

Xarxa Punt TIC



MÒDUL 1 NIVELL AVANÇAT

Recerca i recuperació de la informació a Internet (avançat)

El web opac

→ ÍNDEX

| | |
|---|----|
| ÍNDEX | 2 |
| B. El web opac..... | 3 |
| Metodologia de la investigació | 4 |
| Elecció de paraules clau per al posicionament web | 5 |
| Anàlisi dels factors del posicionament web..... | 8 |
| Freqüència d'aparició i posició de les paraules clau | 8 |
| Metadades..... | 8 |
| Popularitat, textos d'ancoratge i tràfic de visites..... | 10 |
| Conclusions | 11 |
| Taller pràctic d'indexació..... | 13 |
| Per què necessito bons continguts?..... | 13 |
| Haig d'actualitzar constantment els continguts?..... | 14 |
| Quines paraules clau (keywords) utilitzo? | 14 |
| Hauria de tenir el meu propi domini?..... | 16 |
| Quin domini haig d'escollir?..... | 16 |
| Quins són els principals tipus de pàgines dinàmiques? | 17 |
| En què em pot beneficiar utilitzar pàgines dinàmiques? | 18 |
| Per a quin navegador dissenyo el meu lloc web? | 19 |
| Per què cal aconseguir enllaços?..... | 20 |
| Com puc conèixer el PageRank d'una pàgina?..... | 21 |
| Haig d'utilitzar els sistemes automàtics d'enviament a cercadors? | 21 |
| Com puc aconseguir aparèixer a DMOZ? | 22 |
| Tenen alguna relació Google i DMOZ? | 22 |

→ B. El web opac

Pàgines que poden ser indexades, però no són incloses en els cercadors. Els motius perquè els cercadors “decideixen” no incloure-les poden ser:

- Profunditat d'exploració: els llocs tenen “profunditat”. La *home page* o pàgina principal és el primer nivell; allà arriben tots els cercadors. Des d'allà s'enllacen a pàgines internes del lloc, aquest seria el segon nivell, al qual no arriben els directoris i alguns motors de cerca. Aquestes pàgines, alhora, enllacen amb pàgines més internes, que no estaven a la pàgina principal. A aquest nivell arriben molt pocs cercadors. Quant més profund sigui el nivell, menys cercadors l'indexaran.
- Freqüència d'exploració: un lloc pot canviar tots els dies, però molts dels robots dels cercadors que exploren els llocs els visiten un cop al mes, o menys, per una qüestió de cost. Tots els canvis entre una visita i una altra no figuren en els cercadors.
- Supera el nombre màxim de resultats: cada cercador defineix quantes pàgines d'un lloc mostrarà. Si un lloc té més pàgines que les que el cercador inclou, les altres quedaran sense indexar.
- Errors d'exploració: pot haver un problema en el lloc o en el robot del cercador (o en la compatibilitat entre els dos) que impedeix que una pàgina (o fins i tot un lloc sencer) s'inclougui en la base de dades del cercador.

Avui dia la majoria dels accessos a llocs web es realitza a través de motors de cerca, per tant, és fonamental per als seus responsables assegurar-se que apareixen ben posicionats en els resultats de cerca, tant des del punt de vista del màrqueting com per donar un millor servei als seus usuaris.

En conseqüència, el lloc que un web ocupa en el llistat de resultats d'un motor de cerca quan un usuari realitza una consulta en el cercador, és un aspecte molt rellevant per als responsables de llocs web, i les accions per a la seva millora s'expliquen pel “posicionament web” (Codina, 2004; Codin, Marcos, 2005; Arbildi, 2005). L'eix central és preguntar-se quines seran les paraules que previsiblement utilitzaran els usuaris potencials en les seves cerques. Una vegada determinat això, es podran utilitzar les tècniques lícites d'optimització perquè un determinat lloc web aparegui en una bona posició quan els usuaris busquin informació relacionada amb els continguts d'aquest lloc web.

Els llocs web que allotgen bases de dades terminològiques també poden i han d'utilitzar les tècniques de millora del posicionament per facilitar als seus possibles usuaris l'accés a aquests recursos a través dels cercadors. La nostra investigació ha pres com a mostra deu bases de dades terminològiques la

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

consulta de les quals és lliure al Web, presenten multilingüisme i pertanyen a diferents temàtiques (taula 1).

| Bases de dades | URL |
|--|---|
| CercaTerm(català) | http://www.termcat.net |
| Eurodicautom | http://europa.eu.int/eurodicautom |
| EuskalTerm | http://www1.euskadi.net/euskalterm |
| OncoTerm | http://www.ugr.es/~oncoterm/ |
| TerminoBanque | http://www.cfwb.be/franca/bd/bd.htm |
| TIS: Terminological Information System | http://tis.consilium.eu.int |
| UBTerm | http://www.ub.edu/slc/ubterm |
| UNTerm | http://unterm.un.org |
| WTOTerm | http://wtoterm.wto.org |

Taula 1. Bases de dades terminològiques estudiades.

Metodologia de la investigació

Per facilitar que els cercadors els trobin, els llocs web han de millorar tots els aspectes que aquests cercadors consideren a l'hora d'establir el rànquing de resultats. Sense entrar a explicar en detall les tècniques per a l'optimització de llocs web, esmentem els criteris que estant usant els cercadors en els seus rànquings de resultats (Codina; Marcos, 2005):

- Freqüència (absoluta o relativa) de l'expressió o terme buscat a la pàgina web (la qual anomenarem "paraula clau"), sempre que no es caigui en una repetició abusiva que els cercadors considerin *spam*. En aquest estudi s'han plantejat tres possibles paraules clau per a cada lloc web.
- Posició: lloc on es troba el terme dins la pàgina; es tindrà en compte que les metadades i el primer paràgraf tenen més pes que altres parts de la pàgina.
- Metadades: a més de les metadades de la secció *head* (principalment *title*, *description* i *keywords*), hi ha altres etiquetes que també proporcionen informació descriptiva útil per als cercadors, per exemple el

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

títol dels enllaços i de les imatges, així com el text alternatiu (alt) de les imatges.

- Popularitat: nombre d'enllaços externs que rep la pàgina web. Aquest criteri està relacionat amb el *PageRank* determinat per la barra d'eines de Google.
- Anclatge: text que serveix com a enllaç per arribar a aquesta pàgina web.
- Tràfic de visites que rep la pàgina web, considerant alhora el temps de permanència dels usuaris. Ve donat pel *TrafficRank* de la barra d'eines d'Alexa.

D'aquests factors, la nostra investigació està centrada fonamentalment en les metadades i la popularitat. Els altres factors també s'han estudiat, però no han donat resultats que ens ajudin a valorar el grau d'importància que tenen per al posicionament web dels llocs analitzats.

L'estudi ha considerat dos tipus d'anàlisi: en primer lloc, s'ha dut a terme una anàlisi empírico-descriptiva, atenent els aspectes que la bibliografia sobre posicionament web indica que cal tenir en compte. A partir dels resultats obtinguts en aquesta fase, s'han plantejat algunes hipòtesis que s'han posat a prova mitjançant l'anàlisi estadística ANOVA; es tracta d'una anàlisi multivariable en què una variable dependent és creuada amb algunes variables independents.

Per a la seva aplicació s'ha utilitzat el programa Statgraphics Plus 5.0. La variable dependent considerada ha estat el posicionament web dels llocs web, i les variables independents amb les quals s'ha creuat han estat cadascun dels llocs web estudiats, els cercadors sobre els quals s'han fet les consultes, les paraules clau de les consultes i les metadades *title*, *description* i *keywords*. L'encreuament de dades dona com a resultat el p-valor, que és el nivell de significació empíric en un contrast d'hipòtesis. Es considera que una hipòtesi és nul·la en els casos en què el p-valor és inferior a 0,05, i que és alternativa sempre que supera aquesta xifra.

Elecció de paraules clau per al posicionament web

Abans d'analitzar les causes del posicionament en llocs web, ja que aquesta dada constitueix la variable dependent en l'anàlisi estadístic, s'ha comprovat quin és aquest posicionament per a tres consultes per a les quals aquests llocs web haurien d'estar ben posicionats, entenent per una bona posició els 10 primers resultats dels cercadors més utilitzats avui dia. S'ha valorat de forma especialment positiva (2 punts) que el lloc web aparegui en el primer o el segon lloc, de forma parcial (1 punt) si apareix entre els llocs 3 i 10, i no s'ha valorat (0 punts) si està més enllà del resultat número 10.

Les paraules de cerca escollides s'han traduït a l'idioma principal de la interfície de cada eina terminològica, seguint aquests criteris:

- Paraula clau 1 (PC1): nom de la base de dades.

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

- Paraula clau 2 (PC2): sintagma que descriu el tipus de recurs terminològic de què es tracta en cada cas.
- Paraula clau 3 (PC3): sintagma que descriu una qualitat específica del recurs terminològic de què es tracta en cada cas.

| | PC2 | PC3 |
|---------------|-----------------------------------|--------------------------------------|
| CercaTerm | Base de dades terminològica | Terminologia catalana |
| WTOTerm | Terminology database | World trade terminology |
| TerminoBanque | Banque de donées | Terminologie spécialisée |
| Eurodicautom | Terminology database | Multilingual specialized terminology |
| EuskalTerm | Banco terminológico | Terminología euskera |
| TIS | Terminological information system | Consilium terminological database |
| UBTerm | Base de dades terminològica | Terminologia catalana |
| UNTerm | Terminology database | United Nations terminology |
| OncoTerm | Base terminológica de oncología | Terminología oncológica |

Taula 2. Paraules clau utilitzades per buscar en els 6 motors.

El mes de gener de 2006 es van realitzar les tres cerques en sis dels cercadors més utilitzats avui dia (Google, Yahoo! Search, MSN Search, Altavista, Teoma i Vivísimo). El resultat obtingut és molt diferent per a cadascuna de les paraules clau (taula 3):

Les cerques per la PC1 obtenen amb més freqüència el lloc web esperat entre els 10 primers resultats, en especial amb Yahoo! Search i Google, tot i que no tant amb Teoma i Vivísimo, que només aconseguen localitzar aquests llocs web en un 50% de les cerques.

El resultat canvia per a la PC2, ja que dos llocs web no s'han trobat entre els 10 primers resultats, i d'altres no ocupen les primeres posicions. Google i Yahoo! Search mostren aquests llocs web en millors posicions que els altres cercadors. Teoma, en canvi, només mostra en aquest primer grup de resultats 4 dels 10 llocs web.

La cerca per la PC3 mostra resultats diferents als anteriors: tot i que segueix en la línia de la PC2 i alguns cercadors no col·loquen aquests llocs web entre els

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

10 primers resultats, el cercador Teoma aquest cop ofereix les millors posicions, seguit de Vivísimo i Google.

| Bases de dades | PC1 | PC2 | PC3 | Mitjana de cada base de dades |
|---------------------|-------|-------|-------|-------------------------------|
| Eurodicautom | 100,0 | 50,0 | 100,0 | 83,3 |
| TIS | 33,0 | 91,6 | 100,0 | 74,8 |
| WTOTerm | 100,0 | 25,0 | 91,6 | 72,2 |
| EuskalTerm | 100,0 | 83,3 | 0,0 | 61,1 |
| CercaTerm | 91,6 | 0,0 | 50,0 | 47,2 |
| OncoTerm | 41,6 | 100,0 | 0,0 | 47,2 |
| UNTerm | 91,6 | 25,0 | 16,6 | 44,4 |
| UBTerm | 75,0 | 33,0 | 0,0 | 36,0 |
| TerminoBanque | 100,0 | 0,0 | 0,0 | 33,3 |
| Promedio de cada PC | 72,3 | 50,8 | 45,8 | 83,3 |

Taula 3. Classificació de les bases de dades en funció de la mitjana del posicionament que ocupa cada lloc web al cercador per la PC1, PC2 i PC3. El valor màxim, 100, indica que es troba en el primer o el segon lloc; els valors intermedis, propers a 50, indiquen que es troba entre el lloc 3 i el 10, i el valor 0 indica que no és entre els 10 primers resultats.

L'anàlisi estadística corrobora que existeix una diferència significativa entre el posicionament dels llocs web estudiats i les paraules clau per a les quals s'ha provat el seu posicionament, ja que per a la PC1 el posicionament obtingut és molt millor que per a les PC2 i PC3 (figura 1). Al mateix temps, mostra que l'ús d'un cercador o un altre no provoca diferències significatives en els resultats de posicionament, ja que Google, Yahoo! Search i Vivísimo presenten aquests llocs web més ben posicionats que els altres tres cercadors utilitzats.

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

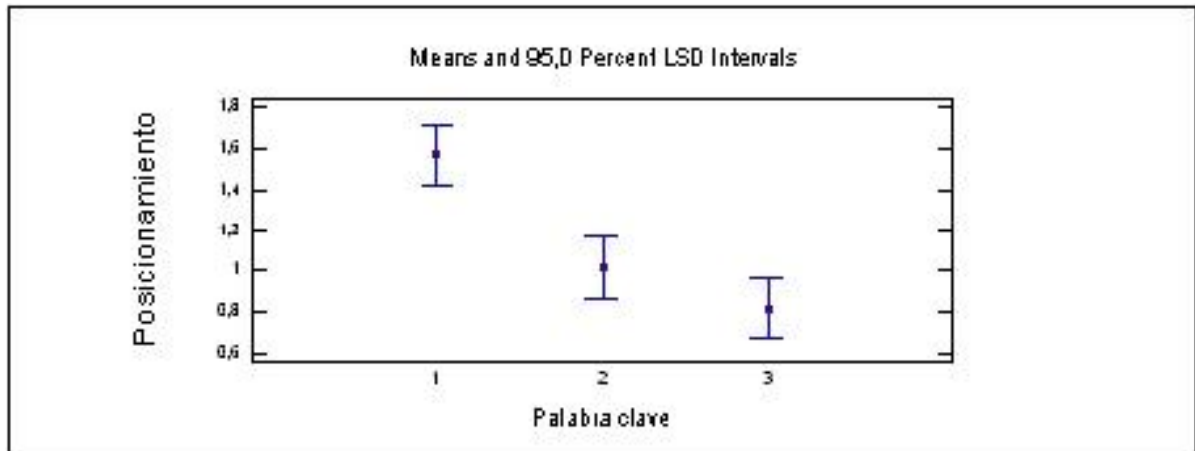


Figura 1. Posicionament dels llocs web en relació amb PC1, PC2 i PC3.

Anàlisi dels factors del posicionament web

A partir dels punts indicats en l'apartat 2, presentem els resultats més rellevants obtinguts en l'anàlisi empírico-descriptiva i en l'anàlisi estadística.

Freqüència d'aparició i posició de les paraules clau

Tot i que estem d'acord amb la importància d'aquests factors per millorar el posicionament web, hem decidit no considerar-ho en aquest estudi per tres motius: en primer lloc, les interfícies de les eines de cerca d'informació terminològica no compten amb un volum de text suficient per poder determinar valors de freqüència òptims; i en segon lloc, amb l'excepció del nom de la base de dades, les altres paraules clau establertes com a consultes (PC2 i PC3) no solen aparèixer reflectides en els sistemes estudiats, és per això que aquest criteri no ens permetrà establir comparacions.

Metadades

Pel que fa a les metadades dels llocs web estudiats, hem comprovat que el 90% tenen el camp *title* emplenat, el 40% inclou el camp *keywords* i només el 20% presenta el camp *description* (taula 3). No s'han considerat les etiquetes *title* i *alt* dels enllaços i les imatges, ja que en els llocs estudiats el seu nombre era molt baix o fins i tot nul, per tant, no afectaria al posicionament web d'aquests llocs.

| Bases de dades | Title | Keywords | Description |
|----------------------|-------|----------|-------------|
| Base de Terminologie | X | - | - |
| CercaTerm | X | X | - |
| Eurodicautom | X | X | - |
| EuskalTerm | X | X | X |

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

| | | | |
|---------------|---|---|---|
| OncoTerm | X | X | - |
| TerminoBanque | X | - | - |
| TIS | X | - | X |
| UBTerm | X | - | - |
| UNTerm | - | - | - |
| WTOTerm | X | - | - |

Taula 4. Ús de metadades en els llocs web de les bases de dades estudiades.

En les anàlisis estadístiques, l'encreuament del posicionament dels llocs web amb la informació de metadades posa de manifest que només el camp *description* implica diferències significatives de posicionament entre llocs web (figura 2), mentre que l'existència de *keywords* no influeix tant en els resultats obtinguts (figura 3). No s'ha pogut obtenir un resultat fiable sobre la repercussió de l'etiqueta *title*, ja que 9 de cada 10 llocs web la tenien, és per això que no ens deixa marge per establir comparacions.

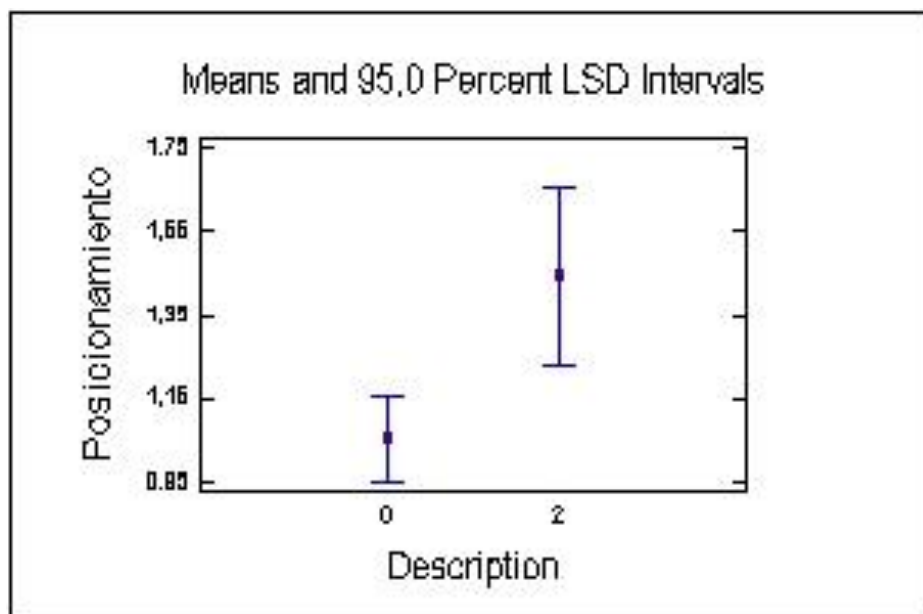


Figura 2. Posicionament dels llocs web considerant si tenen l'etiqueta meta *description* (2) i si no la tenen (0).

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

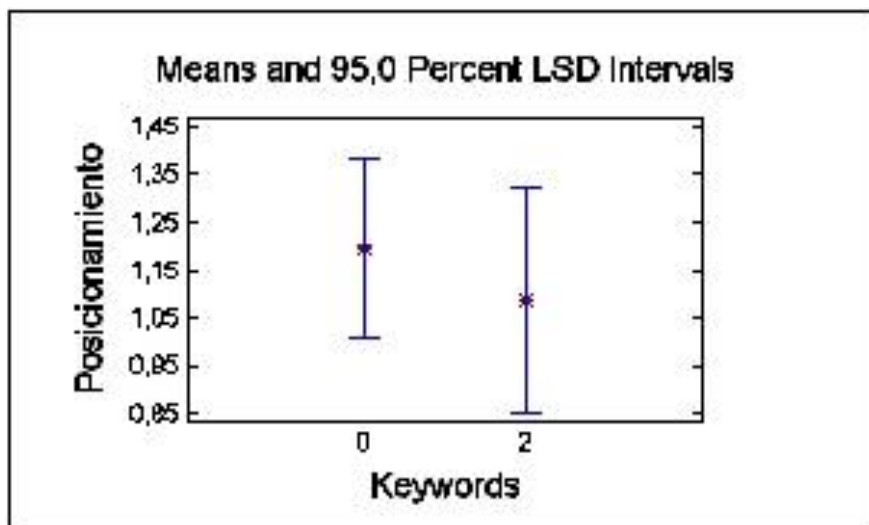


Figura 3. Posicionament dels llocs web considerant si tenen l'etiqueta meta *keywords* (2) i si no la tenen (2).

Popularitat, textos d'ancoratge i tràfic de visites

La popularitat, entesa com el nombre d'enllaços que apunten cap a un lloc web, es pot conèixer parcialment a través de la cerca amb el limitador *link:* que ofereixen alguns cercadors. Mostrem els valors que dona Yahoo! Search (taula 5), que són xifres molt més altes que les que proporciona Google.

| Bases de dades | Enllaços d'arribada (Google) |
|----------------------|------------------------------|
| Eurodicautom | 2.340 |
| CercaTerm | 1.290 |
| Base de Terminologie | 1.030 |
| TIS | 458 |
| EuskalTerm | 243 |
| TerminoBanque | 163 |
| WTOTerm | 104 |
| UNTerm | 79 |
| UBTerm | 49 |
| OncoTerm | 9 |

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

Taula 5. Classificació de les bases de dades en funció del nombre d'enllaços que apunten a cada lloc web segons Yahoo! Search.

Si comparem els resultats que obtenim en les taules 3 i 5, podem veure com existeix una relació bastant directa entre el nombre d'enllaços d'arribada de cada lloc web i els resultats que han donat per al posicionament per les tres paraules clau. Les bases de dades amb més "cites" són alhora les més ben posicionades per a aquestes paraules. A l'extrem contrari, UBTerm i UNTerm, que obtenen menys "cites", també apareixen en una posició pitjor. El cas de CercaTerm, que no està tan ben posicionada i en canvi rep moltes "cites", és perquè la xifra del nombre d'enllaços d'arribada que té no fa referència a la pàgina principal de la base de dades, sinó a la de la seva institució (TermCat), ja que per arribar a la base de dades cal passar per un formulari a la pàgina d'inici de la institució, cosa que dificulta l'accés directe als cercadors. En el cas d'aquest sistema, si no canvia la manera d'accedir a la base de dades, s'haurà d'optimitzar la pàgina d'inici de la institució per millorar-ne el posicionament en les cerques relacionades amb la seva eina terminològica.

Pel que fa als textos usats en els ancoratges d'aquests enllaços, s'ha revisat els cinc primers de la llista que ofereix Google i el resultat obtingut ha estat que, amb l'excepció de la PC1 (el nom de la base de dades), no s'han utilitzat les paraules clau escollides en aquest estudi. Per tant, és un criteri que no ens serveix per establir comparacions i detectar la rellevància que té en el posicionament de llocs web.

També s'ha anotat el valor que atorga Google com a *PageRank*, tot i que finalment no l'hem considerat d'interès per a aquest estudi, ja que es tracta d'un valor obtingut amb un càlcul poc transparent. El mateix criteri ens ha portat a no tenir en compte el valor que dona Alexa com a *TrafficRank*, que, a més, no és específic per a la pàgina web que s'analitza, sinó que agafa el valor de la seva pàgina d'inici.

Conclusions

Tot i que les tècniques d'optimització per a cercadors acostumen a emmarcar-se en el màrqueting, la nostra visió d'aquest concepte és àmplia, ja que creiem que no només es tracta d'una qüestió de publicitat, sinó també de servei als usuaris. En el cas de serveis d'ús gratuït, com és el de les bases de dades terminològiques que estudiem, també haurien de preocupar-se de posar les mesures necessàries per posicionar-se bé, ja que és el primer pas perquè es facin servir.

Quant a la metodologia utilitzada, l'anàlisi empírico-descriptiva ha estat útil per comprovar que cap dels llocs web estudiats ha realitzat una campanya de posicionament. De fet, en la majoria dels casos no s'han emplenat els camps de metadades. Dels camps bàsics que hem mencionat, *title* és el més freqüent (és en el 90% dels llocs web analitzats), seguit de *keywords* (40%) i, molt allunyat, *description* (20%).

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

No s'ha seguit una política d'enllaços per aconseguir una visibilitat millor, tant per als possible usuaris com per als cercadors. D'altra banda, l'anàlisi estadística multivariable (ANOVA) ha estat vàlid per comprovar les observacions realitzades i verificar hipòtesis. L'aplicació conjunta de les dues metodologies ens du a afirmar que la metadada *description* juga un paper important en el posicionament de llocs web, mentre que la metadada *keyword* no presenta aquesta evidència (figura 4). De totes maneres, podem afirmar que l'ús de metadades no és decisiu per al posicionament web, perquè es dona el cas de llocs web ben posicionats que no han fet servir metadades, és per això que no podem deixar de banda altres factors que influeixen en el posicionament, com ara la popularitat.

De fet, s'ha pogut observar que existeix una correlació entre la popularitat dels llocs web analitzats (el nombre d'enllaços que apunten cap a ells) i el seu posicionament per a les paraules clau escollides, que indica que el criteri de popularitat juga un paper molt important en la posició que els llocs web ocupen en els resultats de la cerca en motors. Aquesta popularitat hauria d'estar potenciada per uns textos d'ancoratge apropiats, de manera que es consolidin les paraules clau que les institucions escullin per al seu posicionament en el Web. En el cas dels llocs estudiats no s'ha observat aquesta política.

Pel que fa als cercadors utilitzats en l'estudi, Google és el que ha presentat un comportament més regular en relació amb l'ús de les metadades en els llocs web, i Google, Yahoo! Search i Vivísimo són els tres motors que han mostrat millors posicionaments per a les paraules clau escollides entre aquests 10 llocs web. En la majoria dels cercadors els llocs web apareixen posicionats quan es realitzen consultes per paraules més específiques (PC1), a excepció de Teoma i Vivísimo, en què els llocs web obtenen millors posicionaments en cerques més generals (PC2 i PC3). Per exemple, en el cas d'Eurodicautom, la cerca per la PC3 no presenta aquest lloc web entre els primers resultats de Yahoo! Search i, en canvi, per a Teoma ocupa el segon lloc.

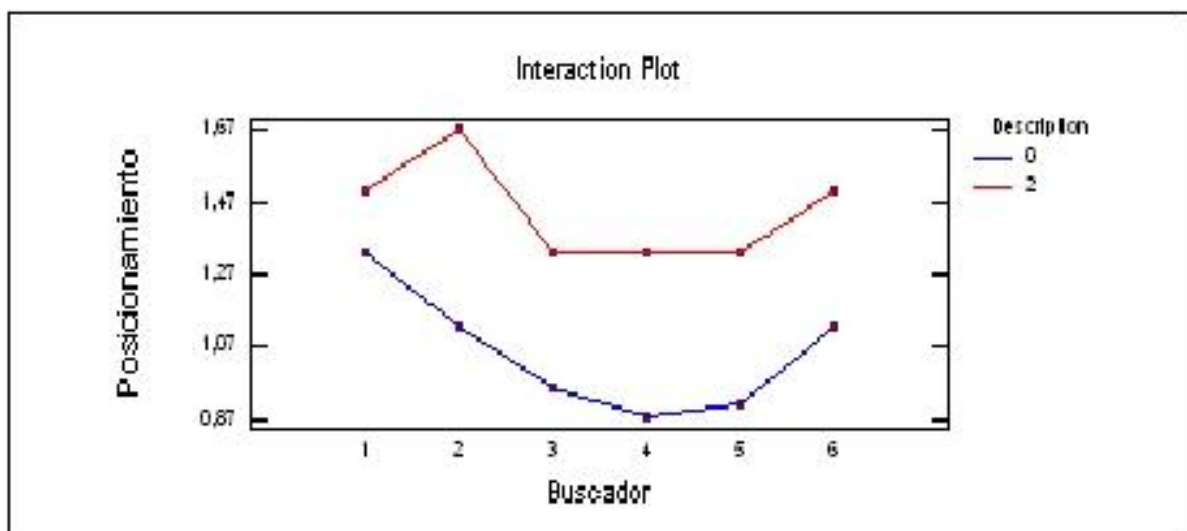


Figura 4. Posicionament dels llocs web en sis cercadors, considerant si tenen l'etiqueta meta description (línia vermella) o no la tenen (línia blava): Google (1), Yahoo! Search (2), MSN Search (3), Altavista (4), Teoma (5) i Vivísimo (6).

Taller pràctic d'indexació

SEO és la tasca d'aconseguir aparèixer en els primers resultats dels cercadors per a determinades cerques, sigla en anglès de *Search Engine Optimization*, optimització per a motors de cerca.

Per aconseguir-ho s'apliquen diferents tècniques i estratègies que poden ser molt diverses. Segons aquestes tècniques, podem classificar SEO en dos grans grups: [SEO Black Hat](#) (barret negre) i [SEO White Hat](#) (barret blanc).

Tot i que els objectius són els mateixos per a les dues, sortir en les primeres posicions d'un cercador, les seves tècniques són molt diferents. Mentre que el *SEO Black Hat* intentarà amb tots els mitjans aconseguir el posicionament, el *SEO White Hat* també intentarà el mateix, però d'una altra manera més subtil i natural per tal de no posar el *website* en perill. Sobre aquest punt aprofundiré molt més en altres articles.

Avui dia el treball d'un **SEO** és tan ampli i variat que abasta des de sòlids coneixements de programació fins a la psicologia per saber com reaccionaran els usuaris. Per exemple, segons el títol que surti en els *serps* de Google, entraran o no. Un SEO mai deixa d'aprendre o deixarà de ser SEO, ja que els cercadors intenten evitar la seva manipulació dels resultats i corregeixen el seu algoritme segons el que veuen fer als SEO.

Aquí et presentem un petit manual de posicionament web a Google. Amb ell podràs aconseguir aparèixer a les primeres posicions dels resultats. Recorda que només podràs estar a la primera posició si t'esforces molt.

Tenir bons continguts és fonamental per a l'èxit segur d'un lloc web. Per una banda, aconseguiràs atraure un gran nombre de visitants que accediran a les teves pàgines regularment. D'altra banda, si saps com redactar aquests continguts, podràs fins i tot atraure més visites gràcies a Google. Intenta redactar triant determinades paraules clau (keywords) i aprèn on situar-les dins cada pàgina web.

Per què necessito bons continguts?

El contingut és el més important en un lloc web. Podràs conèixer tots els trucs i podràs aconseguir enganyar Google, però la manera d'aconseguir realment visites és amb uns bons continguts.

A més, si els continguts mereixen la pena aconseguiràs més enllaços dels *webmasters* d'altres llocs web. Com veurem més endavant, tenir molts enllaços és fonamental per aconseguir un bon posicionament a Google. No deixis de generar continguts i intenta construir pàgines regularment, amb bona informació.

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

Haig d'actualitzar constantment els continguts?

És una bona idea actualitzar periòdicament els continguts del teu lloc web per dos motius:

- -A Google li agraden els llocs que renoven i actualitzen els seus continguts. Valora que son llocs "vius" i que es pot comptar amb ells.
- Pots aconseguir que el robot Freshbot passi regularment pel teu lloc web. Aquest robot passa per les pàgines amb els continguts més "frescos" i actualitza aquests continguts a la base de dades de Google en unes hores. D'aquesta manera, pots modificar ràpidament els continguts del teu lloc web (per exemple, amb un nou producte o noves paraules clau) sabent que apareixerà a Google en un parell de dies.

Escull les paraules amb què vols aparèixer a la primera posició dels resultats de Google quan es busca per elles. Per exemple, "coches usados", "abogados en caracas" o "sms gratis".

Planeja amb antelació cada pàgina web i destina dues o tres paraules clau (keyword) per pàgina. És a dir, no intentis que la mateixa pàgina web aparegui en les primeres posicions de Google cercant per moltes paraules. Serà molt difícil aconseguir-ho.

Quines paraules clau (keywords) utilitzo?

Pot ser que tinguis un lloc web dedicat als negocis de cotxes, però no sàpigues quines paraules clau escollir. Una eina molt útil ens la proporciona Google en el [KeywordSandbox](#). En realitat es tracta d'una ajuda per escollir paraules amb el programa AdWords, però ens pot ajudar molt a l'hora d'escollir les nostres paraules clau.

Per exemple, quan introduïm la paraula "cotxes", aquesta eina ens suggereix "lloguer de cotxes", "cotxes usats" o "cotxes nous", a banda d'altres. A partir d'això, ens haurem de plantejar una estratègia amb les paraules clau.

Altres eines que suggereixen paraules clau:

- <http://es.esspotting.com/popups/keywordgenbox.asp>
- <http://inventory.overture.com/d/searchinventory/suggestion/>
- http://www.7search.com/scripts/advertiser/sample_get.asp

Tampoc s'han d'oblidar les pàgines que fan una classificació de les paraules més buscades o més populars, com el Zeitgeist de Google. Et podran servir d'ajuda per a noves paraules clau.

Altres pàgines que mostren les paraules més buscades:

- <http://sp.ask.com/docs/about/jeevesiq.html>
- <http://50.lycos.com/>

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

- <http://buzz.yahoo.com/>

TITLE: probablement el lloc més important. Intenta que en títol de la pàgina web apareguin les paraules clau desitjades. A més, fes un esforç per escriure títols no molt llargs (que no superin els 50 caràcters) i per no repetir més de tres vegades la mateixa paraula (Google ho pot considerar *spam*).

ALT: l'etiqueta ALT està present dins les etiquetes d'imatges d'aquesta manera:

```
<IMG src="mi_imagen.gif" ALT="Mi comentario">.
```

El text de l'etiqueta ALT va sorgir quan hi havia navegadors que no incloïen les imatges, i es mostrava aquest text en comptes de la imatge. Avui dia, alguns navegadors (com MS Internet Explorer) el mostren quan passem el ratolí per damunt la imatge.

Google considera aquest text, sobretot si la imatge és un enllaç a una altra pàgina web. Per això és convenient introduir paraules clau dins l'etiqueta ALT.

METATAGS: Google no considera els continguts dels següents *metatags*:

META NAME=keywords

META NAME=description

Aquest últim, però, és utilitzat de tant en tant per Google en comptes de l'*snippet* (la petita descripció que sol aparèixer en els resultats de Google) quan el contingut META coincideix amb la cerca realitzada.

URL: Google sí que valora que la URL (adreça de la pàgina web) contingui les paraules clau, tot i que no li dóna molta importància. Intenta que contingui les keywords desitjades, però no abuis i no intentis que el domini, el subdomini i el nom de la pàgina continguin aquestes paraules clau. Pots aconseguir que Google et penalitzi.

En les URL intenta separa els noms amb guions "normals" (-) i no amb un guió baix (_). Intenta escriure "mi-pagina.html" i no "mi_pagina.html".

En la resta de la teva pàgina web intenta situar diverses vegades les paraules clau que vols optimitzar. Tampoc abuis d'això, perquè els teus textos seran més difícils de llegir (recorda que dissenyes les pàgines per als usuaris, no per als cercadors).

A més, Google estima que determinats *TAGS* (etiquetes) reflecteixen millor la importància del text. Per exemple, situar un text entre les etiquetes <H1> i </H1> el destaca quan l'usuari veu la pàgina, però també Google considera que aquestes paraules són més importants, i ho tindrà en compte. Passa el mateix

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

amb les etiquetes <H2> (<H3>, <H4>,...), (negreta) i <I> (cursiva). Convé que repassis el vell HTML que has oblidat o que donis un cop d'ull a algun programa d'aprenentatge d'HTML.

Intenta dissenyar les pàgines web i els seus continguts perquè les paraules clau apareguin a les etiquetes, però tampoc abuisis d'això, ja que Google ho pot considerar *spam* i et pot penalitzar.

Per altra banda, hi ha eines a Internet que obtenen la densitat de paraules clau del teu lloc web. Pots trobar-les en aquesta cerca:

<http://www.google.com/search?q=keyword+density+analyzer>.

Cadascuna de les eines et donarà un resultat diferent, perquè en realitat Google utilitza un algoritme bastant complicat per estimar en quin grau una pàgina s'ajusta a determinades paraules clau. De totes maneres, pots utilitzar alguna de les eines suggerides i intentar que la densitat de les teves paraules clau a la teva pàgina web sigui del 5 al 20%.

A més, intenta que les paraules clau que has seleccionat apareguin en els enllaços que apunten cap a les teves pàgines web.

I, evidentment, vigila l'ortografia de les teves paraules clau. Ja sabem que molta gent escriu malament les paraules clau o que es confonen quan escriuen, però si busquem "abogados en caracas", Google ens suggerirà ràpidament "abogados en caracas".

Hauria de tenir el meu propi domini?

Sí. A banda de la millor imatge que pots oferir als teus visitants, pots optimitzar el teu posicionament a Google gràcies als enllaços.

Tenir el teu lloc web a "paginas.sitios-gratis.com/miempresa/" dóna una imatge bastant dolenta, i el preu d'un domini ja no és excusa perquè no en tinguis un de propi. Pots comprar-lo per menys de 10 dòlars l'any. Compara alguns preus dels [registrars](#) acreditats (és bastant més barat que altres venedors de dominis) i escull el teu.

Quin domini haig d'escollir?

La nostra recomanació és que siguis tu mateix i tinguis la teva pròpia marca. Com es comenta en aquest tutorial, has de dissenyar el teu lloc web per als visitants, no per als cercadors. Si el lloc és bo, la gent recordarà "tunombre.com", però difícilment "abogados-baratos-en-caracas.com".

Fixa't en els exemples de Google i Yahoo!. Són noms de marques no gaire senzills, però han aconseguit que els usuaris les recordin fàcilment. Fins i tot "got.com" va canviar el seu nom per "Overture".

Ara bé, tenir un domini del tipus "abogados-catalunya.com" et dóna l'opció que des d'altres pàgines web t'enllacin així:

<Ahref=http://www.abogados-catalunya.com>abogados-catalunya.com.

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

Això et donarà la possibilitat d'optimitzar el teu lloc web per a les paraules "abogados catalunya".

A més, si vols aparèixer en els resultats de Google que fan referència a un determinat país, hauràs de tenir un domini del tipus, per exemple, "midominio.cat" (si vols aparèixer en els resultats de Catalunya) o "midominio.cat" (en els d'Espanya). Google també t'inclourà en aquests resultats si el servidor on allotges les teves pàgines web està físicament en aquests països.

Dóna un cop d'ull a la categoria de DMOZ de [registre per països](#). Trobaràs el més adequat per al país en el qual vols aparèixer.

Les pàgines dinàmiques són pàgines HTML generades a partir de llenguatges de programació (*scripts*) que són executades en el mateix servidor web. A diferència d'altres *scripts*, com el JavaScript, que s'executen en el navegador de l'usuari, els "Server Side" *scripts* generen un codi HTML des del mateix servidor web.

Aquest codi HTML pot ser modificat, per exemple, per una petició realitzada per l'usuari en una base de dades. Segons els resultats de la consulta en la base de dades, es generarà un codi HTML o un altre, mostrant diferents continguts.

Quins són els principals tipus de pàgines dinàmiques?

Les pàgines dinàmiques s'executen en el servidor web propi, Per tant, dependran del tipus de servidor que tinguem. Per exemple, si tenim un servidor amb "Microsoft Windows Server", generalment trobarem un servidor web "Internet Information Server" (IIS) que executarà *scripts* "Active Server Pages" (ASP). Tot i que això no és sempre així, perquè actualment hi ha paquets de software que executen tots els *scripts* en tots els servidors, sempre estarem condicionats pels llenguatges dissenyats especialment per a cada sistema operatiu.

- **CGI**: abreviatura de Common Gateway Interface. És un estàndard per a la interacció entre aplicacions externes i servidors web. Gràcies a ell, podem adaptar qualsevol programa que hàgim realitzat en qualsevol llenguatge perquè actuï amb el nostre servidor. Tanmateix, Perl s'ha convertit en el llenguatge més popular per desenvolupar aplicacions CGI, tot i que també se sol utilitzar C, C++ o Fortran.
- **PHP**: llenguatge *script* de codi obert. Molt utilitzat sobre el servidor web Apache.
- **ASP**: llenguatge *script* creat per Microsoft per al seu servidor web 'Internet Information Server' (IIS), i basat en "Visual Basic Script". L'última versió "ASP.net" forma part del *Framework* ".net".
- **JSP**: llenguatge *script* creat per Sun, basat en la tecnologia Java. No cal que l'usuari disposi de la màquina virtual de Java, ja que es troba en el servidor que crea les pàgines HTML. No té res a veure amb els *applets* de Java, y tampoc amb JavaScript. Els *scripts* JSP són un cas particular dels *servlets*.

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

- **Cold Fusion:** llenguatge *script* creat per la companyia Allaire (adquirida més tard per Macromedia). Els *scripts* tenen l'extensió “.cfm”.

En què em pot beneficiar utilitzar pàgines dinàmiques?

Les pàgines dinàmiques ens poden ajudar a gestionar més fàcilment els continguts del nostre lloc web i a interactuar amb bases de dades.

Per exemple, si tenim un o diversos menús a les nostres pàgines i volem modificar-los, no haurem d'anar pàgina per pàgina per editar-los, sinó que només ho haurem de fer una vegada. En la resta de pàgines només caldrà incloure (en PHP, per exemple): `include 'menu-izquierda.html'`.

A més, tots els llenguatges *script* comentats tenen components per a la connexió amb la majoria de bases de dades (mySQL, Oracle, SQL Server, etc.). Això ens pot servir per a emmagatzemar els nostres continguts dins d'una base de dades, en lloc de realitzar cada pàgina web una per una.

Informa't de les capacitats de cadascun d'aquests llenguatges *script*, i dóna un cop d'ull als tutorials, que pots trobar a la Xarxa, de CGI Perl, PHP, ASP, JSP i Cold Fusion.

HTML és un estàndard proposat pel Consorci W3C, que té per objectiu aconseguir que tots els documents web siguin compatibles en qualsevol navegador (no només en ordinadors, sinó també en qualsevol dispositiu).

CSS són les sigles de *Cascade StyleSheet*, i especifica la forma del disseny dels documents. Una mateixa pàgina web (un mateix document HTML, per exemple) pot ser vista de diferents formes en un PC que una PDA, gràcies als diferents fulls d'estil CSS.

Utilitzar HTML+CSS et pot ajudar a millorar el teu posicionament web a Google. Per una banda, aconseguiràs que el codi de les teves pàgines web sigui més ampli i clar a ulls del robot de Google. Facilitar la tasca a aquest robot sempre és un punt a favor nostre.

Per altra banda, augmentaràs la densitat de les paraules clau dins els continguts (veieu “on situar les *keywords*”), ja que moltes de les etiquetes t'ocuparan molt menys espai. Això també suposa menys pes per a les teves pàgines web, cosa que Google agrairà. I podràs, alhora, canviar ràpidament els estils d'algunes paraules i modificar així la importància que els hi vols atorgar.

A més, complir amb l'estàndard HTML t'obrirà les portes a dissenyar pàgines web per a dispositius mòbils o per a noves tecnologies que vagin sorgint. I l'ús de CSS et permetrà canviar l'aspecte d'aquestes pàgines web en pocs minuts. En combinació amb les pàgines dinàmiques, pots aconseguir un lloc web realment eficient.

No li ho posis difícil al robot de Google. Si insereixes informació en els següents elements, assegura't que no seran reconeguts:

- JavaScript
- DHTML
- Flash

→ Recerca i recuperació de la informació a Internet (avançat)

El web opac

- Frames
- Session IDs
- Applets de Java
- Imatges: no insereixis textos dins d'elles.

Això no significa que no pugis utilitzar-los per al disseny del teu lloc web. Simplement que la informació que aparegui dins d'ells no apareixerà en les cerques de Google.

Per a quin navegador dissenyo el meu lloc web?

No intentis que les teves pàgines web es vegin millor amb un navegador d'Internet determinat o amb un altre. Intenta que es vegin correctament amb tots, però, sobretot, amb el navegador de Google, és a dir, amb el seu robot.

Si vols saber com veu el robot de Google les teves pàgines web, pots utilitzar el navegador Lynx. Es tracta d'un navegador en mode text, que no contempla les imatges ni els elements superflus com el JavaScript, Flash, etc.

Si utilitzes el Sistema Operatiu Linux, probablement el tindràs instal·lat per defecte, i hauràs d'escriure això en la shell:

```
# lynx http://www.mi-sitio-web.com
```

Si estàs en un altre sistema operatiu (MS Windows, MAC, etc.), el més senzill és accedir via web a un emulador de Lynx:

<http://www.delorie.com/web/lynxview.html>

La memòria cau de Google ("caché") només emmagatzema un límit de 101 kb. Si mires la informació que guarda Google de qualsevol pàgina, veuràs que no supera aquesta xifra. Se sospita que no s'indexa més enllà d'aquests límits, però no està demostrat.

Aquests 101 kb fan referència només a codi HTML (en el qual s'inclouen tots els textos i la informació). No es tenen en compte les imatges, els gràfics Flash, etc.).

De tota manera, i com a recomanació, intenta que les teves pàgines no tinguin un pes excessiu, és a dir, que ocupin poc espai. Si pot ser menys de 30 kb, millor. Tingues en compte que a la gent no li agrada esperar massa quan accedeix a una pàgina web, i molts es cansen d'esperar després de 3 o 4 segons i marxen a altres pàgines. La connexió per cable o banda ampla no està encara massa estesa i la majoria d'usuaris es connecten a Internet amb un mòdem o compartint la connexió de la seva empresa o la universitat.

PageRank (PR) és un valor numèric que representa la importància que una pàgina web té a Internet. Google considera que quan una pàgina col·loca un enllaç (*link*) a una altra és un vot per aquesta última.

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

Quants més vots tingui una pàgina, serà considerada més important per Google. A més, la importància de la pàgina que emet el seu vot també determina el pes d'aquest vot. D'aquesta manera, Google calcula la importància d'una pàgina gràcies a tots els vots que rebí, considerant també la importància de cada pàgina que emet el vot.

PageRank (desenvolupat pels fundadors Larry Page i Sergey Brin) és la manera que té Google de decidir la importància d'una pàgina. És una dada valuosa, perquè és un dels factors que determina la posició que tindrà una pàgina en els resultats de la cerca. No és l'únic factor que Google utilitza per classificar les pàgines, però n'és un dels més importants.

Cal tenir en compte que Google no considera tots els *links*. Per exemple, Google filtra i descarta els enllaços de pàgines dedicades exclusivament a col·locar *links* (anomenades *link farms*).

A més, Google admet que una pàgina no pot controlar els enllaços que apunten cap a ella, però sí que pot controlar els que aquesta pàgina col·loca cap altres pàgines. Per això, *links* cap a una pàgina no poden perjudicar-la, però sí que poden ser perjudicials per al seu *PageRankTM* els enllaços que una pàgina col·loqui cap a llocs penalitzats.

Si un lloc web té PRO, generalment és un web penalitzat, i serà poc intel·ligent col·locar un *link* cap a ella.

Una manera de conèixer el *PageRankTM* d'una pàgina és descarregant la barra de cerca de Google (només disponible per a MS Iexplorer). Apareix una barra on es mostra en color verd el valor del *PageRankTM* en una escala de 0 a 10. Llocs web amb PR10 són Yahoo!, Microsoft, Adobe, Macromedia i Google. Teniu una llista completa amb els llocs amb PR10.

L'algoritme de *PageRank* va ser patentat als Estats Units el dia 8 de gener de 1998, per Larry Page. El títol original és *Method for node ranking in a linked database*, i se li va assignar el número de patent 6.285.999.

Aconseguir enllaços és una de les tasques més crítiques en el posicionament web a Google. En funció del nombre d'enllaços que obtinguem, tindrem major *PageRank* o popularitat.

S'ha de ser extremadament curós amb la manera que realitzem l'enllaç. En funció de com es faci aquest enllaç, promocionarem unes paraules clau o unes altres.

A més, cal saber on aconseguir enllaços. Hi ha determinats llocs web on és més fàcil obtenir-los, però gairebé sempre cal esforçar-se per aconseguir-los i s'ha de fer un seguiment.

Per què cal aconseguir enllaços?

Google atorga un valor numèric a cada pàgina web que insereix a la seva base de dades. A aquest valor numèric l'anomena *PageRank*. Quant més gran sigui el *PageRank* d'una pàgina, més importància li haurà donat Google.

I aquest valor creix quants més llocs web enllacin a la teva pàgina (veieu "transmissió del *PageRank*"). Has de pensar que Google considera cada *link*

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

com un “vot”. A més, si aquests llocs web que t'enllacen tenen un *PageRank* elevat, el valor creixerà més, perquè el “vot” és de més qualitat.

Com puc conèixer el *PageRank* d'una pàgina?

Per conèixer aquest valor, pots descarregar-te la barra de cerca de Google (només disponible per a MS IExplorer), <http://toolbar.google.com/> en la qual hi ha un espai per mostrar el *PageRank* (PR) de cada pàgina que visites. Aquest valor varia entre 0 i 10.

Tanmateix, no has de centrar tots els teus esforços en aconseguir un PR elevat. El valor del *PageRank* d'una pàgina és important, però t'hauràs adonat que hi ha pàgines que, tot i tenir menys PR, estan posicionades per sobre d'altres amb un PR més gran per a determinades cerques.

Això ho han aconseguit, entre d'altres coses, perquè han optimitzat millor els continguts de les seves pàgines, perquè han realitzat uns bons enllaços i perquè han aconseguit inserir aquests enllaços en pàgines web bones.

Tot i que Google és el cercador per excel·lència (més del 60% dels usuaris l'utilitzen), no és l'únic. També hauries d'intentar aconseguir tràfic des d'altres cercadors. No tindran un *PageRank* tan gran como Yahoo! o DMOZ, però es tenen en compte tots els enllaços.

Dóna un cop d'ull a les següents categories de DMOZ. Pots trobar alguns directoris i cercadors que et poden ser útils:

1. <http://dmoz.org/Computers/Internet/Searching/Directories/>
2. http://dmoz.org/Computers/Internet/Searching/Search_Engines/
3. http://dmoz.org/World/Español/Referencia/Buscadores_y_directorios/

Haig d'utilitzar els sistemes automàtics d'enviament a cercadors?

La nostra recomanació és que no utilitzis els sistemes automàtics que prometen que donaran d'alta el teu lloc web a centenars o milers de cercadors. És millor que ho facis personalment, perquè molts cercadors detecten enviaments automàtics i descarten les pàgines web enviades amb aquests mètodes.

A més, aquests sistemes automàtics estan dissenyats per a un determinat moment. La majoria de cercadors modifiquen els formularis i requeriments per donar-s'hi d'alta, i els sistemes automàtics no ho modifiquen alhora. Pot passar també que simplement no funcioni o que, per exemple, no et donis d'alta en la categoria adequada.

Recorda que el posicionament web demana molt esforç i dedicació. Has de tenir paciència i ho has de fer tot personalment.

➔ Recerca i recuperació de la informació a Internet (avançat)

El web opac

Com puc aconseguir aparèixer a DMOZ?

Hi ha categories de dmoz.org que tenen un *PageRank* de 7 o 8, això vol dir que aconseguir un enllaç a DMOZ és molt valuós. A més, Google el pren com a referència per construir el seu propi directori “directory.google.cat”. No escullis la categoria amb un *PageRank* més alt, sinó la que més s’ajusti a la temàtica del teu lloc web.

Però aparèixer a dmoz.org no és una tasca fàcil, ja que els editors només inclouen llocs web de qualitat i que realment tinguin una relació amb la temàtica de cada categoria.

Navega per les categories de DMOZ. Descobreix quina és la que més s’ajusta a la temàtica del teu lloc web i pitja l’enllaç “agregar URL”. Si tens un lloc web en espanyol, el que més et convé és escollir la categoria a “World > Español”.

Has de tenir paciència amb la teva sol·licitud. Solen tardar varies setmanes, però no has d’aclaparar els editors. De totes maneres, pots contactar i debatre amb ells a “www.resource-zone.com”. Si en el termini de sis mesos no has aconseguit que t’enllacin, podries tornar a suggerir-los el teu lloc web.

Convé que esperis a tenir el teu lloc web realment llest abans de suggerir un enllaç a la gent de DMOZ. Si els suggereixes un lloc web “en construcció” segur que no te l’accepten i, possiblement, no te’l revisaran la propera vegada que els el suggereixis. No et precipitis i fes les coses amb calma.

Tenen alguna relació Google i DMOZ?

No. DMOZ permet la reproducció lliure dels seus directoris (amb una llicència especial) i Google simplement es limita a recollir els seus continguts des de l’any 2000.

Molts altres llocs fan el mateix que Google en el seu “directory.google.cat”, i recullen en les seves pàgines web un directori de categories amb els enllaços que té DMOZ. Google tracta aquests enllaços de la mateixa manera que la resta dels links, però, en aparèixer a més llocs web –a causa d’aquests “clons” de DMOZ–, el nombre d’enllaços es multiplica.