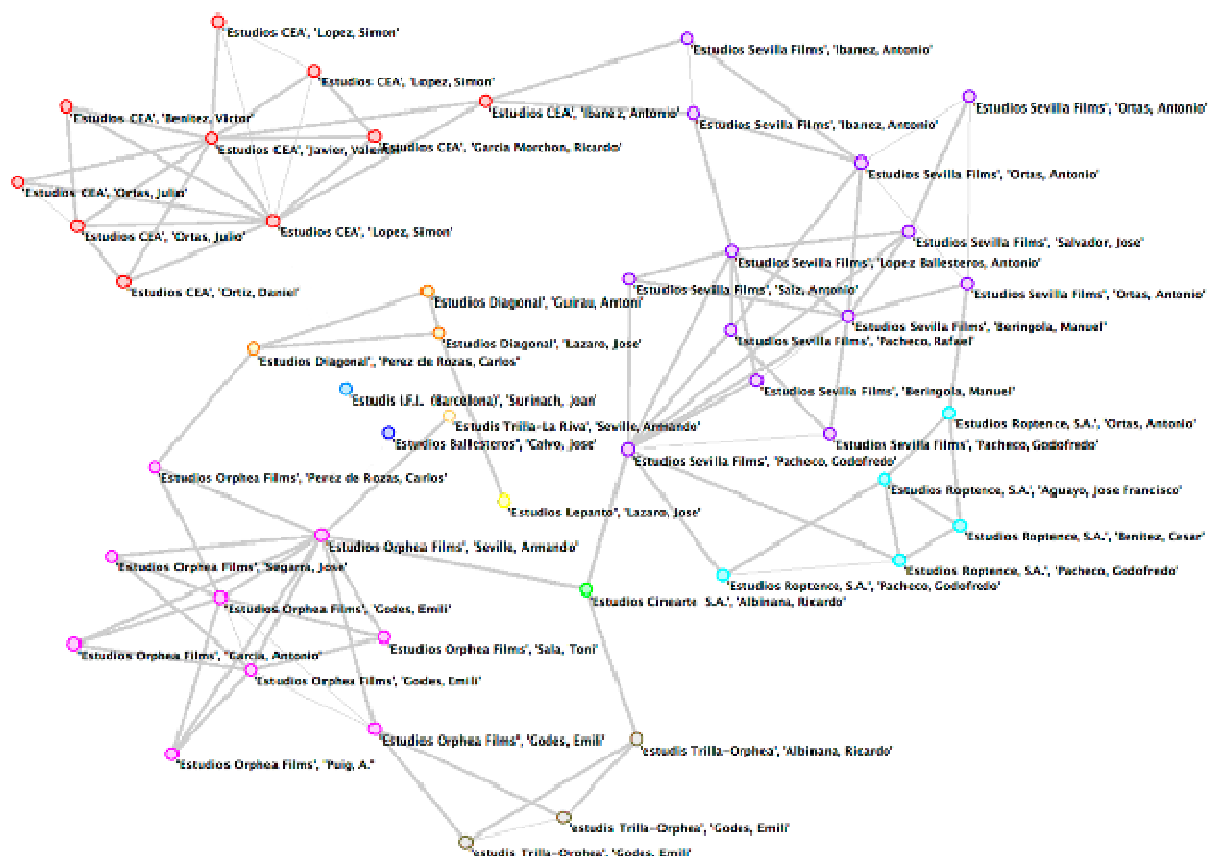


Figura 10. Xarxa amb les connexions entre els estudis cinematogràfics i els foto fixa.

A la xarxa de la **figura 10** podem veure una sèrie de punts que corresponen cadascun a un film del qual es coneix l'estudi cinematogràfic on va ser rodat i el foto fixa. S'ha marcat amb el mateix color cada estudi per a veure'n millor les agrupacions. Les línies mostren les connexions entre aquests punts i conformen la xarxa de relacions entre els fotògrafs i els estudis.

El primer que sorprèn en veure la xarxa és que només hi ha dos punts aïllats, els estudis I.F.I. i els Ballesteros, això vol dir que els foto fixes van passant d'un estudi a un altre i que hi ha una relació entre ells. Fins i tot alguns estudis que tenen un sol film utilitzen fotògrafs que han treballat a d'altres estudis. Això pot reflectir que la feina de foto fixa era molt especialitzada i controlada. De fet durant la dictadura existia un carnet amb la classificació de foto fixa del *Sindicato Nacional del Espectáculo*, sense el qual no es podia exercir aquesta professió.

Sempre hem de comptar, però, amb que les dades reflectides en aquestes gràfiques són una mostra del cinema espanyol de l'època reduïda al material conservat a un fons, Reproducciones Sabaté, i del qual s'ha pogut descobrir el lloc de rodatge. Això vol dir que la xarxa pot estar mostrant en realitat una relació entre els fotògrafs i el laboratori. Seria interessant poder veure com és la relació en el total de films produïts, i no només els conservats a la Filmoteca de Catalunya, i comparar-ho amb aquesta gràfica.

4. Problemàtica del Descobriment de Coneixement en Bases de dades

Un dels perills més evidents que poden aturar el descobriment de coneixement és el procés de preparació, apart de ser la part més gran de feina com hem esmentat anteriorment. En aquest moment és quan surten tots els errors i les petites trampes de la catalogació, ja que aquests es fan evidents per varis factors: (a) les dades es veuen connectades entre elles i en participar d'un mateix coneixement es fan més evidents els errors i (b) és necessari fer processos

automatitzats que no funcionen si tots els valors no estan dins d'uns paràmetres. És en fer-se tan evidents els errors que poden desmoralitzar o evitar que es realitzi el procés, ja que suposarà aturar la catalogació per dedicar-se a la seva correcció.

Un altre dels problemes és que per a que sigui d'utilitat és necessari aturar la catalogació mentre és realitza la mineria de dades, a no ser que sigui un fons tancat com el que hem utilitzat en aquest cas, ja que el procés modifica les dades. Aquest fet fa que per aplicar-lo en un equipament patrimonial de forma continuada hi hauria d'haver una connexió entre la base de dades i el programa de mineria de dades per a que aquesta modificació quedés reflectida en les dades.

5. Aportacions del procés de Descobriment de Coneixement en Bases de dades

L'aplicació del Descobriment de Coneixement en Bases de dades aporta les següents millores en la gestió d'un fons fotogràfic:

- **Qualitat.** Aporta una auditoria objectiva de les condicions i característiques dels objectes de la col·lecció i la seva catalogació que té diversos efectes:
 - La millora de les dades que aporta el procediment en ell mateix
 - Reflexió sobre la qualitat de les dades
 - Repensar els protocols de catalogació
 - Aprendre dels errors de planificació de la catalogació o de les decisions conceptuals i d'estructura de les dades
- **Presa de decisions.** Ajuda a prendre decisions informades, ja que aporta:
 - Dades objectives i comprensibles sobre les necessitats de conservació i actuacions sobre la col·lecció
- **Valor.** Augmentar el valor de la col·lecció a partir del seu coneixement:
 - Veure quin coneixement sobre el cinema i la gènesi de les fotografies ens proporciona la catalogació i aportar una eina de coneixement per a historiadors del cinema a partir de la relació entre els objectes i les dades de la seva creació.
 - Veure de forma gràfica quines són les fortaleses de la col·lecció
- **Accés.** Crear noves formes d'accés a la col·lecció, més enllà de la cerca tradicional:
 - Donar ajudes a la cerca a partir de la mineria de dades sobre els textos i paraules claus que descriuen la col·lecció.
 - Donar accés al coneixement aportat per les visualitzacions generades amb la mineria de dades, anant un pas més enllà en la simple mostra dels fons de la col·lecció.

Bibliografia

- Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *Questiò (Quaderns d'Estadística i Investigació Operativa)*, 25(3), 479–498.
- Baeza-Yates, R., & Ribeiro, B. de A. N. (1999). *Modern information retrieval* (p. 513). Reading, Mass.: Addison-Wesley Longman. Retrieved from <http://biblioteca.uoc.edu/lilibres/16423.htm>
- Bates, M. J. (2006). Fundamental forms of information. *Journal of the American Society for Information Science and Technology*, 57(8), 1033–1045.
- Belkin, N. J., & Robertson, S. E. (1976). Information science and the phenomenon of information. *Journal of the American Society for Information Science*, 27(4), 197–204.
- Bessière, I., & Gili, J. A. (Eds.). (2004). *Histoire du cinéma: problématique des sources*. Paris: Institut National d'Histoire de l'Art.

- Boadas i Raset, J., Casellas, L.-E., & Suguet, M. À. (2001). *Manual para la gestión de fondos y colecciones fotográficas. Biblioteca de la imagen* (Vol. 3, p. 425). Girona: CCG: Centre de Recerca i Difusió de la Imatge CRDI.
- Brecht, B. (1963). Las cinco dificultades para decir la verdad (1934). *Boletín Del Seminario de Derecho Político*, (29-30). Retrieved from <http://labyrintho.uma.es>
- Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351–360.
- Calle Buendía, J. (2005). La Catalogación en una Filmoteca: Arxiu de la Filmoteca de Catalunya. *Cuadernos de Documentación Multimedia*, 16, 64–76.
- Campagnoni, D. P. (2006). The Preservation, Care and Exploitation of Documentation Related to the Cinema: An Unresolved Issue. *Film History*, 18(3), 306–318.
- Capurro, R., & Hj rland, B. (2003). The concept of information. *Annual Review of Information Science and Technology*, 37(1), 343–411.
- Carreras Monfort, C. (2009). *Evaluación TIC en el patrimonio cultural metodologías y estudio de casos*. Barcelona: Editorial UOC.
- Cordón García, J. A., López Lucas, J., & Vaquero Pulido, J. R. (2001). *Manual de investigación bibliográfica y documental: teoría y práctica*. Madrid: Pirámide
- Demšar, J. (2010). Material for Data Mining Lectures, Kyoto University, July 2010. Retrieved June 10, 2014, from <http://www.birolab.si/janez/kyoto/>
- Diez Puertas, E. (2003). *Historia social del cine en España*. Madrid: Editorial Fundamentos.
- DLM-Forum. (1997). *Guidelines on best practices for using electronic information*. Luxemburg: Office for Official Publications of the European Communities.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI Magazine*, 37–54.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). Knowledge Discovery and Data Mining: Towards a Unifying Framework. In *Second International Conference on Knowledge Discovery and Data Mining*.
- Gallardo Arancibia, J. A. (2009). *Metodología para la definición de requisitos en proyectos de data mining (ER-DM)*. (Ó. Marbán Gallego, Ed.). Universidad Politécnica de Madrid Departamento de Lenguajes y Sistemas Informáticos e Ingeniería de Software, Madrid. Retrieved from http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf
- Iglésias Franch, D. (2008). *La fotografía digital en los archivos. Qué es y cómo se trata*. Gijón: Ediciones Trea.
- Iglésias Franch, D. (2012). La definició d'un mapa d'informació conceptual per a la gestió de l'arxiu fotogràfic digital. In *Imatge i Recerca: Jornades Antoni Varés*. Girona: CRDI. Retrieved from <http://www.girona.cat/sgdap/docs/vu23efriglesias.pdf>
- Lytard, J.-F. J.-F. (1979). *The Postmodern Condition: A Report on Knowledge*. (B. Massumi, Ed.) *Poetics Today* (Vol. 10, p. 886). University of Minnesota Press.
- Manovich, L. (2011). What is Visualization? *Visual Studies*.
- Manovich, L. (2012). Media Visualization: Visual Techniques for Exploring Large Media Collections. *Media Studies Futures*.
- Manovich, L., & Douglass, J. (2009). *Visualizing temporal patterns in visual media*.
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25:2, 137–166.
- Martín Carmona, R., & Repiso Caballero, R. (n.d.). IMDB y su utilidad para la investigación cinematográfica: ejemplos de uso de datos desde la metodología de análisis de redes sociales. In *IV Congreso Internacional sobre análisis fílmico*.
- Oliva i Barceló, N. (2010). *Metodología cualitativa d'anàlisi d'imatges*. Universitat Oberta de Catalunya. Retrieved from <http://hdl.handle.net/10609/1265>
- Oyarce, S. (2012). *In pursuit of image: How we think about photographs we seek*.
- Pérez-Montoro Gutiérrez, M. (2003). Gestión del conocimiento y documentación digital: un estudio de caso. *Hipertext.net*.
- Pintanel, M. À. (2008). Foto fixa. Més enllà dels marges. A 10es Jornades Imatge i Recerca.

- Pintanel, M. À. (2010). 11 anys de catalogació de fotos fixes a la Filmoteca de Catalunya. A 11es Jornades Imatge i Recerca.
- Riambau, E., & Torreiro, C. (2008). *Productores en el cine español: estado, dependencias y mercado*. Madrid: Filmoteca Española.
- Sangüesa Solé, R. (2010). El procés de descobriment de coneixement a partir de dades. Objectius, fases i problemàtica. In *Mineria de dades*. Barcelona: Universitat Oberta de Catalunya.
- Sangüesa Solé, R., Mor i Pera, E., & Molina Félix, L. C. (2010). *Mineria de dades*. Barcelona: Universitat Oberta de Catalunya.
- Tufte, E. R. (1997). *Visual explanation: images and quantities, evidence and narrative* (p. 156). Cheshire, Conn.: Graphics Press. Retrieved from <http://biblioteca.uoc.edu/llobres/9019.htm>
- Trybula, W. (1997). Data Mining and Knowledge Discovery. *Annual Review of Information Science and Technology*, 32.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining. Practical Machine Learning Tools and Techniques* (3rd ed.). Burlington: Morgan Kaufman Publishers.

Notes

¹ Són còpies d'altres materials positius destinades, sobretot, a processos fotomecànics.

² De fet, la millor posició es trobaria utilitzant com a lloc de rodatge l'Albufera de València, però aquest no és un element que es pugui comparar amb altres autors ja que només Emili Godes fa fotografies a aquest lloc, i per tant no hi ha correlacions possibles amb altres autors, i, a més, el nombre de fotografies que hi realitza és elevat. Aquesta eina és utilitzada per a diagnòsi mèdica comparant diversos símptomes (Demšar, 2010) i està pensada per a comparar més dades i amb relacions no tan marcades com l'esmentada. Cal dir que la finalitat d'aquestes eines de mineria de dades és mostrar la tendència a partir dels exemples coneguts de quines són les probabilitats i, per tant, si en tots els exemples de l'Albufera de València l'autor ha estat Godes és fàcil que es demostrés aquest fet des del punt de vista teòric.