Content Management Systems (CMS)

Lecture 14: Search engines vs. Content Management (3)

Mariusz Trzaska, Ph. D.

mtrzaska@mtrzaska.com, *http:// www.mtrzaska.com*

Continuation of the previous lecture…

Domain for a site

- Start with a few keywords.
- Then check their combinations.
- Try it to be:
  - easy to remember and write,
  - short.
- Take care of domain uniqueness - names similar to existing domains are not recommended.
- Start searching with with domain * .com.

Domain for a site (2)

- international, e.g.
  - .pl,
  - .us,
  - .ai (Anguilla),
  - .dj (Djibouti),
  - .cd (Congo),
  - .cv (Cape Verde),
  - .fm (Micronesia),
  - .io (Indian Ocean),
- new domains, e.g.:
  - .academy,
  - .adult,
  - .app,
  - .art,
  - .attorney,
  - .auto, .car,
  - .bar,
  - .bike,
  - .biz,
  - .bot,
  - .cheap,
  - ...

https://en.wikipedia.org/wiki/List_of_Internet_top-level_domains

Domain for a site (3)

- Avoid numbers and "-".
- Be careful with current trends. Domain is chosen for years!
- Useful tools:
  - https://www.namemesh.com/
  - https://instantdomainsearch.com/
  - https://www.panabee.com/
  - https://namelix.com

Domain for a site (4)

- The top 20 most expensive domains sold

1	Voice.com	$	30 000 000,00
2	Sex.com	$	13 000 000,00
3	Tesla.com	$	11 000 000,00
4	Fund.com	$	9 999 950,00
5	Porn.com	$	9 500 000,00
6	Porno.com	$	8 888 888,00
7	HealthInsurance.com	$	8 133 000,00
8	We.com	$	8 000 000,00
9	Diamond.com	$	7 500 000,00
10	Z.com	$	6 784 000,00

11	Slots.com	$	5 500 000,00
12	Toys.com	$	5 100 000,00
13	Clothes.com	$	4 900 000,00
14	IG.com	$	4 700 000,00
15	HG.com	$	3 770 000,00
16	MI.com	$	3 600 000,00
17	Ice.com	$	3 500 000,00
18	Whisky.com	$	3 100 000,00
19	Sex.xxx	$	3 000 000,00
19	Vodka.com	$	3 000 000,00
19	California.com	$	3 000 000,00

Source: https://www.dnjournal.com/archive/domainsales/dnjournal-all-time-top-20-cash-domain-sales.htm

Domain for a site (5)

- The most expensive PL domains (up to 2017)

1.	co.pl	1 008 000,00 PLN	29.09.2011
2.	opony.pl	960 000,00 PLN	26.02.2010
3.	moneta.com	450 000,00 PLN	30.09.2013
4.	urlopy.pl	400 000,00 PLN	10.07.2008
5.	grydladzieci.pl	156 000,00 PLN	28.05.2013
6.	kino.pl	150 000,00 PLN	16.02.2009
7.	lekarze.pl	135 000,00 PLN	25.09.2009
8.	magazyny.pl	132 000,00 PLN	16.08.2010
9.	lista.pl	98 000,00 PLN	05.05.2011
10.	domain.pl	97 313,00 PLN	20.06.2008
11.	work.pl	95 000,00 PLN	01.06.2007
12.	makijaz.pl	92 500,00 PLN	06.2016
13.	l.pl	85 000,00 PLN	10.07.2012
14.	material.pl	80 000,00 PLN	22.07.2010
15.	sms.pl	73 800,00 PLN	12.02.2013
16.	deser.pl	73 792,00 PLN	30.06.2008
17.	ss.pl	72 105,00 PLN	25.09.2008
18.	baza.pl	72 000,00 PLN	30.04.2009
19.	adwokaci.pl	70 000,00 PLN	04.10.2008
20.	eventy.pl	70 000,00 PLN	01.07.2009
21.	CasinoOnline.pl	66 000,00 PLN	03.2016
22.	historia.pl	61 000,00 PLN	24.02.2009
23.	telefony.pl	59 600,00 PLN	20.04.2009
24.	choinki.pl	53 100,00 PLN	02.01.2013
25.	garnki.pl	52 800,00 PLN	16.05.2013
26.	lekarze.pl	52 188,00 PLN	17.11.2008
27.	lampa.pl	50 200,00 PLN	04.05.2012
28.	meble.pl	50 000,00 PLN	01.10.2003
29.	oczyszczalnie.pl	50 000,00 PLN	28.09.2009
30.	wedkarski.pl	50 000,00 PLN	25.10.2010

Source: https://sferadomen.pl/top-100-sprzedanych/

Domain for a site (6)

- Number of the PL domains in 2001 - 2019

	Number	Change
2001	134 721	51,44%
2002	159 509	18,40%
2003	195 842	22,78%
2004	276 687	41,28%
2005	443 249	60,20%
2006	557 091	25,68%
2007	762 656	36,90%
2008	1 323 454	73,53%
2009	1 629 929	23,16%
2010	1 994 120	22,34%
2011	2 278 722	14,27%
2012	2 402 107	5,41%
2013	2 461 499	2,47%
2014	2 524 538	2,56%
2015	2 681 752	6,23%
2016	2 703 165	0,80%
2017	2 576 063	-4,70%
2018	2 596 748	0,80%
2019	2 513 491	-3,21%

Source: https://danieldryzek.pl/2020/05/29/co-90-sekund-rejestrowana-jest-nowa-domena-pl/

Content management

Placement of keywords

- It seems that the title page is currently the most important: tag `title`.
  - Keywords at the beginning, the company name at the end.
  - Length (including spaces): Up to 65 characters.
  - Use longer phrases rather than single words.
  - Sometimes separators can be used: |, >, -, :

Placement of keywords(2)

- Page title – *continued*
  - If possible, the title should be an incentive to act, for example: "Buy a smartphone."
  - Ensure the integrity of the titles in the portal.
- Meta tag: `description`
  - Used by search engines on SERP pages;
  - Make sure it matches reality - do not exaggerate!
  - It's best to fit in 160 characters (Google) or 200 (Bing).

Placement of keywords(3)

- Meta tag: `description` - *continued*
  - o Treat it like an advertisement: brief and informative.
  - o Test different versions until you are satisfied with the result.
  - o Use matching keywords.
- Headline tags: `H1, H2, H3`
  - o `H1` as a single full page header. Do not use it to describe the logo / name of the portal.
  - o `H2, H3` are the section titles on the page. They can be found in several places.

Placement of keywords(4)

- Content of the document
    - Once the placement of keywords in the text was the basis of SEO.
    - Now it is not so important.
    - This is due to the immunization of search engines for numerous attempts to scam.
- Image file names and `alt` attribute.
- Tags `<b>` and `<strong>`. Some people think it's worth it (or rather will not cause any harm)

Keywords cannibalization

- It uses the same keywords across multiple pages and / or page titles.

- As a result, the search engine must "choose" which one (page) to include in the SERP.

- In other words, such a portal competes with itself.

- To avoid this, we need to ensure that the portal is properly and consistently implemented.

Duplicate content issue

- Spammers download content from various sources and (after minor modifications) publish them on their pages counting on:
  - o Attracting visitors,
  - o Getting links to their sites.
- As a result, the search engine must select which site to return to SERP.

Duplicate content issue (2)

- Sometimes this problem also occurs within the same portal:
  - A search engine indexing the same content repeatedly can waste too much time and will not process anything valuable;
  - Links to duplicate content may cause "fragmentation" of our content.
  - You can use robots.txt or meta tags to prevent this: `name=„robots" content=„noindex, follow"`.

Duplicate content issue(3)

- To detect duplicate content, you can use the tool: http://copyscape.com/

- In this situation, you can report a DMCA (Digital Millennium Copyright Act) violation and let the search engines know.

- Possibly take a civil trial.

Cloaking

- It consists of delivering two (or more) versions of the site depending on the recipient :
  - o Search engine robots,
  - o Users.
- Detecting is based on, for example, IP addresses and / or *UserAgent.*
- This technique can be used to cheat the search engines. In such a situation, you must be aware of a strict penalty.

Cloaking (2)

- You can also use it to improve search engine collaboration (but be careful)
  - Tests A/B for users and one easy (simple) version for a robot,
  - Content requiring registration,
  - Non-robotic navigation system (e.g. JavaScript)
  - Duplicate content
  - Content dependent e.g. on geographic location.

Choosing the right CMS

- Possibility to modify page title.
- Friendly URLs with keywords.
- Modify meta tags.
- Applying your own attributes / HTML tags.
- Change anchor descriptions.
- Flexible category system with descriptions.
- Paging control.
- Redirect 301.

Choosing the right CMS (2)

- RSS support.
- Image management (including `alt` attribute).
- Content cache.
- URL without parameters (e.g. session).
- Tagging (taxonomies).
- Breadcrumbs navigation.
- Sitemap generator (*sitemap.xml*).
- Validation of HTML5.

Multinational portals

- Domain for a specific country (pl., com.pl).
- Hosting in the country area.
- Physical (geographic) address on many pages of your site.
- Country setting in *Google Webmaster Central*.
- Targeted links from different national sites.
- Use of the national language.

Multinational portals (2)

- If your main page has a strong brand name, you can create national versions using:
  - Sub-domains, e.g. *pl.mycompany.com/*. Each subdomain (language) has its own value;
  - Virtual/logical directories like *mycompany.com/pl/*. The "value" of subdirectories (languages) depends on the general "value" of the domain.

Links

- Links to a site are one of the most important factors in determining quality.
  - Number,
  - Quality of the source site,
  - Descriptive text.

PageRank algorithm

- In simplification:
  - Each site has its own "quality";
  - When we link to the sites, we vote for them;
  - The „strength" of this voice depends on the quality of our site;
  - It is spread over all links ("votes"). So the more links, the smaller the unitary voice.
- Things get complicated when we include return links and any combination between sites.
- The PR algorithm is constantly changing.

Link value

- Descriptive text (*anchor*)
  - Used to determine what the target site is.
  - It can be very valuable when the indicated portal uses a technology that is not processed by search engines (e.g. JavaScript).

Link value (2)

- Descriptive text (*anchor*) – *continued*
  - *Google Bombing* – as a result, the site for the query containing the keywords in the description is very high in the SERP. In 2009 Google introduced the appropriate changes to the algorithm.
- Relevance. Links created under the same topic range are much more valuable.

Link value (3)

- Authority

  - *Hubs* – sites that have links to important sites;

  - *Authorities* - sites that are indicated by pages with a specific topic;

  - Initially introduced as Teoma in Ask.com. Currently implemented in most search engines.

Link value (4)

- Trust
  - Takes into account whether linking pages / domains are reliable (e.g. difficult to hack by spammers).
  - Trust in target websites.
  - We start with a bunch of manually selected sites and we check how far (*clicks*) the site is evaluating (*TrustRank*).

Link value (5)

- Sources.
  - Links within the same site are of course of lower value.
  - Similarly, there are many reciprocal links between several sites.
  - Also includes information on the ownership / dependency of *whois.*
- Diversity of sources - it is worth to have links from e.g.: blogs, forums, portals, social networking sites, etc.

Link value (6)

- Time aspect
  - When did the link appear?
  - When did it disappear?
  - How long was it visible?
  - How fast did new links appear?
- Proximity of other links on the same page.
- Location on the site (*prominence*).
- The text surrounding the link (including the header).
- The topic of the page content and the site.

Link value (7)

- TLD domain type
  - .edu, .gov, .mil, .com;
  - currently, it is not directly addressed.
- Independent estimation:
  - *PageRank* (absolute value) of the link's publishing page (or home page),
  - Website authority ("value" in the context of a specific topic),
  - PageRank, and Landing Page Authority,

Link value (8)

- Independent estimation – *continued*
  - o The number of outbound links on the source page.
  - o Relevance (similar topic) of the page and the source site.

Obtaining links

- The author found interesting content on our site
  - He/she wants his/her visitors to know about it,
  - How to propagate information about our content?
    - Sharing parts of the portal/article in social media, e.g. Facebook, Instagram, Twitter, LinkedIn, Pinterest, Reddit),
    - blogs,
    - ~~directories~~.
  - The most valuable from the search engines point of view.

Obtaining links (2)

- Emotional reaction of a publisher caused by e.g. funny drawing or controversial content.
  - *Link baiting,*
  - You have to decide whether it is consistent with the image of the organization.
- Business dependencies – e.g.
  - Sales network (put an appropriate record in the cooperation agreement)
  - Subcontractors.

Obtaining links (3)

- Direct invitation / request to create a link to our portal
  - Directed to the site (personalized), which may be interested in our topic,
  - Filling in a contact form or sending an email,
  - Give specific reasons / benefits from creating such a link.
- Create a link on social media, such as Linked in, Facebook, Instagram, wikipedia, Twitter.

Obtaining links(2(4)

- Exchange between sites
  - *Reciprocal linking*.
  - Limited value for search engines.
- Paid links
  - From the point of view of search engines they are worthless.
  - In some situations such behavior can be punished with lower scores in the SERP.
- Traditional communication, e.g. an offer of organizing (free) seminar (which will be advertised on the site - link).

Obtaining links – doubtful ways

- Using them can result in search engine penalties (until removed from the index / SERP).
- Buying links
  - Easy,
  - Google suggests that ads use `NoFollow` (no value for search engines)

Obtaining links – doubtful ways (2)

- Buying links - *continued*
  - o Possibilities
    - Direct shopping,
    - Link brokers
    - Grants to organizations (in return for sponsorship links).
- Link farms
  - o Viewed as spam,
  - o (Possibly) effective in the short term,

Obtaining links – doubtful ways (3)

- Bots creating links
  - Publishing entries in:
    - comments,
    - discussion forums.
  - Prevents with captcha, registration, etc.
  - Link effects can be leveled using `NoFollow`.
- Widgets in JavaScript that contain a link in the `NoScript` tag (only this part is processed by the search engines).

Link strategy

- Potential candidates for linking to our site:
  - Uncompetitive sites from the same market (e.g. industry information)
  - Media,
  - Bloggers,
  - Universities,
  - Government websites,
  - Hobbyists,
  - Parties that have links to our competition (success if we achieve 10% effectiveness).

Link strategy (2)

- Where do our competitors take links from?
  - o Tools,
  - o Note the most important ones (high PageRank, mozRank / mozTrust)
- Evaluation of own resources
  - o Which are worth promoting,
  - o What new can we create?
- Endless process.

Link strategy (3)

- Comparison of the effectiveness of different approaches.
    - Creation of new content costs X PLN,
    - Maybe it will be cheaper to buy ads, for example?

Google Ads – the most expensive keywords (payment per click):
1. Business Services: $58.64
2. Bail Bonds: $58.48
3. Casino: $55.48
4. Lawyer: $54.86
5. Asset Management: $49.86
6. Insurance: $48.41
7. Cash Services & Payday Loans: $48.18
8. Cleanup & Restoration Services: $47.61
9. Degree: $47.36
10. Medical Coding Services: $46.84
   Source: *https://www.socialmediatoday.com/news/these-are-the-10-most-expensive-google-keywords/512893/*

Tracking SEO effects

- SEO is a long-term process that is not always based on clear premises.
- Determining the initial conditions (so that we can compare)
  - o Fixing keywords for attracting more visitors,
  - o Division of traffic resulting from organic results based on keywords,
  - o The most popular site sections / content,

Tracking SEO effects (2)

- Establishment of initial conditions - *continued*
  - Not very popular content,
  - Activity of searching robots,
  - Number of indexed pages.
- Measures of success
  - Improvement of sales results,
  - Increased Ad visibility

Tracking SEO effects (3)

- Measures of success
  - Increasing number of
    - subscribers to RSS, newsletter,
    - registrations,
    - issues,
    - contact attempts,
    - links to our site.
  - Improve website traffic
  - Acquisition of grants.

Tracking SEO effects (4)

- Tools
  - Analytical made available by the hosting company (log analysis)
  - [Google Analytics](#),
  - [https://www.bing.com/toolbox/webmaster](#)
  - [https://www.adobe.com/experience-cloud.html](#) (paid)
  - [http://www.unica.com](#) (paid),
  - [http://www.webtrends.com/](#) (paid),
  - …

Tracking SEO effects (5)

- More important information:
  - Traffic according to the search engine,
  - Traffic based on keywords,
  - Source pages (*referring sites*)
  - Top sites / content,
- Common mistakes - making decisions based on:
  - Too small sample of data,
  - Data from unrepresentative time period (eg December in online store).

Tracking SEO effects (6)

- Branding. If most of the traffic from search engines comes from queries about our brand then we probably have a problem
  - Do not use general keywords,
  - No queries about the so-called long tail (important!).

Valuable resources for SEO

- [Search Engine Land](#),
- [Search Engine Watch](#),
- [SEOmoz](#),
- [Web Site Magazine](#),
- [Google Webmaster Central Blog](#),
- [Microsoft Bing Blogs](#),
- [Google Webmaster Help](#),
- [Webmaster World](#),
- [Search Engine Roundtable](#).

Glossary

- Cloaking (in SEO) - sharing different versions of your site depending on the recipient.
- PPC (Pay Per Click) - sponsored results, ads.
- ROI - Return on Investment
- SERP - Search Engine Result Page.
- TLD - Top Level Domain, e.g. com.

Summary

- SEO is a very important aspect of contemporary content management.

- Very often, success of our project depends on the SEO.

- This is an endless process that requires extensive interdisciplinary knowledge.

- It may use spamming techniques although it is usually unprofitable in the long run.