slido

Audience Q&A Session

Crawling Budget

Gennadiy Vorobyov

Founder & CEO at Netpeak Bulgaria

SoftUni Digital

SoftUni Digital

https://digital.softuni.bg

Sli.do Code

#SEO01

Gennadiy Vorobyov aka Genrih

SoftUni Digital

- Founder& CEO @ **Netpeak Bulgaria**

- Author @https://t.me/b4business

- Happy manager with 40+ happy netpeakers and 120+ active projects

- Subscribe to **Netpeak Bulgaria** Telegram channel- https://t.me/netpeakbg

- Providing professional performance digital marketing solutions for the last 6 years.

- Author & lecturer of Softuni Digital courses in 2021:

  - SEO Fundamentals, march 2021 + september 2021

  - Google Analytics Fundamentals, april 2021

  - CRO & Automation, june 2021

  - Google Analytics Advance, july 2021

  - SEO Advance, october 2021

Table of Content

1. Toolbar pagerank history

2. Internal pagerank calculation

3. Interlinking: methodology, examples & tools

4. Crawling budget: definitions, optimization

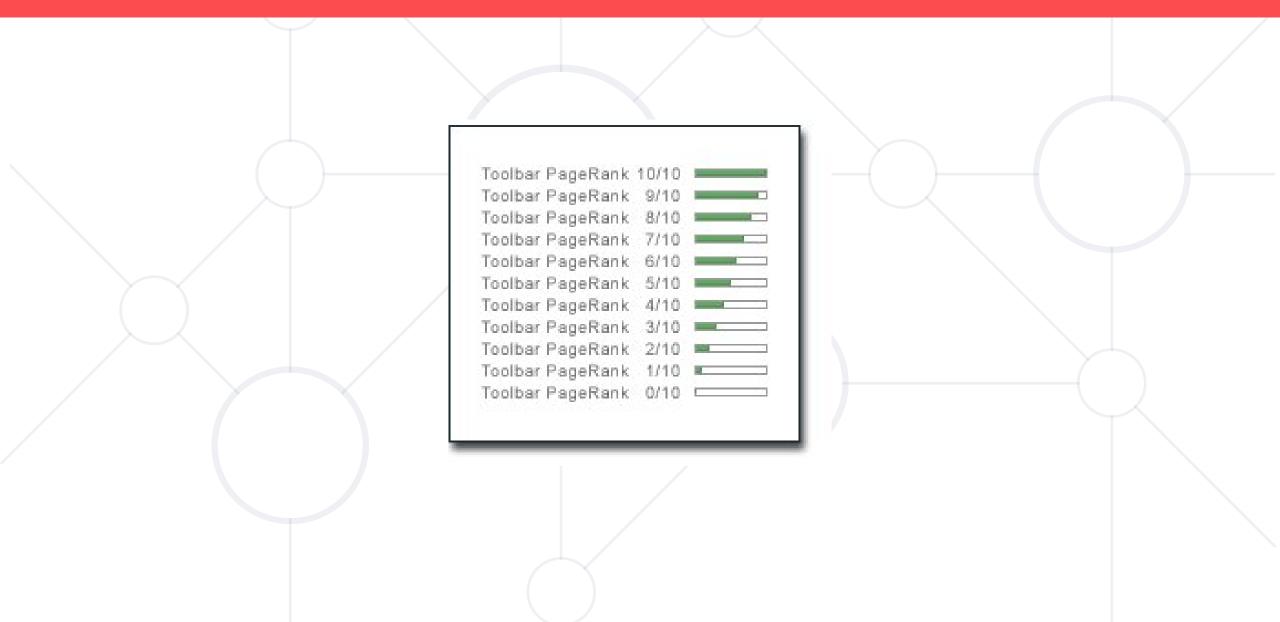5. Related pages/products by category, price

6. Netpeak Cloud demo

Toolbar Pagerank

Toolbar Pagerank

Toolbar Pagerank

The Google Toolbar long had a PageRank feature which displayed a visited page's PageRank as a whole number between 0 and 10.

The most popular websites displayed a PageRank of 10.

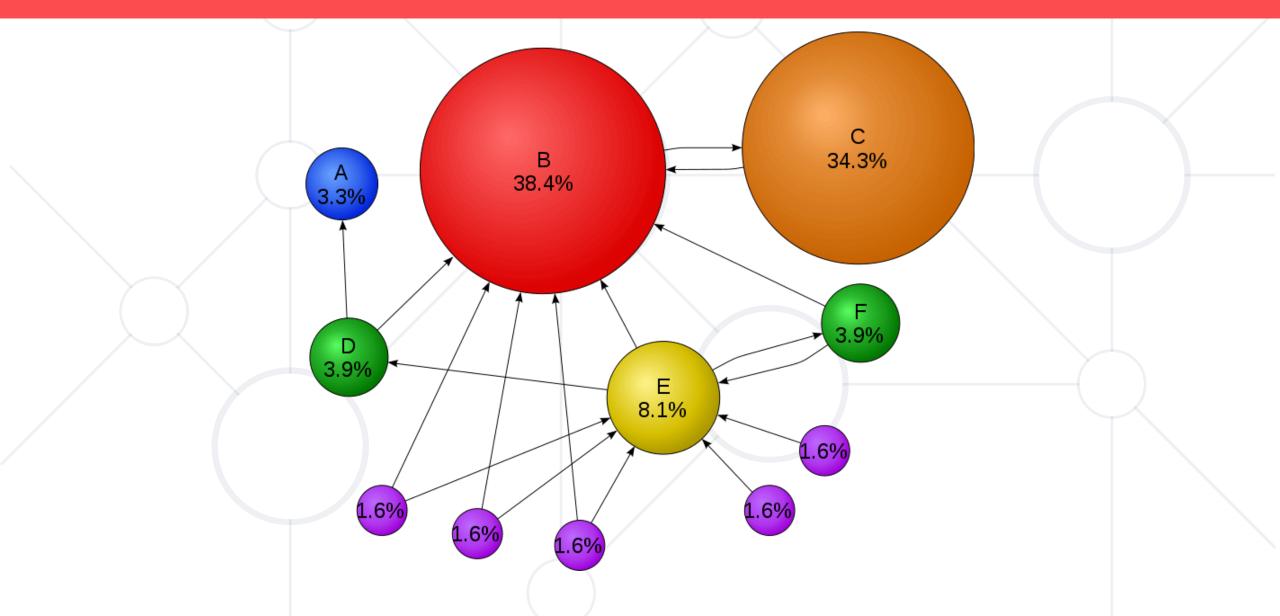The least showed a PageRank of 0.

Google has not disclosed the specific method for determining a Toolbar PageRank value, which is to be considered only a rough indication of the value of a website.

In March 2016 Google announced it would no longer support this feature, and the underlying API would soon cease to operate.

https://en.wikipedia.org/wiki/PageRank#Google_Toolbar

PageRank

SoftUni Digital

A
3.3%

B
38.4%

C
34.3%

D
3.9%

E
8.1%

F
3.9%

1.6%

1.6%

1.6%

1.6%

1.6%

1.6%

PageRank

PageRank (PR) is an algorithm used by Google Search to rank web pages in their search engine results. PageRank is a way of measuring the importance of website pages. According to Google:
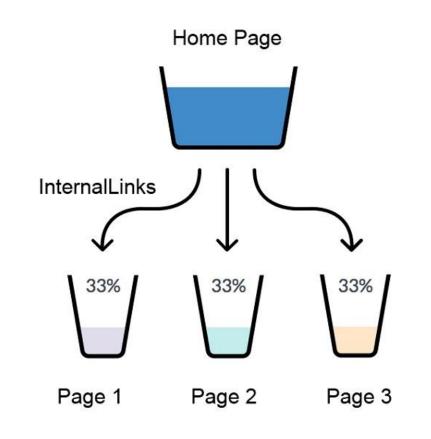
> PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.

https://en.wikipedia.org/wiki/PageRank

Internal PageRank

The internal PageRank determines the importance of the page based on internal links

Formula

$$PR (A) = (1 - d) / N + d * (PR(B) / L(B) + PR(C) / L(C) + ...)$$

- N – number of pages in the calculation;
- d – dumping coefficien (usually 0.85)
- L – number of outgoing links of specific pages

The PageRank algorithm outputs a **probability distribution** used to represent the likelihood that a person **randomly clicking on links will arrive at any particular page**.

Calculations
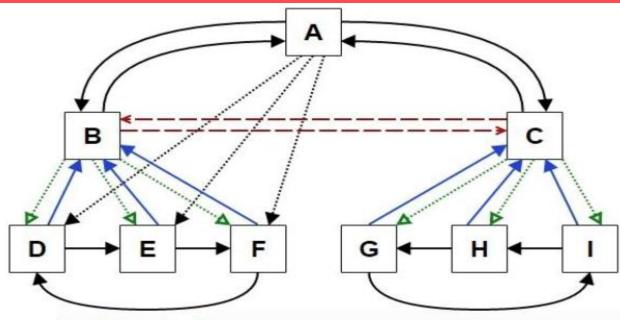
- with crawlers like Netpeak Spider, Screaming frog etc

- Manual – work sheet

$$PR\ (A) = (1 - d)\ /\ N + d * (PR(B)\ /\ L(B) + PR(C)\ /\ L(C) + ...)$$

- N – number of pages in the calculation;
- d – dumping coefficien (usually 0.85)
- L – number of outgoing links of specific pages

N = 9
d = 0.85
L - индивидуално
за всяка страница

L	Pages		Iterations										
			0	1	2	3	4	5	6	7	8	9	10
5	A	=	0,111	0,054	0,083	0,077	0,078	0,078	0,078	0,078	0,078	0,078	0,078
5	B	=	0,111	0,196	0,189	0,190	0,195	0,195	0,196	0,197	0,197	0,198	0,198
5	C	=	0,111	0,196	0,165	0,172	0,165	0,165	0,163	0,163	0,162	0,162	0,162
2	D	=	0,111	0,102	0,102	0,106	0,107	0,109	0,109	0,110	0,110	0,110	0,110
2	E	=	0,111	0,102	0,102	0,106	0,107	0,109	0,109	0,110	0,110	0,110	0,110
2	F	=	0,111	0,102	0,102	0,106	0,107	0,109	0,109	0,110	0,110	0,110	0,110
2	G	=	0,111	0,083	0,085	0,081	0,080	0,079	0,078	0,078	0,077	0,077	0,077
2	H	=	0,111	0,083	0,085	0,081	0,080	0,079	0,078	0,078	0,077	0,077	0,077
2	I	=	0,111	0,083	0,085	0,081	0,080	0,079	0,078	0,078	0,077	0,077	0,077
Total PageRank			1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

NETPEAK

Why to calculate PageRank

1. To understand how internal PageRank is distributed.

1. To find "over-pageranked" landing pages.

1. To find pagerank leaks.

3 main issues

- **PageRank: Dead End** → These pages doesn't have any internal link.

- **PageRank: Redirect** → Pages with redirect or canonical to other page.

- **PageRank: Orphan** → Pages with no any incoming link.

How to calculate internal PageRank with Netpeak Spider

SoftUni Digital

Action!

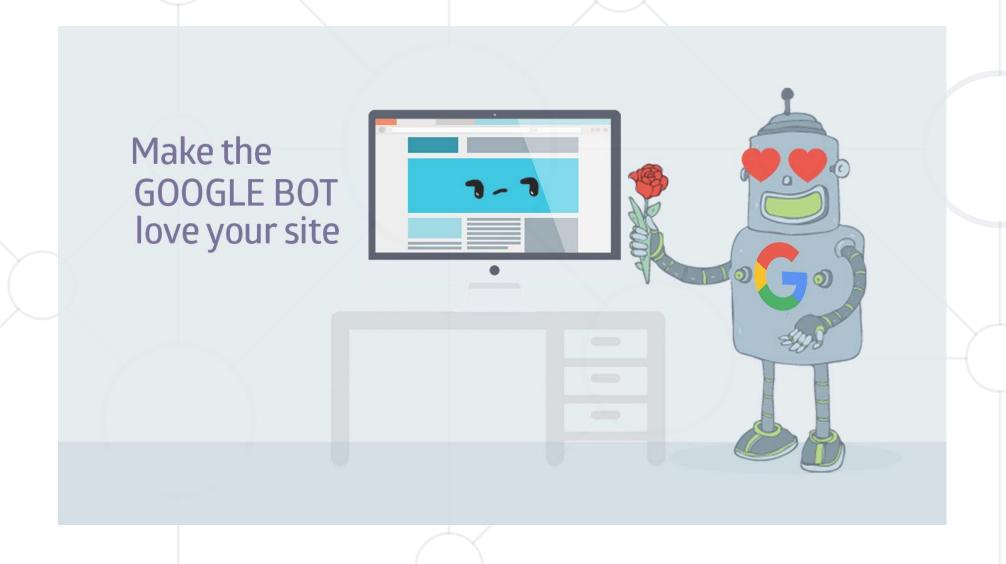Netpeak Spider

Type of pages in NS

3 main statuses

- **Compliant -** HTML file with 2xx http header
- **Non-compliant** - HTML file with http header different than 2xx, noindexed files.
- **Non-html** – other documents with low chanches to be ranked (excl. images).

Make the GOOGLE BOT love your site

What does Google says:

"Определянето на приоритет на това какво да обходим, кога и колко ресурс сървърът, който хоства сайта, може да разпредели за обхождане е по-важно за по-големите сайтове или тези, които автоматично генерират страници въз основа на URL параметри" - Gary Illyes

Google:

Useful: What Crawl Budget Means for Googlebot

What does SEO Specialists say:

Generally

- how many pages Googlebot may crawl (per day, week, month);

- Crawling budget correlate with internal and external PageRank of the main page.

Correlations

- organic traffic reflects of crawling budget;
- crawling budget reflects on number of crawled pages;
- pagespeed reflects of crawling budget
- new & updated content reflects

Statistics

SoftUni Digital

https://search.google.com/search-console/not-verified?original_url=/search-console/settings/crawl-stats&original_resource_id

Statistics

SoftUni Digital

Статистически данни за обхождането

Дейност на Googlebot за последните 90 дни

Обходени страници на ден	Високо	Средно	Ниско
	324 456	122 439	5 640

Изтеглени килобайта на ден	Високо	Средно	Ниско
	94 594 866	35 700 892	680 273

Време за изтегляне на страница (в милисекунди)	Високо	Средно	Ниско
	5 736	2 045	588
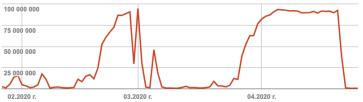
Crawl rate control in GSC

SoftUni Digital

You can control crawl rate

https://www.google.com/webmasters/tools/settings

3 main reasons for low crawling budget:

- Duplicated content
- Internal 404/410
- Internal redirects

Duplicated content

1. Main redirects
2. Pagination
3. Sorting and filtering options
4. Main version: www/non-www
5. Protocols: https/http
6. Mobile version

Internal 404

- Crawling with Spiders
- GSC;
- Ahrefs.

Internal 301 redirects

- Crawling with Spiders

- Can be generated when migrate from http to https

- Wrong addresses.

Стар адрес	301 редирект към:
http://www.sait.bg/Include/Documents/newspaper/easytimes_02_2010.pdf	https://www.sait.bg/novini/vestnik-izi-tayms-yuni-2018
http://www.sait.bg/include/Documents/newspaper/easytimes_09_2010.pdf	https://www.sait.bg/novini/vestnik-izi-tayms-yuni-2018
http://www.sait.bg/Include/Documents/newspaper/easytimes_07_2008.pdf	https://www.sait.bg/novini/vestnik-izi-tayms-yuni-2018

SoftUni Digital
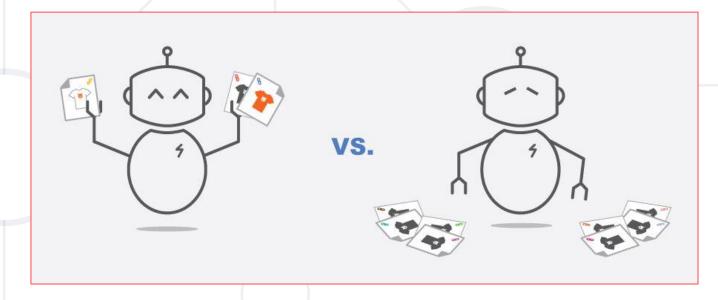
Crawling budget optimization

2 ways to control

- Natural;

- Semi-natural

Natural ways

SoftUni Digital

We can control it building:

- Related products

- Up-sale products

- Related articles

- Breadcrumbs

- Other way for interlinking and internal pagerank distribution

- HTML sitemap also can help.

Semi-natural ways

Noindexing:

- robots.txt, but Googlebot make a request to load disallowed page

- **Meta tag noindex** - Google bots check HTML-header and leave no indexed page.

* nofollow doesn't an option, because Googlebot check the content.

- **X-robots-tag:** Googlebot check HTTP headers only. But, you should be very careful

```
HTTP/1.1 200 OK
Date: Tue, 25 May 2010 21:42:43 GMT
(...)
X-Robots-Tag: noindex
(...)
```

The X-Robots-Tag in the Header Response

```
HTTP/1.1 200 OK
Date: Tue, 25 May 2010 21:42:43 GMT
(...)
X-Robots-Tag: googlebot: nofollow
X-Robots-Tag: otherbot: noindex, nofollow
(...)
```

Example:

www.laptopbg.net - Превод на страницата

Няма налична информация за тази страница.

Научете защо

За да ви покажем най-подходящите резултати, пропуснахме някои, които
са много подобни на вече изведените 1.
Ако искате, можете да повторите търсенето заедно с пропуснатите
резултати.

You can confuse the crawler using Disallow + Noindex

- **SEO Hide**
- **Clever Hide**

SEO Hide

1. ● using the online decoder Base64 we encode the sourcecode of the item we want to hide;
2. paste the code in content.js, and connect each new element to its new attribute hashstring;
3. include jquery.js, base64.js & content.js in <head> of html;
   a. Your link should be look like
4. Disallow /js/ in Robots.txt.
5. Disallow /js/ in Robots.txt.

SEO Hide Technical task - HERE

SEO Hide

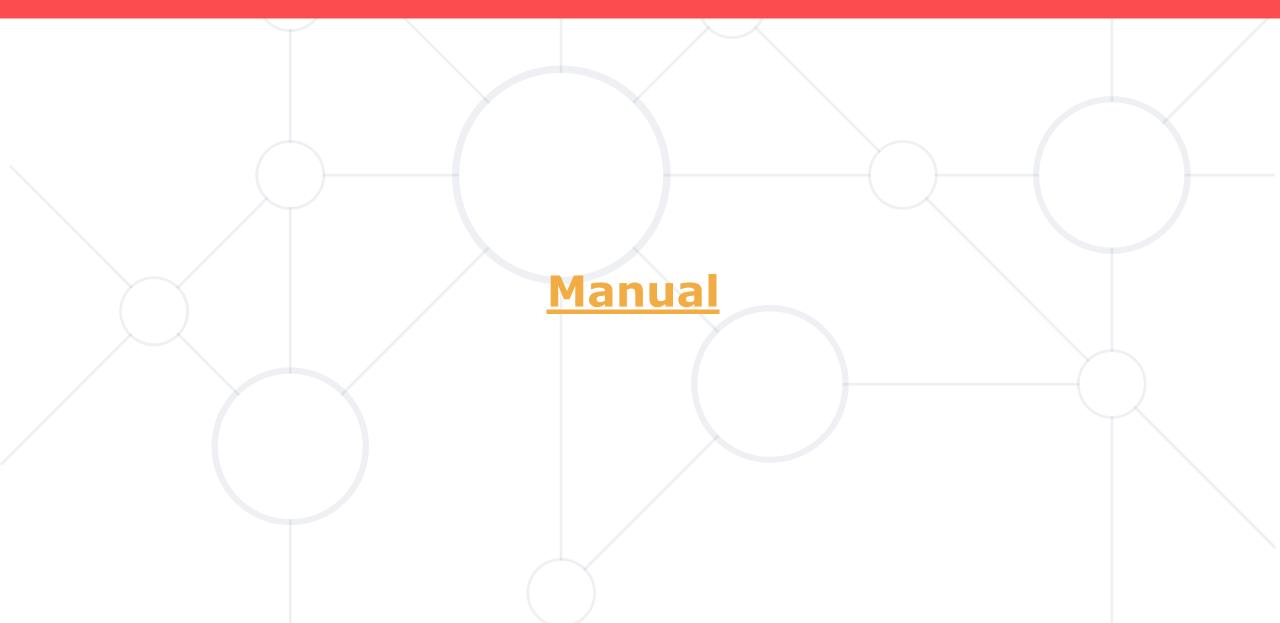- не трябва да затваряме през robots.txt целия JavaScript на сайта;

- връзките са активни и функционални за потребителите;

*Бонус
Линк за изтегляне на библиотеките на SEO Hide: ЦЪК
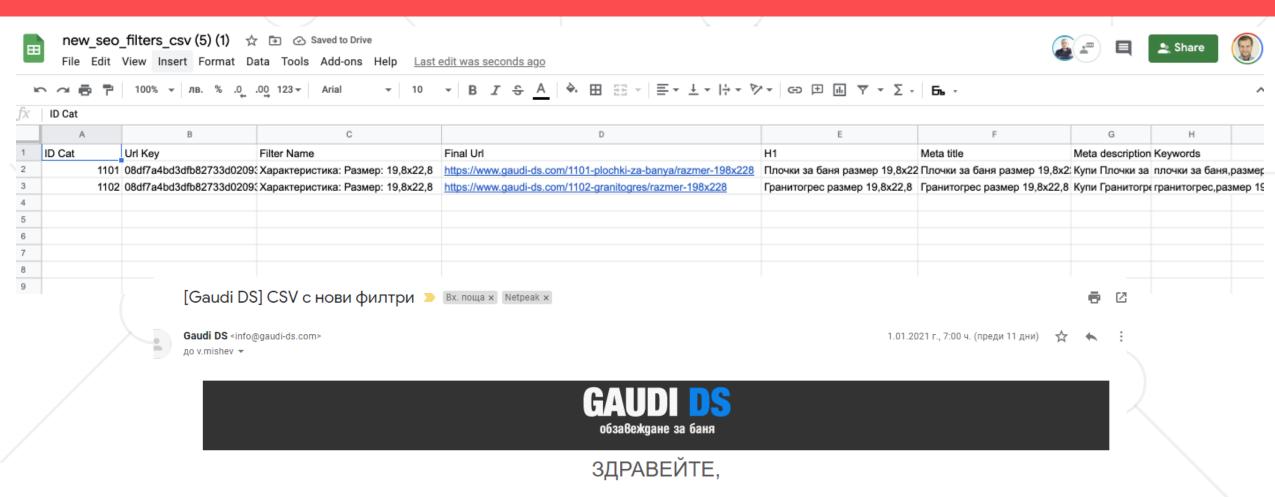Пример тук

Clever hide

Manual

Examples, Netpeak Cloud

1. gaudi-ds.com
2. 4sales.bg
3. medina-med.com
4. teodor.bg
5. ciela.com – Троцки kap_trotsky
6. superimoti.bg – Троцки

new_seo_filters_csv (5) (1) ☆ Saved to Drive

File Edit View Insert Format Data Tools Add-ons Help   Last edit was seconds ago

Share

100% лв. % .0 .00 123 Arial 10 B I S A

ID Cat

	A	B	C	D	E	F	G	H
1	ID Cat	Url Key	Filter Name	Final Url	H1	Meta title	Meta description	Keywords
2	1101	08df7a4bd3dfb82733d02093	Характеристика: Размер: 19,8х22,8	https://www.gaudi-ds.com/1101-plochki-za-banya/razmer-198x228	Плочки за баня размер 19,8х22	Плочки за баня размер 19,8х2	Купи Плочки за	плочки за баня,размер
3	1102	08df7a4bd3dfb82733d02093	Характеристика: Размер: 19,8х22,8	https://www.gaudi-ds.com/1102-granitogres/razmer-198x228	Гранитогрес размер 19,8х22,8	Гранитогрес размер 19,8х22,8	Купи Гранитогре	гранитогрес,размер 19
4								
5								
6								
7								
8								
9								

[Gaudi DS] CSV с нови филтри ⯈ Вх. поща ✕ Netpeak ✕

Gaudi DS <info@gaudi-ds.com>   1.01.2021 г., 7:00 ч. (преди 11 дни) ☆ ↩ ⋮
до v.mishev ▾

GAUDI DS
обзавеждане за баня

ЗДРАВЕЙТЕ,

ФАЙЛЪТ С НОВИТЕ ФИЛТРИ ЗА 01.2021 Е ГЕНЕРИРАН.

Използвайте следния линк, за да го свалите: Нови филтри в Gaudi DS.

Gaudi DS се управлява с PrestaShop™

Resume

1. Information about Pagerank
2. Information about Crawling budget.
3. Ways to control crawling budget.
4. Interlinking examples.

Homework

1. Using Netpeak Spider, calculate an internal PageRank with 20 iterations of up to 100 pages of the website you analyze. As a result, make an export of PageRank amount changes table.
    a. If using another software, make a common export of Pagerank calculations.
2. Make a Doc with ideas for interlinking of the website you analyze.

Sli.do Code

#SEO01

Organization Partners

SoftUni Digital

VASSILENA VALCHANOVA

IMPULSE MEDIA

createX

BOOKMARK

NETPEAK
Performance Marketing for Business

ABC DESIGN & COMMUNICATION

1 FOR FIT

TIEN
PR & Digital

StartupMasters.

Aleikov
studio

License

- This course (slides, examples, demos, exercises, homework, documents, videos and other assets) is **copyrighted content**

- Unauthorized copy, reproduction or use is illegal

- © SoftUni – https://about.softuni.bg

- © SoftUni Creative – https://digital.softuni.bg

About SoftUni Digital

- SoftUni Digital – High-Quality Education, Profession and Job for Designers

    - digital.softuni.bg

- SoftUni Digital @ Facebook

    - https://facebook.com/SoftUniDigital

- SoftUni Digital Forums

    - softuni.bg/forum/categories/116/softuni-creative