

Příjmení: Vacho	Jméno: Peter	Obor: Softwarové Inženýrství
Předmět: <b>Softcomputing a Datamining</b>	Datum: 24.11.2024	
Název protokolu: PageRank algoritmus		

## Scraper

Ako prvé som vytvoril scraper funkciu, ktorá získava obsah danej stránky a hľadá v jej HTML obsahu všetky anchor (<a>) prvky, z ktorých následne extrahuje href parameter, obsahujúci odkaz na tejto stránke. Všetky takéto nájdené odkazy sú potom vrátené v set objekte (pre zabránenie duplicitných linkov). V tejto funkcii som taktiež zabezpečil správne fungovanie pre relatívne adresy v href parametri (naúr.: /ahoj namiesto <https://example.com/ahoj>), ktoré sú konvertované na absolútne adresy na základe začiatkovej URL.

Druhou časťou bolo vytvorenie ďalšej funkcie, ktorá využíva tú predošlú, no rekurzívne prechádza nájdené stránky až do určitej hĺbky. V tejto funkcii som taktiež pridal logiku pre obmedzenie prehľadávania stránok ktoré nespĺňajú určitú filtrovaciu podmienku (vo funkcii toto funguje obecné, s pomocou predicate filtrovacej funkcie ktorú funkcia obdrží ako argument a použije pre overenie, či sa táto URL má prehľadávať alebo nie). Taktiež som tu pridal možnosť pre jednoduché potlačovanie výnimiek, ktoré pri scrapovaní môžu nastať (ako napríklad stránka, ktorá vracia kód HTTP 404).

Výstupom tejto funkcie je hash mapa, ktorá mapuje URL k listu nájdených linkov na tejto URL.

## Ukážka fungovania scraperu

```
itsdrake ~/Current/Datamining/tasks $ python -m snc
Scraping...
> Cache not found, proceeding to scrape
> Scraping from (depth=0): https://aialab.fai.utb.cz
> Scraping from (depth=1): https://aialab.fai.utb.cz/contact/
> Scraping from (depth=1): https://aialab.fai.utb.cz/projects/
> Scraping from (depth=1): https://aialab.fai.utb.cz/
> Scraping from (depth=1): https://aialab.fai.utb.cz/resources/
> Scraping from (depth=1): https://aialab.fai.utb.cz/first-competition-on-llm-designed-ea/
> Scraping from (depth=1): https://aialab.fai.utb.cz/wcci24specialsession/
> Scraping from (depth=1): https://fai.utb.cz/en/
> Scraping from (depth=1): https://aialab.fai.utb.cz/thesis-topics/
> Scraping from (depth=1): https://aialab.fai.utb.cz/#content
> Scraping from (depth=1): https://aialab.fai.utb.cz/bibliography/
> Scraping from (depth=1): https://aialab.fai.utb.cz/about-us/
> Scraping from (depth=1): https://aialab.fai.utb.cz/our-team/
> Scraping from (depth=1): https://fai.utb.cz/en/faculty/about-us/structure/departments/the-department-of-informatics-and-artificial-intelligence/about-the-department/
> Scraping from (depth=1): https://www.utb.cz/en/
> Scraping from (depth=1): https://aialab.fai.utb.cz/publicity/
> Scraping from (depth=2): https://aialab.fai.utb.cz/contact/#content
> Scraping from (depth=2): https://aialab.fai.utb.cz/projects/#content
> Scraping from (depth=2): https://aialab.fai.utb.cz/#content
> Scraping from (depth=2): https://aialab.fai.utb.cz/wp-content/uploads/2021/10/SSCI_2021_TABLES.pdf
--> Got Non-HTML Content-Type Header, (application/pdf) from https://aialab.fai.utb.cz/wp-content/uploads/2021/10/SSCI_2021_TABLES.pdf
> Scraping from (depth=2): https://aialab.fai.utb.cz/wp-content/uploads/2021/04/GECCO_2021_Workshop_SWINGA_TABLES.pdf
--> Got Non-HTML Content-Type Header, (application/pdf) from https://aialab.fai.utb.cz/wp-content/uploads/2021/04/GECCO_2021_Workshop_SWINGA_TABLES.pdf
> Scraping from (depth=2): https://aialab.fai.utb.cz/wp-content/uploads/2021/02/SEMCCO_full_data.pdf
--> Got Non-HTML Content-Type Header, (application/pdf) from https://aialab.fai.utb.cz/wp-content/uploads/2021/02/SEMCCO_full_data.pdf
> Scraping from (depth=2): https://aialab.fai.utb.cz/resources/#content
> Scraping from (depth=2): https://aialab.fai.utb.cz/wp-content/uploads/2021/08/AMCS_Figures_Repository.zip
--> Got Non-HTML Content-Type Header, (application/zip) from https://aialab.fai.utb.cz/wp-content/uploads/2021/08/AMCS_Figures_Repository.zip
> Scraping from (depth=2): https://aialab.fai.utb.cz/wp-content/uploads/2021/05/GECCO_2021_Workshop_competition_TABLES_2.pdf
--> Got Non-HTML Content-Type Header, (application/pdf) from https://aialab.fai.utb.cz/wp-content/uploads/2021/05/GECCO_2021_Workshop_competition_TABLES_2.pdf
> Scraping from (depth=2): https://aialab.fai.utb.cz/de_dynamic_switch-py/
> Scraping from (depth=2): https://aialab.fai.utb.cz/first-competition-on-llm-designed-ea/#content
> Scraping from (depth=2): https://aialab.fai.utb.cz/wcci24specialsession/#content
> Scraping from (depth=2): https://fai.utb.cz/en/faculty/international/

--> Skipping redirect (301) from https://www.utb.cz/en/research-and-development/scientific-board/
> Scraping from (depth=2): https://www.utb.cz/en/projects/tbu-for-action-on-the-climate/
> Scraping from (depth=2): https://www.utb.cz/en/news-events/tbu-gala-evening-celebrated-the-international-students-day-and-the-velvet-revolution/
> Scraping from (depth=2): https://www.utb.cz/en/graduate/opportunities-for-cooperation/
> Scraping from (depth=2): https://www.utb.cz/en/news-events/tbu-gala-evening-to-mark-the-35th-anniversary-of-the-velvet-revolution/
> Scraping from (depth=2): https://www.utb.cz/en/cooperation/services/facility-hire/
--> Skipping redirect (301) from https://www.utb.cz/en/cooperation/services/facility-hire/
> Scraping from (depth=2): https://aialab.fai.utb.cz/publicity/#content
> Storing results to cache
Took: 294.5 seconds
```

## PageRank

Implementoval som funkciu pre algoritmus PageRank, ktorá prijíma hash mapu, kde každá stránka (URL) mapuje na množinu stránok, na ktoré odkazuje. Algoritmus spočíva v klasickom výpočte PageRanku, kde každá stránka začína s rovnakým skóre.

Najprv sa vytvorí matica susednosti, ktorá určuje pravdepodobnosť prechodu medzi stránkami na základe počtu odkazov. Ak stránka nemá výstupy, rovnomerne rozdelí pravdepodobnosť medzi všetky stránky. Následne sa matica upraví o "skoky", čo zabezpečuje stabilitu a zabráňuje zamrznutiu na izolovaných stránkach.

Iteratívne sa aktualizujú hodnoty PageRanku pre každú stránku, pričom proces pokračuje, až kým rozdiel medzi novými a starými hodnotami neprekročí toleranciu alebo neuplynie maximálny počet iterácií. (Použitie tolerancie je v pagerank algoritmu časté, pridal som ho aj keď to zadanie nevyžadovalo, vždy je možné nastaviť toleranciu na 0, pri čom by prakticky nikdy nenastali rovnaké výsledky.)

Výsledkom je zoznam stránok s ich hodnotami PageRanku, ktorý odráža ich relatívnu dôležitosť v sieti. Algoritmus využíva matice a lineárnu algebru a zavádza rovnováhu medzi odkazmi a náhodnými skokmi pomocou parametra damping factor (beta).

## Overenie fungovania

V zadaní boli uvedené aj testovacie dáta a očakávané výsledky z pagerank algoritmu pri týchto vstupoch. Vytvoril som tak testovaciu funkciu, ktorá pagerank funkciu spúšťa práve na týchto testovných dátach. Keďže mi prišlo nadbytočné zaraďovať celý unit-test systém do projektu len pre tento jeden test, funkciu som jednoducho pri testovaní spustil manuálne. Funkcia testom prechádza úspešne, a pri implementácii som nenarazil na žiadne závažné problémy.

```
In [1]: from src.pagerank import test_pagerank
In [2]: test_pagerank()
In [3]:
```

Prázdný výstup značí úspěch (zhodu s ukázkovými datami)

```
def test_pagerank() -> None:
    """A simple function to test out the functionality of the pagerank algorithm.

    This uses the provided sample numbers given in the task assignment.
    """
    # The assignment only had the following as input, this test therefore also
    # contains the logic to parse it out into a proper python dict
    inp = "{1 -> 2, 1 -> 3, 2 -> 4, 3 -> 1, 3 -> 2, 3 -> 4, 4 -> 3}"

    link_map: dict[int, set[int]] = {}
    for pairing in inp.strip("{}").split(", "):
        key, val = pairing.split(" -> ")

        key = int(key)
        val = int(val)
        link_map.setdefault(key, set()).add(val)

    # The results were shown in an image, I will just rewrite them here:
    # (iterations count -> results)
    results = {
        0: [0.25, 0.25, 0.25, 0.25],
        1: [0.10833333, 0.21458333, 0.35625, 0.32083333],
        2: [0.1384375, 0.10447917, 0.35625, 0.32083333],
        3: [0.1384375, 0.19727344, 0.36904427, 0.29524479],
        4: [0.14206254, 0.20089848, 0.34729401, 0.30974497],
        5: [0.13589997, 0.19627655, 0.3611598, 0.3066368],
    }

    # To how many digits should the results be verified
    check_precision = 5

    for it_count, expected in results.items():
        # We're using tolerance of 0 to make sure we run for given amt of iterations exactly
        actual = pagerank(link_map, max_iter=it_count, tol=0)
        # convert to a list (use the format from the results)
        actual = [actual[1], actual[2], actual[3], actual[4]]

        for pos, (expected_num, actual_num) in enumerate(zip(expected, actual, strict=True)):
            assert round(expected_num, check_precision) == round( # noqa: S101
                actual_num, check_precision
            ), f"Invalid result for it={it_count} for {pos + 1}: {expected_num} != {actual_num}"
```

Zdrojový kód testovacej funkcie

## Spustenie programu

Program som spustil na stránke z ukážky v zadaní: <https://ailab.fai.utb.cz/> s ohraničovacím pravidlom zabraňujúcim prieskumu domén ktoré sa nekončia s \*.utb.cz (iné linky na týchto doménach boli zozbierané, len nie prehľadané). Tu som dostal nasledovný výsledok pri hĺbke 2:

Top 50 URLs by PageRank Score		
Rank	URL	PageRank Score
1	<a href="https://ailab.fai.utb.cz/">https://ailab.fai.utb.cz/</a>	1.6125e-08
2	<a href="https://www.facebook.com/ailab.fai.utb.cz/">https://www.facebook.com/ailab.fai.utb.cz/</a>	1.6106e-08
3	<a href="https://ailab.fai.utb.cz/contact/">https://ailab.fai.utb.cz/contact/</a>	1.6106e-08
4	<a href="https://ailab.fai.utb.cz/publicity/">https://ailab.fai.utb.cz/publicity/</a>	1.6106e-08
5	<a href="https://ailab.fai.utb.cz/first-competition-on-llm-designed-ea/">https://ailab.fai.utb.cz/first-competition-on-llm-designed-ea/</a>	1.6106e-08
6	<a href="https://ailab.fai.utb.cz/thesis-topics/">https://ailab.fai.utb.cz/thesis-topics/</a>	1.6106e-08
7	<a href="https://ailab.fai.utb.cz/our-team/">https://ailab.fai.utb.cz/our-team/</a>	1.6106e-08
8	<a href="https://twitter.com/A_I_Lab">https://twitter.com/A_I_Lab</a>	1.6106e-08
9	<a href="https://ailab.fai.utb.cz/bibliography/">https://ailab.fai.utb.cz/bibliography/</a>	1.6106e-08
10	<a href="https://ailab.fai.utb.cz/about-us/">https://ailab.fai.utb.cz/about-us/</a>	1.6106e-08
11	<a href="https://ailab.fai.utb.cz/projects/">https://ailab.fai.utb.cz/projects/</a>	1.6106e-08
12	<a href="https://ailab.fai.utb.cz/resources/">https://ailab.fai.utb.cz/resources/</a>	1.6106e-08
13	<a href="https://github.com/TBU-AILab">https://github.com/TBU-AILab</a>	1.6106e-08
14	<a href="https://www.researchgate.net/lab/AiLab-Roman-Senkerik">https://www.researchgate.net/lab/AiLab-Roman-Senkerik</a>	1.6106e-08
15	<a href="https://ailab.fai.utb.cz/wcc124specialsession/">https://ailab.fai.utb.cz/wcc124specialsession/</a>	1.6106e-08
16	<a href="https://fai.utb.cz/en/">https://fai.utb.cz/en/</a>	7.0107e-09
17	<a href="https://moodle.utb.cz/login/index.php?lang=en">https://moodle.utb.cz/login/index.php?lang=en</a>	6.6711e-09
18	<a href="https://twitter.com/UTBZlin">https://twitter.com/UTBZlin</a>	6.5551e-09
19	<a href="https://www.utb.cz">https://www.utb.cz</a>	6.3338e-09
20	<a href="https://www.utb.cz/">https://www.utb.cz/</a>	6.2522e-09
21	<a href="https://mapa.utb.cz/en/">https://mapa.utb.cz/en/</a>	6.2444e-09
22	<a href="https://www.utb.cz/en/privacy-policy/">https://www.utb.cz/en/privacy-policy/</a>	6.1198e-09
23	<a href="https://www.utb.cz/en/cookies/">https://www.utb.cz/en/cookies/</a>	6.1198e-09
24	<a href="https://www.utb.cz/en/physical-security/">https://www.utb.cz/en/physical-security/</a>	6.0666e-09
25	<a href="https://kmz.utb.cz/en/">https://kmz.utb.cz/en/</a>	6.0666e-09
26	<a href="https://fame.utb.cz/en/">https://fame.utb.cz/en/</a>	6.0666e-09
27	<a href="https://www.utb.cz/en/cyber-security/">https://www.utb.cz/en/cyber-security/</a>	5.9995e-09
28	<a href="https://uni.utb.cz/en/">https://uni.utb.cz/en/</a>	5.9655e-09
29	<a href="https://knihovna.utb.cz/en/">https://knihovna.utb.cz/en/</a>	5.9655e-09
30	<a href="https://ft.utb.cz/en/">https://ft.utb.cz/en/</a>	5.9655e-09
31	<a href="https://stag.utb.cz/portal/studium/index.html?pc_lang=en">https://stag.utb.cz/portal/studium/index.html?pc_lang=en</a>	5.8872e-09
32	<a href="https://fmk.utb.cz/en/">https://fmk.utb.cz/en/</a>	5.8532e-09
33	<a href="https://fhs.utb.cz/en/">https://fhs.utb.cz/en/</a>	5.8532e-09
34	<a href="https://flkp.utb.cz/en/">https://flkp.utb.cz/en/</a>	5.7898e-09
35	<a href="http://office365.utb.cz">http://office365.utb.cz</a>	5.6335e-09
36	<a href="https://www.utb.cz/en/">https://www.utb.cz/en/</a>	5.2529e-09
37	<a href="https://www.instagram.com/utbzlin/">https://www.instagram.com/utbzlin/</a>	4.8403e-09
38	<a href="https://www.youtube.com/c/UniverzitaTomaseBatiVeZline">https://www.youtube.com/c/UniverzitaTomaseBatiVeZline</a>	4.8291e-09
39	<a href="https://www.facebook.com/UTBZlin/">https://www.facebook.com/UTBZlin/</a>	4.7280e-09
40	<a href="https://www.linkedin.com/school/tomas-bata-university/">https://www.linkedin.com/school/tomas-bata-university/</a>	4.6615e-09
41	<a href="https://www.utb.cz/en/admissions/study-in-english/apply-now/">https://www.utb.cz/en/admissions/study-in-english/apply-now/</a>	4.4511e-09
42	<a href="https://www.utb.cz/en/employee/">https://www.utb.cz/en/employee/</a>	4.3508e-09
43	<a href="https://www.utb.cz/en/student/">https://www.utb.cz/en/student/</a>	4.1075e-09
44	<a href="https://www.utb.cz/en/university/about-the-university/location-map/">https://www.utb.cz/en/university/about-the-university/location-map/</a>	4.0931e-09
45	<a href="https://jidelnicek.utb.cz/">https://jidelnicek.utb.cz/</a>	4.0702e-09
46	<a href="https://stag.utb.cz/portal/">https://stag.utb.cz/portal/</a>	4.0441e-09
47	<a href="https://www.utb.cz/en/courses-offered/">https://www.utb.cz/en/courses-offered/</a>	4.0340e-09
48	<a href="https://www.utb.cz/en/university/study/studying-without-barriers/">https://www.utb.cz/en/university/study/studying-without-barriers/</a>	4.0009e-09
49	<a href="https://www.utb.cz/en/university/about-the-university/structure/rectorate-division-managed-by-rector/rectors-office-2/ombudsman/">https://www.utb.cz/en/university/about-the-university/structure/rectorate-division-managed-by-rector/rectors-office-2/ombudsman/</a>	3.9808e-09
50	<a href="https://www.utb.cz/en/admissions/">https://www.utb.cz/en/admissions/</a>	3.9808e-09

Následne som skúsil pagerank aplikovať na stránku: <https://mynixos.com/>, taktiež pri hĺbke 2, bez ohraničovacieho pravidla (prehľadávanie všetkých stránok). Scraper našiel 14888 odlišných stránok, po ich zoradení na základe skóre z pagerank algoritmu bolo TOP 50 nasledujúcich:

```
import base64
```

Top 50 URLs by PageRank Score

Rank	URL	PageRank Score
1	<a href="https://mynixos.com/packages">https://mynixos.com/packages</a>	5.5528e-08
2	<a href="https://mynixos.com/categories">https://mynixos.com/categories</a>	5.5528e-08
3	<a href="https://mynixos.com/privacy">https://mynixos.com/privacy</a>	5.5528e-08
4	<a href="https://mynixos.com/flakes">https://mynixos.com/flakes</a>	5.5528e-08
5	<a href="https://mynixos.com/pricing">https://mynixos.com/pricing</a>	5.5528e-08
6	<a href="https://mynixos.com/changelog">https://mynixos.com/changelog</a>	5.5528e-08
7	<a href="https://mynixos.com/">https://mynixos.com/</a>	5.5528e-08
8	<a href="https://mynixos.com/options">https://mynixos.com/options</a>	5.5528e-08
9	<a href="https://mynixos.com/about">https://mynixos.com/about</a>	5.5528e-08
10	<a href="https://mynixos.com/help">https://mynixos.com/help</a>	5.5528e-08
11	<a href="https://mynixos.com/terms">https://mynixos.com/terms</a>	5.5528e-08
12	<a href="https://support.cloudflare.com/hc/en-us/articles/200170016-What-is-Email-Address-Obfuscation-">https://support.cloudflare.com/hc/en-us/articles/200170016-What-is-Email-Address-Obfuscation-</a>	2.0673e-08
13	<a href="https://www.cloudflare.com/5xx-error-landing">https://www.cloudflare.com/5xx-error-landing</a>	2.0673e-08
14	<a href="https://www.cloudflare.com/sign-up?utm_source=email_protection">https://www.cloudflare.com/sign-up?utm_source=email_protection</a>	2.0673e-08
15	<a href="https://support.cloudflare.com/hc/en-us/categories/200275218-Getting-Started">https://support.cloudflare.com/hc/en-us/categories/200275218-Getting-Started</a>	2.0673e-08
16	<a href="https://mynixos.com/help/release">https://mynixos.com/help/release</a>	1.5475e-08
17	<a href="https://mynixos.com/help/nixos-options">https://mynixos.com/help/nixos-options</a>	1.1092e-08
18	<a href="https://mynixos.com/help/build">https://mynixos.com/help/build</a>	1.0527e-08
19	<a href="https://mynixos.com/help/resources">https://mynixos.com/help/resources</a>	9.7534e-09
20	<a href="https://nixos.org/manual/nix/stable/">https://nixos.org/manual/nix/stable/</a>	7.6077e-09
21	<a href="https://mynixos.com/help/install">https://mynixos.com/help/install</a>	7.2072e-09
22	<a href="https://mynixos.com/help/nixos-generate">https://mynixos.com/help/nixos-generate</a>	7.1780e-09
23	<a href="https://mynixos.com/help/remote-mirrors">https://mynixos.com/help/remote-mirrors</a>	7.1741e-09
24	<a href="https://mynixos.com/help/imports">https://mynixos.com/help/imports</a>	6.6348e-09
25	<a href="https://mynixos.com/help/nix-development-shell">https://mynixos.com/help/nix-development-shell</a>	6.3485e-09
26	<a href="https://mynixos.com/cdn-cgi/L/email-protection#d8b9cb5b1b698b5a1b6b1a0b7abf6bbb7b5">https://mynixos.com/cdn-cgi/L/email-protection#d8b9cb5b1b698b5a1b6b1a0b7abf6bbb7b5</a>	6.1634e-09
27	<a href="https://nixos.org/manual/nixos/stable/">https://nixos.org/manual/nixos/stable/</a>	6.0077e-09
28	<a href="https://mynixos.com/cdn-cgi/L/email-protection#056461686c6b45687c6b6c7d6a762b666a68">https://mynixos.com/cdn-cgi/L/email-protection#056461686c6b45687c6b6c7d6a762b666a68</a>	5.9500e-09
29	<a href="https://nycee.gitlab.io/home-manager/">https://nycee.gitlab.io/home-manager/</a>	5.8821e-09
30	<a href="https://mynixos.com/help/nixos-configuration">https://mynixos.com/help/nixos-configuration</a>	5.8693e-09
31	<a href="https://mynixos.com/help/home-manager">https://mynixos.com/help/home-manager</a>	5.8168e-09
32	<a href="https://mynixos.com/cdn-cgi/L/email-protection#2d5e585d5d425f594054434455425e034e4240">https://mynixos.com/cdn-cgi/L/email-protection#2d5e585d5d425f594054434455425e034e4240</a>	5.6890e-09
33	<a href="https://mynixos.com">https://mynixos.com</a>	5.6890e-09
34	<a href="https://daiderd.com/nix-darwin/manual/">https://daiderd.com/nix-darwin/manual/</a>	5.5974e-09
35	<a href="https://mynixos.com/help/nix-packages">https://mynixos.com/help/nix-packages</a>	5.5916e-09
36	<a href="https://mynixos.com/help/search">https://mynixos.com/help/search</a>	5.4071e-09
37	<a href="https://mynixos.com/help/download">https://mynixos.com/help/download</a>	5.3403e-09
38	<a href="https://nixos.org/manual/nixpkgs/stable/">https://nixos.org/manual/nixpkgs/stable/</a>	5.2601e-09
39	<a href="https://mynixos.com/nixpkgs">https://mynixos.com/nixpkgs</a>	4.4121e-09
40	<a href="https://mynixos.com/cdn-cgi/L/email-protection#0d6c696064634d6074636475627e236e6260">https://mynixos.com/cdn-cgi/L/email-protection#0d6c696064634d6074636475627e236e6260</a>	4.3607e-09
41	<a href="https://nixos.org/">https://nixos.org/</a>	3.7908e-09
42	<a href="https://mynixos.com/create">https://mynixos.com/create</a>	3.5343e-09
43	<a href="https://mynixos.com/home-manager">https://mynixos.com/home-manager</a>	3.4448e-09
44	<a href="https://opencollective.com/nixos">https://opencollective.com/nixos</a>	3.3782e-09
45	<a href="https://mynixos.com/cdn-cgi/L/email-protection#58393c353136183521363120372b763b3735">https://mynixos.com/cdn-cgi/L/email-protection#58393c353136183521363120372b763b3735</a>	3.2710e-09
46	<a href="https://mynixos.com/category/os-specific">https://mynixos.com/category/os-specific</a>	3.2475e-09
47	<a href="https://nixos.org/donate.html">https://nixos.org/donate.html</a>	3.2230e-09
48	<a href="https://ico.org.uk/make-a-complaint/">https://ico.org.uk/make-a-complaint/</a>	3.2133e-09
49	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>	3.2131e-09
50	<a href="https://mynixos.com/category/kde">https://mynixos.com/category/kde</a>	3.1900e-09

```
itsdrike ~/Current/Datamining/task4 main+ took 1m 1s $
```

## Záver

Z implementácie algoritmu PageRank som sa naučil, ako tento algoritmus hodnotí dôležitosť webových stránok na základe počtu a kvality odkazov. PageRank je dôležitý nástroj v kontexte vyhľadávačov, ktorý umožňuje určiť, ktoré stránky sú najdôležitejšie. Algoritmus začína s rovnakými hodnotami pre všetky stránky, ktoré sa iteratívne aktualizujú na základe odkazov medzi stránkami, pričom sa zabezpečuje stabilita prostredníctvom damping faktora.

Pri implementácii som nenarazil na žiadne vážne problémy, no pri samotnom scrapovaní stránok sa ukázalo, že program môže chvíľu trvať, najmä pri prehľadávaní väčšieho množstva stránok. To mi ukázalo, aké náročné môže byť spracovávanie rozsiahlych dát, ale tiež som si uvedomil, ako dôležité je optimalizovať tento proces pri práci s veľkými webovými sieťami. Celkovo sa mi podarilo úspešne implementovať PageRank a získať cenné skúsenosti s touto technológiou.