Tècniques de Webmetria,
anàlisi de logs i Google Analytics per a la obtenció de variables mètriques del web de la Biblioteca Virtual (BV) de la Universitat Oberta de Catalunya (UOC)

David Oistrach
Responsable de l'explotació i l'anàlisi de la col·lecció i els serveis bibliotecaris
Av. Drassanes, 3-5 - 08001 Barcelona
Tel. 93 481 72 83 / 93 317 70 75
http://biblioteca.uoc.edu
doistrach@uoc.edu

Resum

En aquesta presentació es vol donar a conèixer l'ús i el context d'aplicació de dues tècniques de webmetria, l'anàlisi de logs i Google Analytics, vigents actualment a la Biblioteca Virtual de la UOC. D'acord amb aquesta finalitat es proporciona en primera instància una introducció global a la webmetria i, just a continuació, abordem el cas de la Biblioteca Virtual de la UOC destacant l'assimilació d'aquestes tècniques i les consideracions implícites al seu ús, i englobant de forma integral el procés de captació, processament i explotació de les dades. En darrer terme també es donen pautes per a la interpretació de les variables mètriques obtingudes.

Paraules clau: Webmetria; Anàlisi de logs; Indicadors; Google analytics; Mesurament del web; Analítica web.

1. Introducció

La BV de la UOC, ha adoptat recentment, noves tècniques de webmetria per a conèixer l'ús i l'explotació de la seva plataforma web i els recursos i serveis basats en llenguatge HTML.

La BV, com totes les biblioteques universitàries, realitza des de la seva fundació una anàlisi estadística de les principals variables d'activitat; tanmateix, atesa la seva presència únicament virtual, sempre ha necessitat poder disposar de forma molt acurada i detallada de la informació i els indicadors relatius a l'ús de la plataforma web, a partir de la qual ofereix serveis i continguts.

En aquesta presentació es pretén donar a conèixer el projecte d'implementació de les tècniques de webmetria utilitzades i les principals dificultats trobades. Les dues tècniques de webmetria que la BV ha adoptat són l'anàlisi transaccional de logs i la implementació de Google Analytics.

2. Els indicadors de la Biblioteca Virtual de la UOC

En una unitat organitzativa, entenent unitat organitzativa de forma àmplia i sintètica com un conglomerat de recursos i un equip de persones que treballen en equip per a aconseguir uns objectius, es poden identificar processos o factors operatius estratègicament claus per a la consecució d'èxit. En el cas d'una empresa privada, aquests factors claus els podríem associar a circumstàncies molt diverses vinculades als seus productes i serveis que conformen la seva cadena de valor. Bons exemples d'aquests factors podrien ser el cost d'adquisició de la matèria prima en un període de temps determinat, el marge de benefici en la venda d'un producte a un segment del seu públic objectiu o el temps que en una consultoria dediquen els equips a projectes d'un determinat client. De factors crítics per a les organitzacions n'hi ha de molt diversos i són molt específics, ja que han d'estar alineats als seus objectius concrets però tots tenen una característica en comú: gràcies a la sistematització i digitalització dels processos mitjançant les tecnologies de la informació, aquests factors es poden representar en una o «n» variables quantitatives i és d'interès per a la direcció de l'organització mesurar-les, i així poder fer-ne seguiment. En el context empresarial aquestes variables que tenen un valor estratègic per a la organització reben el nom de *KPI's* (Key Performance Indicators).

Per poder contextualitzar les tècniques de webmetria objecte d'aquesta comunicació que s'apliquen a la Biblioteca de la UOC és necessari en primer lloc decriure què entenem per *indicador*: és una informació planificada i regular que ens aporta coneixement estratègic útil a partir de l'anàlisi i la presentació d'una o més variables combinades.

És una informació planificada i regular perquè es delimita el seu abast, format i freqüència en concordança i per a respondre a les necessitats d'informació de les persones responsables de la direcció de la Biblioteca. Per a definir un indicador, inicialment s'identifica quina és la informació que necessitem per a prendre decisions encertades i posteriorment s'investiga el procés per a extreure-la dels sistemes d'informació. Volem remarcar la diferència notable entre aquesta perspectiva, proactiva i exigent en la qualitat de la informació de què es vol disposar per a la gestió, d'una altra de més directa que és la d'adaptar-se a la informació que els mateixos sistemes d'informació poden en alguns casos oferir per defecte.

Conseqüentment, els indicadors són el producte resultant de l'aplicació d'unes tècniques en la plataforma tecnològica de la Biblioteca per a recuperar, processar i presentar la informació que prèviament s'ha definit acuradament fruit d'un procés analític i introspectiu d'identificació de les necessitats informatives per a la direcció.

La Biblioteca Virtual de la UOC, afavorida pel fet de ser una unitat organitzativa que fa un ús intensiu de les TIC i viu en l'entorn digital, té definits els seus KPI, a partir d'ara anomenats directament *indicadors*, en tres categories:

— *Indicadors de situació*: ens proporcionen coneixement de l'estat de la qüestió del producte, que en el nostre cas són els documents de la col·lecció física, els

recursos de la col·lecció digital, les accions de formació, la difusió realitzada i els subproductes generats periòdicament, com per exemple: reculls d'actualitat, butlletins de notícies i sumaris i resums digitalitzats de llibres i revistes. Els serveis i el vincle existent entre la Biblioteca i els seus usuaris es basa en la difusió i l'accés a aquest coneixement heterogeni que resideix ens aquests actius. Les tècniques d'explotació de bases de dades són les més habituals per a recuperar i processar la informació que donen forma a aquests indicadors.

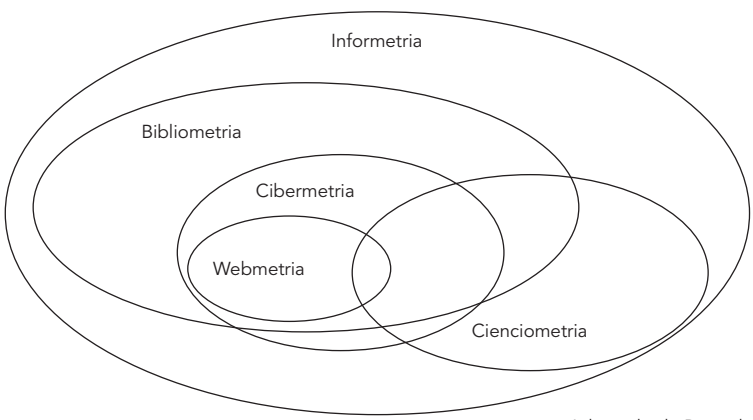
— *Indicadors d'ús*: ens proporcionen coneixement sobre l'ús que es fa dels recursos i serveis que la Biblioteca posa a l'abast dels seus usuaris. En els sistemes d'informació i recursos basats en web s'apliquen majoritàriament tècniques de webmetria.

— *Indicadors de satisfacció*: ens proporcionen coneixement sobre el grau de satisfacció dels usuaris en relació amb els serveis que la Biblioteca posa a la seva disposició i les seves expectatives. L'eina més consolidada per captar aquest *feedback* i la valoració percebuda per part de l'usuari és la distribució d'enquestes. És una tipologia d'indicador en la qual, com a fet distintiu dels altres dos, es fa necessària la col·laboració i participació activa de l'usuari.

3. La webmetria

La webmetria apareix en el context de la cibermetria, disciplina que estudia els aspectes quantitatius en relació amb els continguts i els processos de comunicació que es produeixen en el ciberespai, independentment del sistema d'informació objecte d'anàlisi, i es caracteritza per centrar-se en els sistemes basats en web des de la perspectiva bibliomètrica i informètrica.

Gràfic 1: Webmetria en el context de la informetria (Björneborn 2002).



Adaptado de Björneborn (2002)

Paral·lelament a la consolidació d'Internet, com a mitjà viable per al comerç electrònic, s'ha vist impulsat el desenvolupament de la webmetria amb la finalitat de mesurar i conèixer l'ús que es fa de les plataformes web. Aquest desenvolupament de la webmetria va en la línia de les fortes inversions que les empreses destinen a potenciar la seva presència i posicionament a Internet, on les transaccions en línia van en augment i van adquirint un pes específic rellevant en l'economia mundial. És un fet que un dels factors claus i de màxim al·licient per aplicar tècniques de webmetria és la definició d'objectius: determinar què volem que l'usuari faci en la nostra seu web i saber quin ús n'està fent realment per així poder fer un seguiment lligat al retorn de la inversió.

A la Biblioteca Virtual de la UOC, l'ús de la webmetria encaixa i té interès per a poder avaluar l'ús que es fa de la nostra seu web i així guanyar coneixement sobre quines seccions són més visualitzades i, per tant, de més rellevància per l'usuari final. D'altra banda —una opció sovint més interessant—, també permet identificar quines seccions del web no són consultades i així, un cop identificades gràcies a la webmetria, podem intentar extrapolar-ne les possibles causes i les possibles solucions.

Addicionalment, les tècniques de webmetria també ens possibiliten conèixer característiques dels visitants, des de quin sistema operatiu, navegador i configuració de pantalla utilitzen fins a la seva ubicació geogràfica, el temps de la visita i la recurrència en un període de temps concret.

4. L'ús de la webmetria a la BV UOC

A continuació aprofundirem en com la Biblioteca de la UOC ha adoptat i combina l'ús de dues tècniques de webmetria, l'anàlisi transaccional de logs i la implementació de Google Analytics, per a conèixer l'ús i l'explotació de la seva plataforma web i els recursos i serveis basats en llenguatge HTML.

4.1. L'anàlisi transaccional de web logs

Des del seus inicis la Biblioteca Virtual de la UOC ha tingut present que, atesa la seva presència únicament virtual, era imprescindible mesurar l'ús dels seus recursos i serveis electrònics i la seva plataforma web. És per aquest motiu que des de la seva fundació l'any 1995 es varen implementar uns procediments associats a una tècnica aleshores incipient: l'anàlisi de logs del servidor web.

Els web logs són un tipus d'arxiu de text pla generats pels servidors web on s'enregistra la seva activitat, la qual consisteix en el subministrament d'arxius als clients connectats prèvia petició. Quan un dispositiu es connecta a una seu web es realitza una petició al servidor per a cada arxiu que compon la pàgina web visualitzada i cada petició queda enregistrada en una línia de l'arxiu de log, d'aquesta manera:

86.42.132.114 – – [31/Oct/2005:18:15:16 -0500] «GET /style/css/estilos. css HTTP/1.1» 200 5194 «http://biblioteca.uoc.edu/cat/home.html» «Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.12) Gecko/20050915 Firefox/1.0.7»

La informació enregistrada d'esquerra a dreta és la següent:

1. *86.42.132.114*: IP de la màquina client.
2. *[31/Oct/2005:18:15:16 -0500]*: data i hora de la connexió.
3. *GET /style/css/estilos. css HTTP/1.1*: arxiu demanat especificant el mètode GET i el protocol HTTP/1.1.
4. *200 5194*: codi 200 resultant de la petició, que indica que s'ha tramitat correctament; la xifra *5194* són els bytes que ha generat la tramesa de l'arxiu demanat.
5. «*http://biblioteca.uoc.edu/cat/home.html*»: *referral* o pàgina web des de la qual l'usuari ha demanat l'arxiu, en aquest cas l'arxiu estilos. css.
6. «*Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.12) Gecko/20050915 Firefox/1.0.7*»: informació del sistema operatiu, idioma i navegador de l'usuari; en aquest cas utilitza Windows XP en anglès i ha accedit a la pàgina web de la biblioteca amb Firefox.

Un cop vist aquest exemple hem de tenir present que un arxiu de log conté milers de registres similars a aquest; per tant, per poder processar aquesta informació en brut ens cal disposar d'un programari que ens permeti treballar amb l'arxiu de log, processar els camps que hem vist de forma automatitzada i fer una explotació d'aquesta informació ajustada i parametritzada als nostres criteris. A la Biblioteca de la UOC fem servir el programari Funnel Web Analizer de l'empresa Quest Software.

L'operativa bàsica i comuna de qualsevol programari per a realitzar una anàlisi transaccional de web logs consisteix en la configuració dels filtres d'inclusió i exclusió per a delimitar l'abast de l'anàlisi. Per a fer una anàlisi de logs orientada a les nostres necessitats és vital conèixer prèviament molt bé l'organització i l'estructura de les carpetes del servidor web i saber exactament què volem mesurar.

Disposem de dos tipus de filtres:

— *Filtres de les extensions d'arxiu*: serveixen per especificar els tipus d'arxius que entren dins l'abast de l'anàlisi. Per exemple: HTML, PDF, SWF, etc.
— *Filtres delimitadors de les URL*: el seu ús acota les URL del domini objecte de l'anàlisi. Són necessaris per a aplicar l'anàlisi a una part concreta de la seu web, per exemple, una secció. Si prescindim d'utilitzar-los l'anàlisi es realitza en tota la plataforma web, incloent totes les seves parts components. L'aplicació d'aquests filtres és molt versàtil i els podem definir per inclusió o exclusió d'URL que continguin, comencin, acabin o siguin igual a una cadena de caràcters unívoca a la o les URLs de la part component del web a analitzar.

Un cop definit l'abast, el programa d'anàlisi de logs processa l'arxiu de log del servidor web aplicant les condicions definides en els filtres i podem obtenir les següents variables mètriques.

— *Sessions d'usuari*: nombre i temps mitjà de durada de les visites virtuals que els usuaris han realitzat al web o la part component definida pels filtres. Una sessió va lligada a la visualització d'una o més pàgines.
— *Pàgines visualitzades*: nombre de visualitzacions de documents HTML.
— *Visitants*: nombre estimat de persones un cop comptabilitzades les diferents adreces IP que han accedit i consultat al web o la part component del web subjecte d'anàlisi. Aquesta dada es desglossa en usuaris recurrents i no recurrents en funció de si han accedit un cop o més d'un cop en el període temporal que abasta l'arxiu de log, habitualment un mes.
— *Arxius descarregats*: aplicant els filtres que especifiquen els tipus d'arxiu a tenir en compte en l'anàlisi de logs podem conèixer el nombre de descàrregues d'un tipus d'arxiu determinat, per exemple PDF, en una secció concreta o en el conjunt de seccions de tota la seu web. Aplicant els filtres d'inclusió i exclusió d'URL de forma molt exclusiva podem mesurar fins i tot el nombre de descàrregues d'un determinat arxiu en qüestió, independentment del seu format.

A la BV de la UOC aquestes variables quantitatives sobre la plataforma web les recuperem mensualment aplicant dos procediments.

El primer procediment consisteix a realitzar una anàlisi de log del servidor web sense aplicar cap mena de filtre delimitador del domini biblioteca. uoc.edu però definint els formats i tipus d'arxius vàlids: HTML, DOC, PDF, XML, PPT…, en definitiva, tots els que tenen un valor documental. Els resultats ens proporcionen dades totals sobre l'ús de la plataforma web.

El segon procediment consisteix a executar una seqüència d'anàlisis en l'arxiu de log estratificats pels filtres d'inclusió i exclusió d'URL i seguint el principi d'especificitat amb l'objectiu d'obtenir les variables mètriques particulars sobre:

— L'ús de la plataforma web, en relació amb les seccions principals: informació general, col·lecció digital, serveis i funcionalitats.
— L'ús de funcionalitats com per exemple el motor de cerca global Verity, el catàleg bibliogràfic OPAC VTLS o el servei de difusió electrònica de sumaris DESU.

L'ús d'aquesta tècnica no està exempt d'alguns inconvenients:

— Intervenció manual recurrent: realitzar una anàlisi de logs requereix aplicar uns procediments manuals per a cada execució. Aquest procés implica descarregar mensualment l'arxiu de log del servidor web i processar-lo per obtenir les dades

parcials i globals de la nostra plataforma web. Els resultats per a cada anàlisi es presenten en un informe en format PDF. Per a poder agrupar les dades i explotar-les —com per exemple definint un abast temporal per curs i semestre— és convenient que estiguin dipositades en la taula d'una base de dades. Aquest procés d'inserció de les variables mètriques en una taula de base de dades a partir dels informes PDF resultants de l'anàlisi de l'arxiu de log, és complex d'automatitzar. Actualment s'efectua de forma manual.

— Dificultat per conèixer la navegació dels usuaris: acotant amb els filtres podem saber, entre d'altres variables mètriques, el nombre de sessions, les pàgines visualitzades o els usuaris estimats a partir de l'adreça IP del client. Són dades quantitatives sobre ús però que no ens indiquen clarament la navegació que es realitza a la plataforma web. En seus web amb seccions molt entrellaçades les unes amb les altres es fa impossible saber des d'on ha accedit un usuari a un determinat contingut (*referrals*).
— Impossibilitat de fer seguiment d'enllaços externs: l'arxiu de log només recull l'activitat del nostre servidor web, peticions rebudes i arxius enviats al remitent. Aquest abast té com a conseqüència que no es pugui saber el nombre d'accessos a un enllaç determinat que apunti fora del nostre domini, perquè l'accés a un recurs extern no queda recollit en l'arxiu de log.

4.2. Google Analytics

A principis de l'any 2000 i a mesura que Internet —i, més concretament, la World Wide Web— ha anat integrant llenguatges d'execució en la presentació de continguts estàtics, l'analítica web ha fet un pas endavant en la monitorització de l'ús de pàgines web. Concretament, aquest avenç es consolida amb l'aparició d'una nova tècnica que consisteix a etiquetar amb un script javascript les pàgines HTML de les quals vulguem obtenir dades d'ús.

Aquest script porta implícit un codi alfanumèric únic que capta el tràfic generat en temps real i el comunica a un servidor central on es processen les dades per a obtenir les mètriques de mesura del web. L'usuari pot consultar-les en qualsevol moment, definint l'abast temporal de les dades i acotant el nivell de profunditat i detall, a través d'un plafó de control sense necessitat de coneixements tècnics en informàtica.

Aquesta tècnica, que forma part d'un sistema integrat client-servidor, es va popularitzar l'any 2004 gràcies a Google i el seu servei gratuït Google Analytics, basat en aquesta arquitectura. La Biblioteca Virtual de la UOC va decidir l'any 2006 fer-ne ús i abordar la seva implementació i explotació efectiva.

Per a la seva instal·lació vàrem decidir mesurar de forma diferenciada els tres entorns idiomàtics de la plataforma web, és a dir, les versions catalana, castellana i anglesa. Amb aquest propòsit es van crear tres perfils diferenciats de Google Analytics i es

va inserir en tot el conjunt de pàgines web de cada versió idiomàtica un script de Google Analytics independent.

L'script javascript inserit és similar a aquest:

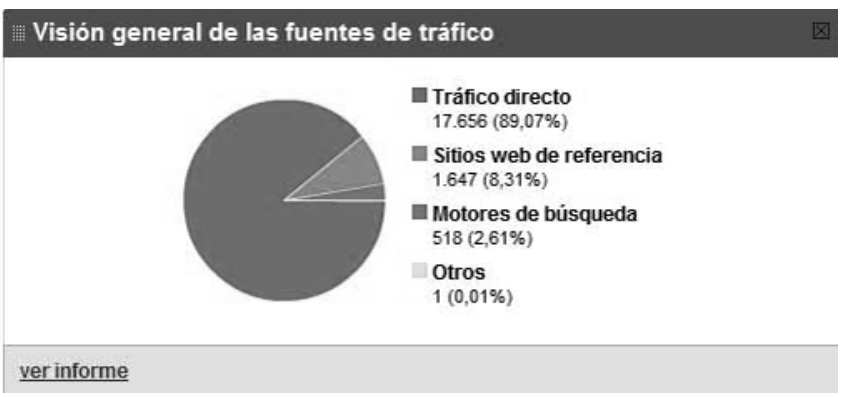
```
<script src=»http://www.google-analytics.com/urchin.js»type=»text/javascript»>
</script>
<script type=»text/javascript»>
_uacct = «UA-xxxxxxx-x»;
urchinTracker();
</script>
```

Finalitzada l'operació mitjançant el plafó de control web accessible a l'URL *http://www.google.com/analytics/es-ES*, vàrem disposar de tot un conjunt de noves variables mètriques addicionals a les mencionades amb anterioritat:

— *Percentatge d'abandonament*: per a cada pàgina individual de la plataforma web de la Biblioteca Virtual de la UOC podem constatar el percentatge estimat del tràfic de xarxa que abandona la versió idiomàtica a la qual pertany. Aquest càlcul es realitza controlant la navegació de cada sessió d'usuari fins al punt on l'usuari tanca el navegador o bé accedeix a una pàgina on no hi ha implementat l'script de Google Analytics corresponent a la versió idiomàtica.

— *Agregació i categorització de les fonts del tràfic* (referrals): es comptabilitzen els orígens del tràfic de xarxa per a cada pàgina monitoritzada i cada accés es categoritza en quatre epígrafs: tràfic directe, motors de cerca, llocs web de referència i altres. Com a resultat obtenim dades quantitatives i percentatges respecte al total d'accessos i en relació amb el lloc des d'on accedeix l'usuari.

Gràfic 2: Fonts de tràfic web BV UOC.



— *Característiques dels equips clients*: ubicació geogràfica, sistema operatiu, idioma, resolució i colors de pantalla del navegador utilitzat, versions de Flash i compatibilitat amb llenguatge Java dels equips dels nostres usuaris. Conèixer totes aquestes característiques dels usuaris que es connecten a la nostra plataforma web ens serveix per definir els requeriments tecnològics a satisfer per a una correcta interacció web-persona en un model de disseny centrat en l'usuari.
— *Rendiment del contingut*: variable interessant que ens permet saber el percentatge del tràfic total que aglutina cada pàgina individual respecte al conjunt de pàgines de què disposa cadascuna de les versions idiomàtiques de la plataforma web. Aquest coneixement ens possibilita inferir la rellevància i el pes específic d'una pàgina individual respecte del conjunt en funció del nombre d'accessos a la mateixa pàgina. Cal destacar, però, que no sempre es pot extrapolar que un alt nombre d'accessos a una pàgina sigui directament proporcional a una major satisfacció de l'usuari en l'experiència de navegació i consulta.

Com s'ha pogut constatar, l'ús de Google Analytics comporta una sèrie d'avantatges a considerar, entre els quals: automatització de la captació i processament de les dades, entorn web dinàmic i complet per a la seva explotació, variables mètriques sobre l'ús del web addicionals i noves funcionalitats per a la interpretació, com per exemple, i entre d'altres, l'opció de «Superposició del lloc web». Aquesta funcionalitat ens permet visualitzar una pàgina del domini analitzat amb Google Analytics i consultar en la mateixa pàgina i en temps real el nombre de clics efectuats en cada enllaç, sempre que estigui programat en llenguatge HTML i s'utilitzi l'etiqueta *<a href>*.

Coneguts els seus punts forts, és pertinent destacar també els punts febles identificats en el transcurs del projecte d'implementació que la Biblioteca Virtual de la UOC va dur a terme:

— Incompatibilitat amb seus web amb marcs: l'ús de marcs (*frames)* en una seu web és una pràctica considerada no recomanable pels problemes associats en la seva indexació en els motors de cerca i la seva recuperació posterior. Tot i això, la plataforma web de la Biblioteca Virtual de la UOC disposa els continguts en diferents marcs afegint un grau de complexitat en la implementació de l'script de Google Analytics. Aquestes dificultats es deuen al fet que si afegim l'script als tres *frames* que componen una visualització de pantalla el nombre de visites, sessions d'usuari i totes les variables mètriques calculades es multipliquen pel nombre de vegades que l'script de Google Analytics s'executa. D'altra banda, si no afegim l'script en els tres *frames* i només monitoritzem el marc principal de visualització de continguts, no podrem conèixer l'ús que es fa dels enllaços dels marcs on no hi hagi l'script de seguiment. Finalment, l'ús de funcionalitats avançades com la «Superposició del lloc web» no és compatible per defecte amb l'ús de marcs, perquè el seu funcionament està vinculat a la càrrega d'un únic document HTML.

— Residència de les dades en els servidors de Google: és convenient ser conscients que les dades s'emmagatzemen i es consulten en els servidors propietat de Google. Encara que podem fer exportacions de les dades en Excel, PDF i XML, el conjunt de dades històriques en brut es troben en poder de Google i aquest fet origina una certa relació de dependència.
— Dificultats de la implementació en sistemes d'informació integrats a la plataforma web: en sistemes de gestió de continguts, com per exemple els gestors de continguts CMS, i per extensió en tots aquells que generin pàgines HTML de forma dinàmica, és necessari fer modificacions en el disseny per tal que generin les pàgines amb el codi de Google Analytics inserit. En cas contrari, quan l'usuari accedeixi a una pàgina del nostre domini generada dinàmicament, aquest tràfic deixa de mesurar-se i es contempla com a percentatge d'abandonament. A la plataforma de la Biblioteca Virtual de la UOC aquestes pàgines generades dinàmicament provenen de les consultes al catàleg (VTLS), a la col·lecció digital (DIMAX) i en la presentació de resultats pel motor de cerca Verity.
— Seguiment orientat i associat a documents HTML estàtics: l'script de Google Analytics només es pot inserir en documents HTML dificultant la medició d'altres formats que s'utilitzen i són presents a Internet, com per exemple els arxius XML per a la sindicació de continguts amb l'estàndard RSS. En aquests casos la tècnica d'anàlisi de logs és el mètode més precís, perquè tot i que podríem monitoritzar els clics realitzats a l'enllaç que apunta al canal RSS, hem de tenir en compte que la sindicació de continguts implica accedir al fitxer XML sense utilitzar directament l'enllaç, a través de la sincronització entre una eina client gestora de canals RSS i el servidor web on resideix l'arxiu XML.

5. Conclusions

Diverses conclusions es poden inferir de l'experiència en l'ús i explotació de les dues tècniques d'analítica web vigents avui i d'aplicació a la Biblioteca Virtual de la UOC.

En primer lloc, no podem interpretar les variables mètriques sobre els visitants i les sessions d'usuari com a dades precises i exactes. Independentment de la tècnica utilitzada, el motiu pel qual no podem fer aquesta interpretació es dóna pel fet que la comptabilització d'aquestes variables es realitza a partir de les IP que es connecten al nostre servidor web. Aquest és el principal problema. És habitual que una mateixa persona disposi d'una IP diferent cada cop que es connecta a Internet a causa de l'assignació dinàmica d'adreces IP que realitzen els proveïdors d'Internet. D'altra banda, amb una mateixa IP pot accedir a Internet més d'una persona si comparteixen l'accés en una xarxa d'àrea local, fet habitual en organitzacions i empreses.

Un altre problema de caire tècnic són els sistemes de proxy-cache que utilitzen els proveïdors que faciliten accés a Internet als seus usuaris. Aquests sistemes s'imple-

menten en els servidors del proveïdor que integren la seva porta d'enllaç a Internet i tenen per funció optimitzar el tràfic de xarxa i l'ampla de banda disponible. Per aconseguir-ho, sovint generen còpies de les pàgines més sol·licitades, i quan l'usuari es connecta a una pàgina web molt consultada, en realitat no la descarrega del servidor web on està ubicada la pàgina, si no que la consulta directament de la replica emmagatzamada en la memòria del sistema proxy-cache. Aquests sistemes tenen una freqüència d'actualització molt elevada, com a màxim d'un parell de minuts, i afecta a les seus web d'ús més recurrent. Quan l'usuari consulta una web emmagatzamada en aquests sistemes de proxy-cache no genera consultes en el servidor web, i per tant, la tècnica d'anàlisi de logs no reflexa la seva activitat. Aquesta casuística no afecta a l'ús de sistemes com Google Analytics gràcies a que la replica emmagatzemada, al ser una còpia literal, també porta l'script de Google Analytics en el seu codi font.

Tot i que hi ha factors que ens impedeixen interpretar els valors com a dades exactes, les desviacions es mantenen en el temps i, per tant, serveixen per conèixer la magnitud estimada del nombre de sessions, visitants i pàgines visualitzades de la plataforma de web o d'una secció concreta al llarg del curs acadèmic. També és una font vàlida per detectar on poden haver-hi problemes d'interacció a partir de tendències d'ús o no ús, o els efectes dels canvis que es realitzen en l'arquitectura del web, de les seves directrius en usabilitat o directament dels seus continguts. Tanmateix, cal tenir present que una major freqüència d'ús no necessàriament implica una major satisfacció de l'usuari en la consulta i navegació de la seu web o d'un recurs electrònic. Fins i tot es pot donar el cas contrari: un recurs difícil d'utilitzar i que no satisfaci uns mínims criteris d'usabilitat pot generar moltes més consultes per desorientació de l'usuari i proves d'assaig-error.

Més concretament, l'experiència en la implementació de tècniques webmetria des dels seus inicis ens ha permès també, d'una banda, identificar l'especial rellevància que té conèixer, assimilar i adoptar el conjunt de bones pràctiques establertes en programació HTML, arquitectura de la informació i disseny usable d'interfícies web per a facilitar l'aplicació i explotació posterior d'aquestes tècniques de webmetria. En aquesta línia, la Biblioteca Virtual de la UOC, fruit del seu interès pel mesurament del web des de la seva fundació, ha acumulat una experiència sòlida en analítica web que li permetrà establir els requeriments tecnològics i formals bàsics a satisfer en els nous projectes de desenvolupament web per a garantir una explotació de dades més acurada. D'altra banda, i fruit d'aquesta experiència, també destacaríem la idoneïtat de fer un ús combinat de les dues tècniques d'anàlisi pel fet que aquestes són plenament compatibles i complementàries l'una amb l'altra. Aquest ús combinat ens permet aprofitar els avantatges de cada tècnica per a nodrir els indicadors d'ús vinculats a sistemes basats en web i així obtenir coneixement que ens faciliti la presa de decisions.

Bibliografia

ARROYO, Natalia (2005). «Cibermetría: estado de la cuestión». A: *Novenas Jornadas Españolas de Documentación, FESABID 2005*. <http://www.fesabid.org/madrid2005/descargas/presentaciones/comunicaciones/arroyo_natalia.pps> [Consulta 09/01/2008]

BUSTOS-GONZÁLEZ, Atilio (2007). «Bibliotecas universitarias: ¿sabemos medir sus resultados e impactos?». *El profesional de la información*, vol. 16, núm. 4 (juliol-agost).

BJÖRNEBORN, Lennart (2004). «Small-world link structures across an academic web space: a library and information science approach». Tesi doctoral. <http://vip.db.dk/lb/phd/phd-thesis-ch2-webometrics.pdf> [Consulta 26/11/2007]

GROVES, Karl (2008). «The limitations of server log files for usability analysys». *Boxes and Arrows* (gener). <http://www.boxesandarrows.com/view/the-limitations-of> [Consulta 09/01/2008]

HARLEY, Diane; HENKE, Jonathan (2007). «Toward an effective understanding of website users: advantages and pitfalls of linking transaction Log analyses and online survey». *D-Lib Magazine*. Vol. 13, núm. 3/4. <http://www.dlib.org/dlib/march07/harley/03harley.html> [Consulta 18/12/2007]

WHITE, Andrew; KAMAL, Eric D. (2006). *E-Metrics for library and information professionals: how to use data for managing and evaluating electronic resource collections*. Londres: Facet Publishing.