



Administrarea conținutului site-urilor Web

Ciprian Dobre
ciprian.dobre@cs.pub.ro



Conținut

1. Factori de optimizare pentru motoarele de căutare
2. Bazele unui design “Search Friendly”
3. Mentenanța site-urilor Web



Alphabet Soup: SEO, SEM

- **SEO** - Search Engine Optimization
 - procesul de creștere a calității și volumului de trafic către o pagină web, trafic ce provine de la motoarele de căutare.
 - creșterea volumului de trafic se poate realiza în primul rând printr-o poziție bună în motoarele de căutare, ținta fiind primele poziții pentru principalele cuvinte cheie.
- **SEM** - Search Engine Marketing

Factori de vizibilitate

(sau cum să determini
Google să-ți placă pagina)





Componentele de bază ale unei pagini

- **Componenta Text (Keyword)**
 - Cuvinte și fraze ce corespund lucrurilor pe care audiența target le tastează în motoarele de căutare
- **Componenta Legături**
 - Structura de navigare (prin site) pe care crawlerii ar trebui să o poată urma ușor
- **Componenta Popularitate**
 - Există site-uri ce includ legături către site-ul meu?



Text: Densitatea de cuvinte cheie (keywords)

- Alegerea unuia sau mai multor cuvinte cheie sau fraze pentru fiecare pagină
- Nu folosiți prea multe cuvinte cheie
 - Folosirea prea multor cuvinte cheie poate avea ca rezultat penalizarea sau ignorarea
- Verificați Densitatea de Cuvinte Cheie:
 - <http://www.webobjectives.com/keyword.htm>
 - <http://www.keyworddensity.com/>



Text: Evidențierea cuvintelor cheie

- Motoarele de căutare plasează valori (“weight”) termenilor în funcție și de locul în care sunt folosiți
- **Plasați cuvinte cheie în ...**
 - Tag-uri titlu
 - Text de tip *heading* sau *emphasize*
 - Text din body vizibil
 - Meta tag-uri de descriere
 - “Alt text” din imagini
- Titlul și tag-urile body sunt cele mai importante



Mitul metadatelor

Citat: “Metadata improves search relevancy”

Fals. Cu excepția tag-urilor titlu și descriere motoarele de căutare ignoră alte metadata.

Citat: “If more agencies were diligent about adding metadata, Google would pay attention”

Fals. Industria se focusează mai mult pe dezvoltarea de algoritmi ce determină relevanța pe baza conținutului decât pe ceea ce un autor sau un creat de metadata spune despre respectivul conținut



Componenta Legături

- Paginile nu vor avea rank bun dacă site-ul nu respectă o schemă de navigare
- Schema de navigare trebuie să fie placută utilizatorilor și motoarelor de căutare
- Creați un “site map”, plănuiți modul în care paginile vor fi legate între ele
- Evitați paginile încălcite



Scheme problematice de navigare

- Codare HTML slabă
- Image maps
- Frames
- JavaScript
- Dynamic Pages
- Flash



Crearea de legături

- Cuvintele cheie din legături indică crawlerelor detalii legate de paginile respective
- Cuvintele cheie din legături influențează relevanța paginii respective
- Evitați link-urile “click here”, creați legături precum:
 - **Bad:** Click [here](#) for more info on famous admirals.
 - **Good:** Visit our [naval history site](#) for more info on [famous admirals](#).



Componenta Popularitate

- Bazată pe numărul de pagini ce au legături către pagina voastră
- Cu cât sunt mai multe pagini populare ce au legături către pagina voastră, cu cât aceasta va deveni mai populară
- Toate motoarele de căutare folosesc diverși algoritmi de determinare a popularității
 - Algoritmul Google se numește *Page Rank*
- Oricare pagină Web primește o valoare a popularității bazată pe legături



Factorii Popularității

- Numărul și Popularitatea Legăturilor
 - Deveniți listați în Yahoo!, DMOZ (<http://www.dmoz.org/>)
 - Creați legături (sociale) cu alte site-uri
 - Faceți ca site-ul vostru să devină un *magnet* pentru alții
- Anchor Text
 - Cuvintele cheie pe care alții le folosesc pentru a descrie legăturile la site-ul vostru
- Popularitatea este asignată per pagină, nu pentru un între site
 - Popularitatea *nu* se moștenește



Legături cu exteriorul

- Popularitatea paginii nu este determinată de site-urile către care ea are legături interne, dar
- Legăturile cu exteriorul afectează popularitatea site-urilor către care aveți legături
- Legăturile interne și externe au cel mai mare impact asupra popularității



Excluderea roboților

- Excluderea Roboților prin Meta-Tag
`<head><meta name="robots" content="no index, nofollow"></head>`
- Fișierul Robots.txt
 - Se plasează în directorul root al serverului
 - Două elemente: User-agent, Disallow
 - Exemplu:
User-agent: *
Disallow: /cgi-bin/
Disallow: /scripts/
Disallow: /images/



Excluderea roboților

- Nu toate motoarele de căutare iau în considerare instrucțiunile indicate în robots.txt
- MSN și Yahoo! iau în considerare instrucțiunile mai des decât Google
- Nu excludeți niciodată msnbot ... nu veți mai avea căutare în siteh

"msnbot/1.1 (+http://search.msn.com/msnbot.htm)"

- Blocking bots is contrary to OMB's guidance

Robots.txt:
User-agent: msnbot
Disallow: /



Studiu de caz: NOAA Fisheries

- Termen: Fisheries
- #1 - FirstGov
- #6 - MSN
- #10 - Google (first page)
- #4 - Yahoo

Termen: Fish

- Nu apare între primele pagini în nici un motor de căutare



NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE



NOAA FISHERIES SERVICE: NORTHEAST REGIONAL OFFICE

[Home](#) | [About Us](#) | [Search](#) | [Directions](#) | [Site Map](#)

Welcome to the Northeast Regional Office



Search

Go fish!

This site All of NMFS

Northeast Regional Office (NERO) | 55 Great Republic Drive | Gloucester, MA 01930

Topics

- ◆ Sustainable Fisheries
- ◆ Protected Resources
- ◆ Habitat Conservation
- ◆ State & Federal Programs
- ◆ Fishery Statistics
- ◆ Environmental Analyses
- ◆ Cooperative Research
- ◆ Management & Budget
- ◆ Outreach & Communications
- ◆ Financial Services

Features

- ◆ Hot News
- ◆ Fishing Industry Homepage
- ◆ Permit Holder Letters
- ◆ Public Comments
- ◆ Permit Information
- ◆ Fishery Regulations
- ◆ Quota, Landing, & Trip Reports
- ◆ Job Announcements
- ◆ Other Links



Studiu de caz: NOAA Fisheries

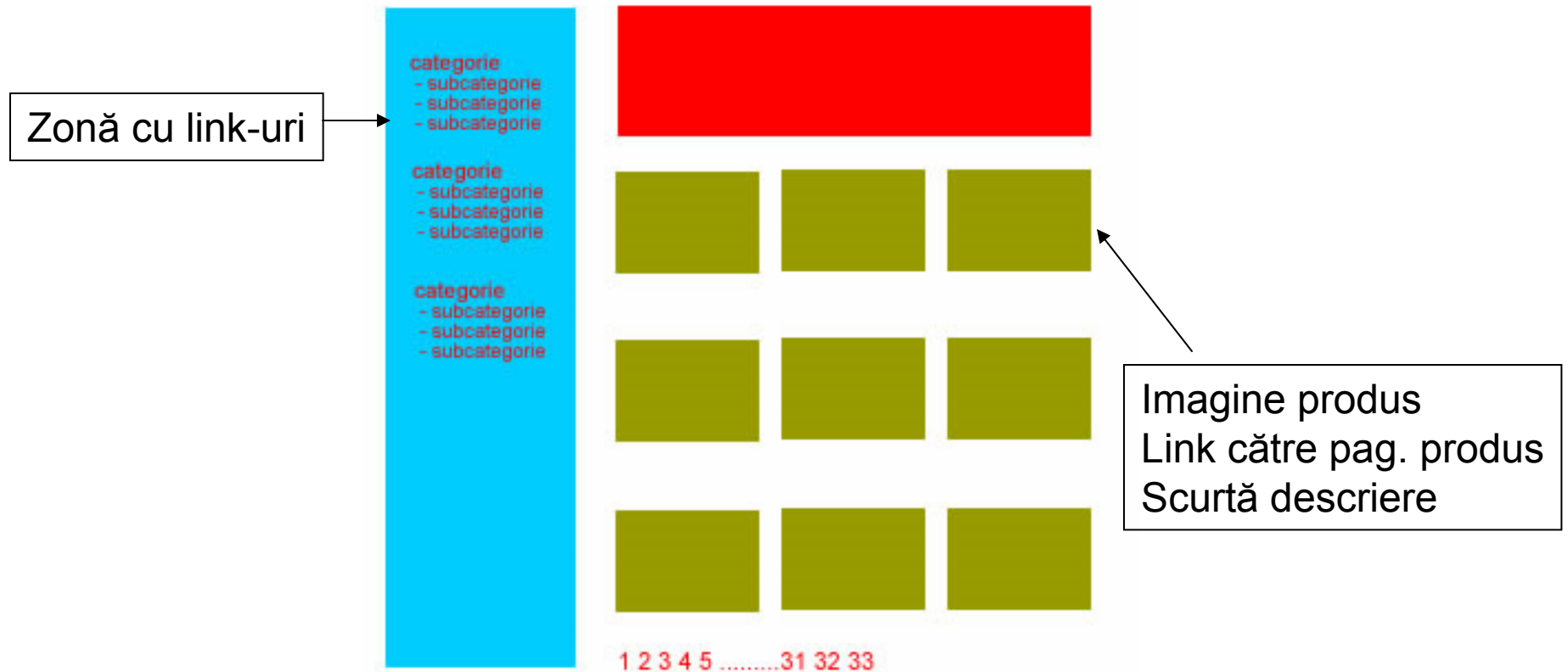
- **Componenta Text:** pagina este bine optimizată pentru “fisheries”, dar nu și pentru “fish”
- **Componenta Legături:** Navigarea de tip “Fly-out” (JavaScript) poate fi o problemă pentru unele crawlere. Adăugați navigare user-friendly, poate de tip search (e.g., site map)
- **Componente Popularitate:** Multe legături interne ce folosesc cuvântul “fisheries” (nu “fish”) ca text anchor



Link juice

- Conceptul exprimă greutatea, puterea, valoarea *(sau alți termeni similari)* pe care un link îl transmite documentului spre care trimite.
- Valoare proporțională cu **importanța paginii care publică linkul**, cu **numărul de legături primite** și cu **numărul de legături publicate** în acest document.

- Una din problemele cu care se confruntă magazinele online este diluarea link juice-ului trimis spre paginile interioare.



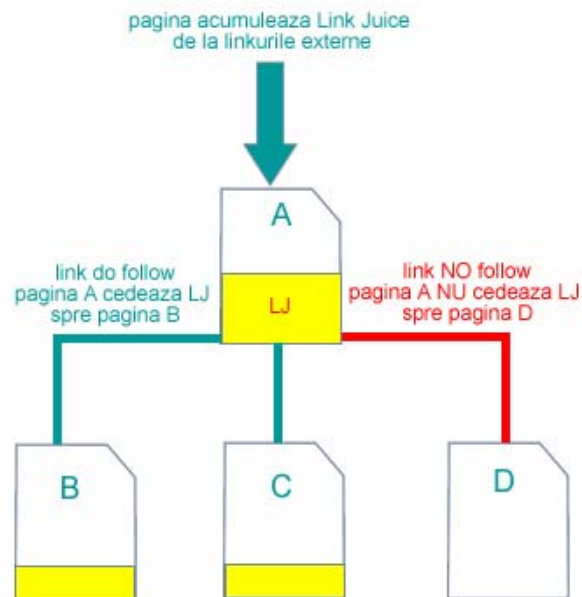


Probleme

- Un magazin online care se respectă are un număr considerabil de produse, probabil câteva sute.
- Prezentare - solutia cea mai folosită și corectă din punct de vedere al utilizabilității este **paginarea**.
 - Programatorii vor pune într-un loc vizibil un șir de linkuri de genul: pagina 1, pagina 2, pagina 30 etc.
- Prin creerea acestor pagini se diluează link juice –ul
 - motoarele de căutare nu vor “aprecia” documentele astfel obținute.
 - Din punctul lor de vedere *pagina 2* va fi la fel ca 3 sau 4 - fără prea mare valoare din punct de vedere al informației și de multe ori cu același titlu sau descriere.
- În al doilea rând: un e-shop cu sute de produse va avea **probleme cu indexarea** pentru că un spiderul are mult de “mers” până “citește” pagina nr. 30. Adio *flat architecture* și *indexare rapidă*.

Atributul *nofollow*

- `nofollow` –atribut html care “instruiește” motoarele de căutare să nu considere acel link ca un “vot” pentru pagina spre care trimit.



< a href="<http://www.site.ro>" rel="nofollow">Ancora Link< /a >



Linkuri ce pot fi publicate cu nofollow

- **Paginile de Register, Login sau Forgot Password**
- **Logo:**
 - Multe situri au un logo publicat ca imagine în header și linkuit către pagina principală
- **Categoriile, Rss , tagurile**
 - În general aceste pagini au parte numai de duplicate content
- **Contact**
 - dacă pagina are numai un formular de contact atunci nofollow este folositor
- **Copyright/TOS/Privacy/Feedback**



Linkuri ce pot fi publicate cu nofollow

- **Un alt caz în care folosirea tagului nofollow este recomandată:**
- Să presupunem că am scris un articol nou : “Voi și Programarea Web” postat pe www.site.com/voi-si-pweb.html. Vreau ca acest articol să apară pe prima pagină și atunci postez primele 2-3 paragrafe. Vom avea :

Titlu - link catre articol

Paragraf 1 - content

Paragraf 2 - content

Paragraf 3 - content

Citeste mai mult - link catre articol

- Linkul “citeste mai mult / read more” este unul care “produce pagube” (deși este absolut necesar) în cazul de față. El ajută la pierderea de link juice. Problema se rezolvă prin folosirea *nofollow*.



Bazele Proiectării “Search Friendly”



Concepte de bază

- **Page Content:** folosirea de conținut și text ce se adresează audienței și atragerea motoarelor de căutare și a legăturilor din alte site-uri
- **Navigation:** furnizarea utilizatorilor și crawlerelor acces facil la conținut
- **Design Considerations:** anularea oricăror artificii de design ce pot submina eforturile SEM
- **Page Rank:** popularitatea legăturilor



Gestionarea Conținutului Paginilor



Tipuri de Pagini

- Home Page
- About
- Contact
- Site Map
- News
- Forms
- Galleries
- FAQs
- Catalogs
- Product pages
- Shopping cart
- Search Results

Modul cum scrieți, proiectați și optimizați o pagină depinde de tipul acesteia



Text inclus

- Cuvinte cheie
 - Folosirea limbajului audienței
 - Folosirea de instrumente pentru alegerea cuvintelor cheie
 - Yahoo!/Overture: <http://inventory.overture.com/>
 - Google: <https://adwords.google.com/select/KeywordSandbox>
- Conținutul paginii
 - Conținutul trebuie să apară focusat
 - Title tag, heading-uri, contextual links, cross-links
 - Body text vizibil, i.e., nu ar trebui să impunem anumite acțiuni utilizatorului pentru ca acesta să vadă conținutul paginii principale



Text Primar vs. Secundar

Text Primar

- Title tag
- Body text
- Textul din partea de sus a paginii
- Textul din și din jurul legăturilor (e.g. anchor text)

Text Secundar

- Alt text
- Description tag
- Domain name și elementele URL



Conținutul ce trebuie inclus

- **Conținutul original**
 - Pros: Original, conținut unic
 - Cons: Time consuming, proces birocratic
- **Conținut preluat**
 - Pros: Modalitate lejeră de generare conținut
 - Cons: Cum să atragem valoare pentru a atrage legături?



Folosirea de *Syndicated Content**

- Tendința de actualizare frecventă
- Crawlerele vizitează paginile actualizate frecvent mai des
- Combinarea de “syndicated content” pentru crearea unei resurse unice
- Crearea propriului conținut exportat (e.g. RSS)
 - Îmbunătățirea vizibilității site-ului și atragerea de trafic
 - Oportunități pentru link-uri externe
- Aplicații care folosesc *syndicated content*: forumuri, programe web mail

*<http://www.rss-specifications.com/syndicated-content.htm>



Gestiunea Arhitecturii Site-ului



Schema de navigare în site

- Legături Text
 - Foarte “search engine friendly”
- Problemele cu Legăturile Text
 - Pot influența negativ densitatea de cuvinte cheie
 - Crawlerele au tendința de a citi legăturile text mai întâi
- Butoanele de Navigare
 - Folositoare (dar cu includerea de alt text)
 - Evitați JavaScript, exceptând cazul în care furnizați o formă de navigare pe care crawlerele o pot urma
 - Recomandare: folosiți alt text și text de navigare la sfârșitul paginii – vă permite plasarea de cuvinte cheie în mai multe locuri



Schema de navigare în site

- Image Map
 - Crawlerele ignoră legăturile din image map-uri
 - Adăugați legăruti text sau butoane de navigare suplimentare
- Meniuri Pull Down
 - În general nu sunt “crawler friendly” pentru că necesită JavaScript sau un program CGI
 - Întotdeauna furnizați două forme de navigare: una pentru utilizatori și una pentru crawlere



Ajutați navigarea crawlerelor

- Creați un site map
- Subscribe la Google Site Maps
 - Un instrument ce permite asistarea crawlerelor Google
 - Analizează informația despre arhitectura site-ului pentru îmbunătățirea căutării



Considerații de Design



Considerații legate de Design

- Artificiile de design (animațiile Flash, JavaScript) îmbunătățesc experiența utilizatorului dar influențează negativ vizibilitatea de căutare
- Implementați cu grijă site-ul pentru a ajuta și păstra rankingul acordat de motoarele de căutare
- Asigurați-vă că echipa de design înțelege conceptele SEM



Folosirea de JavaScript și CSS externe

- Folosind JavaScript pentru navigarea în site poate influența negativ posibilitatea crawlerelor de a naviga în cadrul site-ului
- Crawlerele nu urmează legături include în codul JavaScript; sau limitează tipurile de legături interne pe care le pot vizita
- JavaScript în <body> poate descrește timpul de încărcare a paginii
- Timpuri lungi de încărcare a paginii pot indica că respectivul site este spam și crawlerele pot ajunge să ignore



Folosirea de JavaScript și CSS externe

- Fișierele externe pot duce la o descreștere a timpului de încărcare pentru vizitatori
- Fișierele externe duc la îmbunătățirea timpului de descărcare și pentru crawlere
 - Rețineți că trebuie să dezactivați scripturile de crawling din robots.txt
- Scripturile externe sunt și mai ușor de refolosit



Cadre

- Nu se recomandă proiectarea site-ului din frame-uri
- Crawlerele au probleme în a parcurge structurile de cadre (frameset)
- Frameset-urile nu furnizează crawlerelor cuvinte cheie și legături
- Fiecare pagină este indexată separat
- Tag-ul `<no frames>` este ignorat



Frames Workarounds

1. Adăugarea de Navigation

- Dați tuturor paginilor titluri unice și tag-uri de descriere
- Plasați legături de navigare în cadrul paginilor

2. Adăugarea codului Java Script în tag-ul <head>

```
<script language="javascript">
```

```
<!--
```

```
if (top == self) self.location.href = "index.html";
```

```
// →
```

```
</script>
```

- Forțează browserul să încarce întotdeauna frameset-ul



Flash

- Puține motoare de căutare parcurg legăturile include într-o schemă de navigare Flash
- Site-urile Flash conțin puțin text
- Dacă totși includeți Flash ...
 - Includeți o legătură “Skip” a.î. atât utilizatorul cât și un crawler să poată să ajungă la homepage-ul real
 - Includeți metatag-uri pentru titlu și descriere



Pagini dinamice

- Scripturi având la bază baze de date, create ad-hoc pe baza asp, cfm, php, jsp sau cgi
- Site-urile dinamice sunt compuse din template-uri, de obicei fără conținut original
- Atunci când o pagină este vizualizată template-ul încarcă conținutul din baza de date
- URL-ul conține o serie de parametri ce indică template-ului ce conținut specific să încarce



Pagini dinamice

Exemplu:

<http://smithsonianstore.com/catalog/product.jsp?productId=14273&parentCategoryId=3151&categoryId=3152>

- URL-uri precum acesta sunt dificile de indexat pentru motoarele de căutare deoarece ele nu cunosc parametrii ce definesc o pagină unică
- Cu cât avem mai mulți parametrii cu atât mai slabe sunt șansele ca paginile respective să fie indexate



Pagini Dinamice “Search Friendly”

- Creați pagini HTML statice
- Modificați URL-urile a.î. să nu arate precum paginile dinamice, cu mai puțini parametri
- Folosiți modulul de rescriere a URL-urilor *mod_rewrite*



Optimizarea documentelor PDF

- PDF-urile trebuie să conțină chiar text, nu imagini a textului
- Aceleași reguli de folosire a cuvintelor cheie și a frazelor
- Puneți cel mai important text în titlu și headline
- Minimizați dimensiunea documentului (> 100K)
- Creați pagini HTML optimizate pentru PDF-uri



Gestiunea Page Rank



Page Rank

- Toate motoarele de căutare asignează o valoare site-ului pe baza legăturilor inbound
- Google numește acest factor de relevanță “*Page Rank*” (PR) – sinonim cu rank de popularitate
- Un link inbound reprezintă un vot pentru pagina voastră
- Un link outbound reprezintă un vot pentru pagina spre care aveți legătura



Page Rank

- O pagină primește un PR de îndată ce este indexată
- Aveți grijă ca toate paginile voastre să aibă un link înapoi către site
- Pagina ce primește cele mai multe legături inbound primește page rank-ul cel mai mare



Outbound Link Strategy

- Nu creați pagini având mai multe legături externe decât interne
- Creați legături spre site-uri de calitate, având subiecte înrudite – ajută la stabilirea site-ului ca unul de autoritate sau hub pentru alte site-uri
- Dacă o pagină conține legături externe, adăugați legături și spre paginile interne site-ului
- Nu vă fie frică să adăugați legături spre site-uri având un PR mai mic, dar conținut de calitate
 - Strategia bursieră – bet on small candidates for tomorrow's gold



Legături spre site

- Legături de la Yahoo! și DMOZ
 - Impactul asupra popularității poate fi diferit
 - Ajută la a face site-ul remarcat de crawlers
- Stabilirea de aranjamente de legături reciproce între site-uri cu agenții având domenii apropiate de activitate
- Făceți-vă publici în cadrul unor agenții sau profesioniști în domeniu
- Syndicate



Legături ce nu sunt legături

- Legături create într-o manieră pe care motoarele de căutare nu o pot înțelege
- Legăturile duc alt-undeva și nu direct la propriul site
- Exemplu: FirstGov.gov

<http://www.firstgov.gov/external/external.jsp?url=http://www.noaa.gov/>

- Legătura trece printr-un program JSP ce loghează click-urile și redirectează browser-ul către noaa.gov
- Motoarele de căutare citesc asta ca o legătură internă către FirstGov, nu o legătură externă către NOAA



Site Maintenance



Re-proiectarea site-ului

- Planificați **ÎNAINTE** de a face re-proiectarea
 - Se va modifica arhitectura site-ului?
 - Static sau Dinamic?
 - Cum se modifică conținutul?
- Adăugați ca activitate un review din partea SEO-ului a planului de re-proiectare
- Planificarea și execuția slabă pot conduce la scăderea rank-ului
- Nu așteptați până după ce terminați re-proiectarea!



Schimbarea arhitecturii site-ului

- Încercați să păstrați aceleași nume de fișiere și structură de directoare chiar și după re-proiectare
- Citiți înainte prezentarea MSN: “What to do when your site moves”:
http://search.msn.com/docs/siteowner.aspx?t=SEARCH_WEBMASTER_REF_Redirectcode.htm
- Recomandat: Setati redirectări HTTP 301 ce duc la noul site sau noile pagini



Ștergerea conținutului

- Evitați folosirea paginilor implicite 404
- Pagini 404 personalizate sunt mai *user-friendly*
- Submițați toate URL-urile 404 către motoarele de căutare folosind formularele lor speciale “Add URL” – cea mai rapidă cale de a scoate paginile 404 din index
- Nu uitați să schimbați codul de stare HTTP a paginilor custom de eroare de la 200 la 404



Resurse și Recomandări



Recomandări

- **Citiți Forumurile Industriale**

Webmaster World: <http://www.webmasterworld.com/>

Search Engine Watch: <http://www.searchenginewatch.com/>

Digital Point Forums: <http://forums.digitalpoint.com/>

High Rankings: <http://www.highrankings.com/forum/>

- **Dacă aveți timp, consultați chiar Evenimentele Industriale (lucrurile se schimbă încotinu :=**

Search Engine Strategies:

<http://www.searchenginestrategies.com/>



Bibliografie Recomandată

- Battelle, John. The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed our Culture. New York. Penguin, 2005.
- Kent, Peter. Search Engine Optimization for Dummies. Hoboken, NJ. Wiley, 2004.
- Moran, Mike, Hunt, Bill. Search Engine Marketing, Inc. Upper Saddle River, NJ. IBM Press, 2006.
- Thurow, Shari. Search Engine Visibility. Indianapolis, IN: New Riders, 2003.



Concluzii

- Un site optimizat pentru motoarele de căutare este ușor de găsit.
- Site-ul optimizat va răspunde întrebărilor vizitatorilor noi care se pot transforma în vizitatori fideli.
- O optimizare exagerată sau încercările de a 'păcăli' motoarele de căutare duc la penalizări
- Investițiile de bani în promovarea site-ului pot fi controlate ca buget și optimizate în funcție de rezultate preconizate