

CL-II 5 (2)

July 22, 2025

```
[ ]: '''NAME:Aher Swami Sandip
ROLL NO.01
COURSE: AI&DS
CLASS: BE
SUB:Computer Laboratory-II (Information Retrieval)'''
```

```
[2]: import os
import math
import requests
from bs4 import BeautifulSoup
import networkx as nx

# Step 1: Crawling - Collect links from a page
def get_links_from_url(url):
    try:
        response = requests.get(url)
        soup = BeautifulSoup(response.text, 'html.parser')
        links = set()
        for a_tag in soup.find_all('a', href=True):
            href = a_tag['href']
            if href.startswith("http"):
                links.add(href)
        return links
    except:
        return set()

# Step 2: Build a directed graph from a list of URLs
def build_web_graph(seed_urls, max_depth=1):
    graph = {}
    visited = set()

    def crawl(url, depth):
        if depth > max_depth or url in visited:
            return
        visited.add(url)
        links = get_links_from_url(url)
        graph[url] = links
```

```

        for link in links:
            crawl(link, depth + 1)

    for seed in seed_urls:
        crawl(seed, 0)

    return graph

# Step 3: Compute PageRank using simplified algorithm
def compute_pagerank(graph, damping=0.85, max_iterations=100, tol=1e-6):
    pages = list(graph.keys())
    n = len(pages)
    ranks = {page: 1.0 / n for page in pages}

    for _ in range(max_iterations):
        new_ranks = {}
        for page in pages:
            incoming_links = [p for p in pages if page in graph[p]]
            rank_sum = sum(ranks[p] / len(graph[p]) for p in incoming_links if graph[p])
            new_ranks[page] = (1 - damping) / n + damping * rank_sum

        # Check convergence
        delta = sum(abs(new_ranks[page] - ranks[page]) for page in pages)
        ranks = new_ranks
        if delta < tol:
            break

    return ranks

# Step 4: Run everything together
if __name__ == "__main__":
    seed_urls = [
        "https://en.wikipedia.org/wiki/Web_crawler",
        "https://en.wikipedia.org/wiki/PageRank"
    ]

    print("Building web graph...")
    web_graph = build_web_graph(seed_urls, max_depth=1)

    print("Computing PageRank...")
    pagerank_scores = compute_pagerank(web_graph)

    print("\nPageRank Results:")
    for page, score in sorted(pagerank_scores.items(), key=lambda x: x[1], reverse=True):
        print(f"{page}: {score:.5f}")

```

Building web graph...
Computing PageRank...

PageRank Results:

<https://www.mediawiki.org/>: 0.00160
<https://www.wikimedia.org/>: 0.00154
<https://developer.wikimedia.org>: 0.00151
[https://foundation.wikimedia.org/wiki/Special:MyLanguage/Policy:Universal_Code_o_f_Conduct](https://foundation.wikimedia.org/wiki/Special:MyLanguage/Policy:Universal_Code_of_Conduct): 0.00150
https://foundation.wikimedia.org/wiki/Special:MyLanguage/Policy:Cookie_statement: 0.00139
https://foundation.wikimedia.org/wiki/Special:MyLanguage/Policy:Privacy_policy: 0.00127
https://foundation.wikimedia.org/wiki/Special:MyLanguage/Policy:Terms_of_Use: 0.00115
<https://de.wikipedia.org/wiki/Webcrawler>: 0.00093
https://es.wikipedia.org/wiki/Ara%C3%B1a_web: 0.00091
https://en.wikipedia.org/wiki/Web_crawler: 0.00091
<https://nn.wikipedia.org/wiki/S%C3%B8kerobot>: 0.00091
<https://et.wikipedia.org/wiki/Veebi%C3%A4mblik>: 0.00091
<https://sr.wikipedia.org/wiki/Veb-popisiva%C4%8D>: 0.00091
https://ar.wikipedia.org/wiki/%D8%B2%D8%A7%D8%AD%D9%81_%D9%88%D9%8A%D8%A8: 0.00091
<https://zh.wikipedia.org/wiki/%E7%B6%B2%E8%B7%AF%E7%88%AC%E8%9F%B2>: 0.00091
https://ta.wikipedia.org/wiki/%E0%AE%B5%E0%AE%B2%E0%AF%88_%E0%AE%8A%E0%AE%8D%E0%AE%A4%E0%AE%BF: 0.00091
<https://it.wikipedia.org/wiki/Crawler>: 0.00091
https://ko.wikipedia.org/wiki/%EC%9B%B9_%ED%81%AC%EB%A1%A4%EB%9F%AC: 0.00091
<https://no.wikipedia.org/wiki/S%C3%B8kerobot>: 0.00091
https://cs.wikipedia.org/wiki/Web_crawler: 0.00091
https://az.wikipedia.org/wiki/Axtar%C4%B1%C5%9F_botu: 0.00091
https://uk.wikipedia.org/wiki/%D0%9F%D0%BE%D1%88%D1%83%D0%BA%D0%BE%D0%B2%D0%B8%D0%B9_%D1%80%D0%BE%D0%B1%D0%BE%D1%82: 0.00091
https://pl.wikipedia.org/wiki/Robot_internetowy: 0.00091
<https://ja.wikipedia.org/wiki/%E3%82%AF%E3%83%AD%E3%83%BC%E3%83%A9>: 0.00091
https://ca.wikipedia.org/wiki/Aranya_web: 0.00091
https://he.wikipedia.org/wiki/%D7%96%D7%97%D7%9C%D7%9F_%D7%A8%D7%A9%D7%AA: 0.00091
<https://hy.wikipedia.org/wiki/%D5%88%D6%80%D5%B8%D5%BC%D5%B6%D5%BF>: 0.00091
https://fr.wikipedia.org/wiki/Robot_d%27indexation: 0.00091
https://lv.wikipedia.org/wiki/T%C4%ABmek%C4%BCCa_r%C4%81pu%C4%BCprogramma: 0.00091
<https://th.wikipedia.org/wiki/%E0%B9%80%E0%B8%A7%E0%B9%87%E0%B8%9A%E0%B8%84%E0%B8%A3%E0%B8%AD%E0%B8%A7%E0%B9%8C%E0%B9%80%E0%B8%A5%E0%B8%AD%E0%B8%A3%E0%B9%8C>: 0.00091
https://tr.wikipedia.org/wiki/Arama_robotu: 0.00091
https://id.wikipedia.org/wiki/Perayap_web: 0.00091

https://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%D1%8B%D0%B9_%D1%80%D0%BE%D0%B1%D0%BE%D1%82: 0.00091
https://eu.wikipedia.org/wiki/Web_crawler: 0.00091
[https://ary.wikipedia.org/wiki/%D8%B1%D8%AA%D9%8A%D9%84%D8%A9_\(%D8%A8%D9%88%D8%A8\)](https://ary.wikipedia.org/wiki/%D8%B1%D8%AA%D9%8A%D9%84%D8%A9_(%D8%A8%D9%88%D8%A8)): 0.00091
<https://fi.wikipedia.org/wiki/Hakurobotti>: 0.00091
https://fa.wikipedia.org/wiki/%D8%AE%D8%B2%D9%86%D8%AF%D9%87_%D9%88%D8%A8: 0.00091
https://el.wikipedia.org/wiki/%CE%91%CE%BD%CE%B9%CF%87%CE%BD%CE%B5%CF%85%CF%84%C%E%AE%CF%82_%CE%B9%CF%83%CF%84%CE%BF%CF%8D: 0.00091
<https://bar.wikipedia.org/wiki/Webcrawler>: 0.00091
https://ro.wikipedia.org/wiki/Robot_de_c%C4%83utare: 0.00091
https://ms.wikipedia.org/wiki/Perangkak_web: 0.00091
<https://hu.wikipedia.org/wiki/Keres%C5%91robot>: 0.00091
<https://nl.wikipedia.org/wiki/Spider>: 0.00091
https://hr.wikipedia.org/wiki/Web_pauk: 0.00091
<https://zh-classical.wikipedia.org/wiki/%E7%88%AC%E7%B6%B2%E8%9F%B2%E6%A2%B0>: 0.00091
https://pt.wikipedia.org/wiki/Rastreador_web: 0.00091
https://qu.wikipedia.org/wiki/Web_aysakuq: 0.00091
https://ia.wikipedia.org/wiki/Rastreator_de_rete: 0.00091
<https://af.wikipedia.org/wiki/Webkruiper>: 0.00091
https://simple.wikipedia.org/wiki/Web_crawler: 0.00091
https://lt.wikipedia.org/wiki/Interneto_robotas: 0.00091
https://cy.wikipedia.org/wiki/Ymgripiwr_gwe: 0.00091
https://mhr.wikipedia.org/wiki/%D0%9A%D1%8B%D1%87%D0%B0%D0%BB%D1%88%D0%B5_%D1%80%D0%BE%D0%B1%D0%BE%D1%82: 0.00091
<https://nds-nl.wikipedia.org/wiki/Webkroeper>: 0.00091
[https://sv.wikipedia.org/wiki/Spindel_\(internet\)](https://sv.wikipedia.org/wiki/Spindel_(internet)): 0.00091
<https://www.wikidata.org/wiki/Special:EntityPage/Q45842>: 0.00090
<https://www.wikidata.org/wiki/Special:EntityPage/Q45842#sitelinks-wikipedia>: 0.00090
<https://en.wikipedia.org/wiki/PageRank>: 0.00081
<https://commons.wikimedia.org/wiki/Category:PageRank>: 0.00081
<https://www.wikidata.org/wiki/Special:EntityPage/Q184316>: 0.00080
<https://hi.wikipedia.org/wiki/%E0%A4%AA%E0%A5%87%E0%A4%9C%E0%A4%B0%E0%A5%88%E0%A4%82%E0%A4%95>: 0.00079
<https://zh.wikipedia.org/wiki/PageRank>: 0.00079
<https://ru.wikipedia.org/wiki/PageRank>: 0.00079
<https://kn.wikipedia.org/wiki/%E0%B2%AA%E0%B3%87%E0%B2%9C%E0%B3%8D%E2%80%8D%E0%B2%80%E0%B3%8D%E0%B2%AF%E0%B2%BE%E0%B2%82%E0%B2%95%E0%B3%8D>: 0.00079
[https://ta.wikipedia.org/wiki/%E0%AE%AA%E0%AF%87%E0%AE%9C%E0%AF%8D_%E0%AE%A4%E0%AE%80%E0%AE%BF%E0%AE%9A%E0%AF%88](https://ta.wikipedia.org/wiki/%E0%AE%AA%E0%AF%87%E0%AE%9C%E0%AF%8D_%E0%AE%A4%E0%AE%80%E0%AE%B5%E0%AE%80%E0%AE%BF%E0%AE%9A%E0%AF%88): 0.00079
<https://uk.wikipedia.org/wiki/PageRank>: 0.00079
<https://fr.wikipedia.org/wiki/PageRank>: 0.00079
<https://tr.wikipedia.org/wiki/PageRank>: 0.00079
<https://de.wikipedia.org/wiki/PageRank>: 0.00079
<https://ms.wikipedia.org/wiki/PageRank>: 0.00079

<https://hy.wikipedia.org/wiki/PageRank>: 0.00079
<https://es.wikipedia.org/wiki/PageRank>: 0.00079
<https://hu.wikipedia.org/wiki/PageRank>: 0.00079
<https://eu.wikipedia.org/wiki/PageRank>: 0.00079
<https://pt.wikipedia.org/wiki/PageRank>: 0.00079
<https://bg.wikipedia.org/wiki/PageRank>: 0.00079
<https://ca.wikipedia.org/wiki/PageRank>: 0.00079
[https://ar.wikipedia.org/wiki/%D8%A8%D9%8A%D8%AC_%D8%B1%D8%A7%D9%86%D9%83_\(%D8%A%D9%88%D8%A7%D8%B1%D8%B2%D9%85%D9%8A%D8%A9\)](https://ar.wikipedia.org/wiki/%D8%A8%D9%8A%D8%AC_%D8%B1%D8%A7%D9%86%D9%83_(%D8%A%D9%88%D8%A7%D8%B1%D8%B2%D9%85%D9%8A%D8%A9)): 0.00079
<https://pl.wikipedia.org/wiki/PageRank>: 0.00079
<https://id.wikipedia.org/wiki/PageRank>: 0.00079
<https://it.wikipedia.org/wiki/PageRank>: 0.00079
<https://zh-yue.wikipedia.org/wiki/PageRank>: 0.00079
<https://no.wikipedia.org/wiki/PageRank>: 0.00079
<https://cs.wikipedia.org/wiki/PageRank>: 0.00079
<https://ro.wikipedia.org/wiki/PageRank>: 0.00079
<https://ja.wikipedia.org/wiki/%E3%83%9A%E3%83%BC%E3%82%B8%E3%83%A9%E3%83%B3%E3%82%AF>: 0.00079
<https://he.wikipedia.org/wiki/PageRank>: 0.00079
<https://bn.wikipedia.org/wiki/%E0%A6%AA%E0%A7%87%E0%A6%9C%E0%A6%B0%E2%80%8D%E0%A7%8D%E0%A6%AF%E0%A6%BE%E0%A6%82%E0%A6%95>: 0.00079
<https://fi.wikipedia.org/wiki/PageRank>: 0.00079
<https://vi.wikipedia.org/wiki/PageRank>: 0.00079
<https://th.wikipedia.org/wiki/%E0%B9%80%E0%B8%9E%E0%B8%88%E0%B9%81%E0%B8%A3%E0%B8%87%E0%B8%81%E0%B9%8C>: 0.00079
https://sr.wikipedia.org/wiki/%D0%9F%D0%B5%D1%98%D1%9F_%D1%80%D0%B0%D0%BD%D0%BA: 0.00079
<https://ko.wikipedia.org/wiki/%ED%8E%98%EC%9D%B4%EC%A7%80%EB%9E%AD%ED%81%AC>: 0.00079
<https://lt.wikipedia.org/wiki/PageRank>: 0.00079
<https://fa.wikipedia.org/wiki/%D9%BE%DB%8C%D8%AC%E2%80%8C%D8%B1%D9%86%DA%A9>: 0.00079
<https://sk.wikipedia.org/wiki/PageRank>: 0.00079
https://gu.wikipedia.org/wiki/%E0%AA%AA%E0%AB%87%E0%AA%9C_%E0%AA%B0%E0%AB%87%E0%AA%A8%E0%AB%8D%E0%AA%95: 0.00079
<https://da.wikipedia.org/wiki/PageRank>: 0.00079
<https://kk.wikipedia.org/wiki/PageRank>: 0.00079
<https://nl.wikipedia.org/wiki/PageRank>: 0.00079
https://el.wikipedia.org/wiki/%CE%91%CE%BB%CE%B3%CF%8C%CF%81%CE%B9%CE%B8%CE%BC%C_E%BF%CF%82_PageRank: 0.00079
https://az.wikipedia.org/wiki/Page_Rank: 0.00079
<https://sv.wikipedia.org/wiki/Pagerank>: 0.00079
<https://www.wikidata.org/wiki/Special:EntityPage/Q184316#sitelinks-wikipedia>: 0.00078
<https://wikimediafoundation.org/>: 0.00061
<https://d-nb.info/gnd/4796298-7>: 0.00056
<https://web.archive.org/web/20091213213920/http://wiki.foaf-project.org/w/Scutter>: 0.00056

http://www.redorbit.com/news/education/318480/stanford_earns_336_million_off_google_stock/: 0.00044
http://infolab.stanford.edu/~backrub/google.html: 0.00044
http://www.stanfordalumni.org/news/magazine/2004/novdec/features/startup.html: 0.00042
https://web.archive.org/web/20211206205907/https://webbrowsersintroduction.com/: 0.00042
https://www.wikidata.org/wiki/Q45842#identifiers: 0.00042
http://chato.cl/research/crawling_thesis: 0.00042
http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Agents/spetka/spetka.html: 0.00041
https://webbrowsersintroduction.com/: 0.00041
http://www.cs.washington.edu/homes/pedrod/papers/nips01b.pdf: 0.00041
http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=: 0.00041
https://web.archive.org/web/20060427013310/http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=1999-66&format=pdf&compression=: 0.00041
https://web.archive.org/web/20040903174942/http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Agents/spetka/spetka.html: 0.00040
https://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html: 0.00040
https://web.archive.org/web/20130702063520/https://www.google.com/competition/howgooglesearchworks.html: 0.00040
http://www.wiley.com/legacy/compbooks/sonnenreich/history.html: 0.00040
https://web.archive.org/web/20090310190153/http://www.stanfordalumni.org/news/magazine/2004/novdec/features/startup.html: 0.00040
https://www.google.com/competition/howgooglesearchworks.html: 0.00040
http://oak.cs.ucla.edu/~cho/research/crawl.html: 0.00040
http://wiki.foaf-project.org/w/Scutter: 0.00040
https://patents.google.com/patent/US6285999: 0.00040
http://www.cs.duke.edu/nicl/netecon06/papers/ne06-sybil.pdf: 0.00040
http://searchengineland.com/google-has-confirmed-they-are-removing-toolbar-pagerank-244230: 0.00039
https://www.slideshare.net/denshe/icwe13-tutorial-webcrawling: 0.00039
http://searchengineland.com/what-is-google-pagerank-a-guide-for-searchers-webmasters-11068: 0.00039
https://web.archive.org/web/20160703031514/http://searchengineland.com/what-is-google-pagerank-a-guide-for-searchers-webmasters-11068: 0.00039
http://www10.org/cdrom/papers/pdf/p208.pdf: 0.00039
https://web.archive.org/web/20171224210412/http://www10.org/cdrom/papers/pdf/p208.pdf: 0.00039
http://www-cs-students.stanford.edu/~taherh/papers/secondeigenvalue.pdf: 0.00039
https://doi.org/10.1145%2F358923.358934: 0.00039
https://www.google.com/press/funfacts.html: 0.00039
https://web.archive.org/web/20010715123343/https://www.google.com/press/funfacts.html: 0.00039
https://patents.google.com/patent/US7058628B1/en: 0.00039
https://en.wikiquote.org/wiki/PageRank: 0.00039

http://infolab.stanford.edu/pub/papers/google.pdf: 0.00039
https://support.apple.com/en-us/HT204683: 0.00039
http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf: 0.00039
http://vigna.dsi.unimi.it/ftp/papers/UbiCrawler.pdf: 0.00039
http://vigna.dsi.unimi.it/ftp/papers/ParadoxicalPageRank.pdf: 0.00039
https://web.archive.org/web/20040317210216/http://www.fxpal.com/people/vdberg/pubs/www8/www1999f.pdf: 0.00039
http://www.robotstxt.org/wc/exclusion.html: 0.00039
https://www.slideshare.net/denshe/intelligent-crawling-shestakovwiat13: 0.00039
http://www2003.org/cdrom/papers/refereed/p007/p7-abiteboul.html: 0.00038
http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PT02&Sect2=HITOFF&u=%2Fnetrahtml%2FPT0%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,058,628.PN.&OS=pn/7,058,628&RS=PN/7,058,628: 0.00038
http://patft1.uspto.gov/netacgi/nph-Parser?Sect1=PT01&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetrahtml%2FPT0%2Fsrchnum.htm&r=1&f=G&l=50&s1=6,799,176.PN.&OS=PN/6,799,176&RS=PN/6,799,176: 0.00038
http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PT02&Sect2=HITOFF&u=%2Fnetrahtml%2FPT0%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7,269,587.PN.&OS=pn/7,269,587&RS=PN/7,269,587: 0.00038
http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html: 0.00038
https://doi.org/10.1145/2F1062745.1062789: 0.00038
https://doi.org/10.1145%2F342009.335391: 0.00038
http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=6,285,999: 0.00038
https://doi.org/10.1002%2Fspe.587: 0.00038
https://web.archive.org/web/20100821011759/http://www.cs.duke.edu/nic1/netecon06/papers/ne06-sybil.pdf: 0.00038
https://web.archive.org/web/20090320030833/http://vigna.dsi.unimi.it/ftp/papers/UbiCrawler.pdf: 0.00038
http://www.cs.brown.edu/courses/cs227/2002/cache/Cho.pdf: 0.00038
http://www-cs-students.stanford.edu/~taherh/papers/comparison.pdf: 0.00038
https://www.google.com/technology/: 0.00038
http://news.bbc.co.uk/2/hi/science/nature/8238462.stm: 0.00038
https://web.archive.org/web/20080623233116/http://www.google.com/technology/: 0.00038
http://www.rankdex.com/about.html: 0.00038
http://bradlove.org/blog/cogsci-page-rank: 0.00038
http://bradlove.org/papers/love_sloman_1995.pdf: 0.00038
http://oak.cs.ucla.edu/~cho/papers/cho-thesis.pdf: 0.00038
https://web.archive.org/web/20071107021800/http://www.robotstxt.org/wc/exclusion.html: 0.00038
http://www.robotstxt.org/wc/guidelines.html: 0.00038
https://web.archive.org/web/20050422045839/http://www.robotstxt.org/wc/guidelines.html: 0.00038
http://searchengineland.com/google-toolbar-pagerank-officially-goes-dark-247553: 0.00038
https://doi.org/10.1016%2FS0169-7552%2898%2900110-X: 0.00038
http://www.mattcutts.com/blog/how-to-report-paid-links/: 0.00038
https://web.archive.org/web/20140829134755/http://patft.uspto.gov/netacgi/nph-

Parser?patentnumber=6,285,999: 0.00038
<https://web.archive.org/web/20180331090147/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PT02>: 0.00038
http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2010_Navigli_Lapata.pdf: 0.00038
<http://dbpubs.stanford.edu:8091/diglib/pub/projectdir/google.html>: 0.00038
<http://nmis.isti.cnr.it/sebastiani/Publications/ACL07.pdf>: 0.00038
<https://web.archive.org/web/20190828223450/http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PT02&Sect2=HITOFF&u=%2Fnetahhtml%2FPT0%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=7%2C058%2C628.PN.&OS=pn%2F7%2C058%2C628&RS=PN%2F7%2C058%2C628:0.00038>
<http://dbpubs.stanford.edu:8090/pub/1998-51>: 0.00038
<http://www.mattcutts.com/blog/seo-for-bloggers/>: 0.00038
<http://www.seomoz.org/article/search-ranking-factors>: 0.00038
<https://web.archive.org/web/20171211160207/http://bradlove.org/blog/cogsci-page-rank>: 0.00038
<https://doi.org/10.1145%2F775152.775192>: 0.00037
<https://arxiv.org/abs/cs/0503069>: 0.00037
<https://github.com/bedirhan/wivet>: 0.00037
<https://web.archive.org/web/20150927004511/http://infolab.stanford.edu/pub/papers/google.pdf>: 0.00037
<https://doi.org/10.1002%2Ffasi.20078>: 0.00037
<http://informatics.indiana.edu/fil/Papers/AA98.ps>: 0.00037
<http://informatics.indiana.edu/fil/Papers/ICML.ps>: 0.00037
<https://doi.org/10.1002%2F8SICI%291099-1425%28199806%291%3A1%3C15%3A%3AAID-JOS3%3E3.0.C0%3B2-K>: 0.00037
<https://web.archive.org/web/20081217094957/http://www-cs-students.stanford.edu/~taherh/papers/secondeigenvalue.pdf>: 0.00037
<https://doi.org/10.1038%2F21987>: 0.00037
<https://pubmed.ncbi.nlm.nih.gov/10428673>: 0.00037
https://web.archive.org/web/20090408084918/http://www.redorbit.com/news/education/318480/stanford_earns_336_million_off_google_stock/: 0.00037
<http://pages.stern.nyu.edu/~panos/publications/icde2005.pdf>: 0.00037
<http://www.chato.cl/papers/baeza02balancing.pdf>: 0.00037
<http://www.mccurley.org/papers/fractal.pdf>: 0.00037
<https://doi.org/10.1145%2F572326.572328>: 0.00037
<https://web.archive.org/web/20050905013119/http://pages.stern.nyu.edu/~panos/publications/icde2005.pdf>: 0.00037
<https://web.archive.org/web/20110511180227/http://www.mattcutts.com/blog/pagerank-sculpting/>: 0.00037
<https://web.archive.org/web/20140625233510/http://www10.org/cdrom/papers/210/index.html>: 0.00037
<http://www10.org/cdrom/papers/210/index.html>: 0.00037
<http://ilpubs.stanford.edu:8090/347/>: 0.00037
<http://dollar.biz.uiowa.edu/~pant/Papers/crawling.pdf>: 0.00037
<https://web.archive.org/web/20090320030832/http://dollar.biz.uiowa.edu/~pant/Papers/crawling.pdf>: 0.00037
<http://www.mattcutts.com/blog/pagerank-sculpting/>: 0.00037
<https://web.archive.org/web/20200116103001/https://patents.google.com/patent/US7>

058628B1/en: 0.00037
<https://web.archive.org/web/20050612001742/http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html>: 0.00037
<https://doi.org/10.1137%2FS1064827502412875>: 0.00037
<http://www.eecs.harvard.edu/cs286r/courses/fall11/papers/AT%2705.pdf>: 0.00037
<https://www.google.com/insidesearch/howsearchworks/algorithms.html>: 0.00037
<https://www.wired.com/2007/01/tax-takers-send-in-the-spiders/>: 0.00037
<https://www.canada.ca/en/revenue-agency/services/about-canada-revenue-agency-cra/protecting-your-privacy/privacy-impact-assessment/xenon-web-crawling-initiative-privacy-impact-assessment-summary.html>: 0.00037
<http://www.cindoc.csic.es/cybermetrics/pdf/68.pdf>: 0.00037
<http://clgiles.ist.psu.edu/papers/VLDB-2000-focused-crawling.pdf>: 0.00037
<https://web.archive.org/web/20010904075500/http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/pinkerton/WebCrawler.html>: 0.00037
<https://citeseerkx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.185>: 0.00037
<http://www.fxpal.com/people/vdberg/pubs/www8/www1999f.pdf>: 0.00037
<https://web.archive.org/web/20060219085958/http://www.cindoc.csic.es/cybermetrics/pdf/68.pdf>: 0.00037
http://www.scit.wlv.ac.uk/~in7803/publications/cothey_2004.pdf: 0.00037
https://www.researchgate.net/publication/220724572_A_Fast_Community_Based_Algorithm_for_Generating_Web_Crawler_Seeds_Set: 0.00037
<https://arxiv.org/abs/2210.17300>: 0.00037
<https://arxiv.org/abs/math/0307056>: 0.00037
<https://arxiv.org/abs/1610.01544>: 0.00037
<https://arxiv.org/abs/1206.6904>: 0.00037
<https://arxiv.org/abs/physics/0612011>: 0.00037
<https://arxiv.org/abs/0805.3322>: 0.00037
<https://web.archive.org/web/20160310093405/http://searchengineland.com/google-has-confirmed-they-are-removing-toolbar-pagerank-244230>: 0.00037
https://www.forbes.com/forbes/2009/1005/technology-baidu-robin-li-man-whos-beating-google_2.html: 0.00037
<https://archive.org/details/googlestory00vise/page/37>: 0.00037
<https://doi.org/10.1109%2F4236.707687>: 0.00037
http://www.iicm.tugraz.at/Ressourcen/Papers/dangers_google.pdf: 0.00037
<https://doi.org/10.1016%2F0022-2496%2877%2990033-5>: 0.00037
https://web.archive.org/web/20140116234734/http://www.iicm.tugraz.at/Ressourcen/Papers/dangers_google.pdf: 0.00037
<https://search.worldcat.org/issn/0169-7552>: 0.00037
<https://doi.org/10.1016%2F0306-4573%2876%2990048-0>: 0.00037
<https://web.archive.org/web/20150525015816/http://www.rankdex.com/about.html>: 0.00037
<https://arxiv.org/abs/2302.14596>: 0.00037
https://web.archive.org/web/20140318134631/http://semanticweb.com/more-semantics-for-google-ita-software-acquisition-comes-with-needlebase-too_b19329: 0.00037
<https://arxiv.org/abs/1407.5107>: 0.00037
<https://api.semanticscholar.org/CorpusID:7587743>: 0.00037
<https://stats.wikimedia.org/#/en.wikipedia.org>: 0.00037

https://donate.wikimedia.org/?wmf_source=donate&wmf_medium=sidebar&wmf_campaign=en.wikipedia.org&uselang=en: 0.00037
<https://arxiv.org/abs/0804.1630>: 0.00037
<https://arxiv.org/abs/1002.2858>: 0.00037
<https://arxiv.org/abs/1205.1960>: 0.00037
<https://techcrunch.com/2012/05/08/swifttype-launch/>: 0.00037
<https://arxiv.org/abs/1208.3071>: 0.00037
<https://web.archive.org/web/20160421224919/http://searchengineland.com/google-toolbar-pagerank-officially-goes-dark-247553>: 0.00037
<https://web.archive.org/web/20091017081100/http://www.google.com/support/forum/p/Webmasters/thread?tid=6a1d6250e26e9e48&hl=en>: 0.00037
<https://web.archive.org/web/20141203194914/http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf>: 0.00037
<https://web.archive.org/web/20170413154033/https://www.searchenginejournal.com/google-pagerank-official-shuts-doors-public/161874/>: 0.00037
<http://ilpubs.stanford.edu:8090/697/1/2005-33.pdf>: 0.00037
<https://www.searchenginejournal.com/google-pagerank-official-shuts-doors-public/161874/>: 0.00037
<https://web.archive.org/web/20070528054725/http://www.mattcutts.com/blog/how-to-report-paid-links/>: 0.00037
<https://web.archive.org/web/20100628120615/http://www.cs.washington.edu/homes/pe/drod/papers/nips01b.pdf>: 0.00036
<https://web.archive.org/web/20061128230555/http://dbpubs.stanford.edu:8091/diglib/pub/projectdir/google.html>: 0.00036
<http://www.seomoz.org/blog/pagerank-sculpting-parsing-the-value-and-potential-benefits-of-sculpting-pr-with-nofollow>: 0.00036
https://web.archive.org/web/20101214110431/http://www.dsi.uniroma1.it/~navigli/pubs/PAMI_2010_Navigli_Lapata.pdf: 0.00036
<https://web.archive.org/web/20100207231048/http://www.mattcutts.com/blog/seo-for-bloggers/>: 0.00036
<https://web.archive.org/web/20090507232106/http://www.seomoz.org/article/search-ranking-factors/>: 0.00036
<https://web.archive.org/web/20070628113247/http://nmis.isti.cnr.it/sebastiani/Publications/ACL07.pdf>: 0.00036
<https://web.archive.org/web/20110512060125/http://news.bbc.co.uk/2/hi/science/nature/8238462.stm>: 0.00036
<https://web.archive.org/web/20101216084254/http://www-cs-students.stanford.edu/~taherh/papers/comparison.pdf>: 0.00036
<https://web.archive.org/web/20110514201211/http://www.seomoz.org/blog/pagerank-sculpting-parsing-the-value-and-potential-benefits-of-sculpting-pr-with-nofollow>: 0.00036
https://en.wikipedia.org/w/index.php?title=Web_crawler&oldid=1301855004: 0.00036
<https://www.researchgate.net/publication/264620349>: 0.00036
<https://support.google.com/webmasters/bin/answer.py?hl=en&answer=174992>: 0.00036
<https://api.semanticscholar.org/CorpusID:9362566>: 0.00036
<https://doi.org/10.1145%2F2389936.2389949>: 0.00036
<https://doi.org/10.1145%2F2380718.2380762>: 0.00036
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.36.6087>: 0.00036

<https://web.archive.org/web/20121221113630/http://informatics.indiana.edu/fil/Papers/AA98.ps>: 0.00036
<https://api.semanticscholar.org/CorpusID:18513666>: 0.00036
<https://api.semanticscholar.org/CorpusID:325714>: 0.00036
<http://www.mendeley.com/download/public/1423991/3893295922/dc0f7d824fd2a8fbcc84f6fdf9e4f337d343987d/dl.pdf>: 0.00036
https://doi.org/10.1007%2F978-3-642-02457-3_74: 0.00036
<https://api.semanticscholar.org/CorpusID:3710903>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.5538>: 0.00036
<https://api.semanticscholar.org/CorpusID:6416041>: 0.00036
<https://api.semanticscholar.org/CorpusID:147958>: 0.00036
<https://doi.org/10.1002%2Fcpe.2980>: 0.00036
<https://doi.org/10.1002%2Ffasi.20388>: 0.00036
http://semanticweb.com/more-semantics-for-google-ita-software-acquisition-comes-with-needlebase-too_b19329: 0.00036
<https://ui.adsabs.harvard.edu/abs/2005cs...3069N>: 0.00036
<https://api.semanticscholar.org/CorpusID:205690364>: 0.00036
<https://ui.adsabs.harvard.edu/abs/1999Natur.400..107L>: 0.00036
http://chato.cl/papers/baeza05_crawling_country_better_breadth_first_web_page_ordering.pdf: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.5877>: 0.00036
<https://api.semanticscholar.org/CorpusID:16718130>: 0.00036
<http://web.archive.org/web/20020106130042/http://www.wolfbot.com/>: 0.00036
<https://web.archive.org/web/20140706120641/https://oa.doria.fi/handle/10024/38506>: 0.00036
<https://api.semanticscholar.org/CorpusID:10316730>: 0.00036
<https://doi.org/10.1142%2F3725>: 0.00036
<https://www.researchgate.net/publication/44241179>: 0.00036
<https://doi.org/10.1016%2Fs1389-1286%2899%2900052-3>: 0.00036
<https://doi.org/10.1016%2Fs0169-023x%2804%2900107-7>: 0.00036
<https://web.archive.org/web/20161222185208/https://www.wired.com/2007/01/tax-takers-send-in-the-spiders/>: 0.00036
<https://doi.org/10.1145%2F371920.371960>: 0.00036
<https://api.semanticscholar.org/CorpusID:4347646>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1018.1506>: 0.00036
<http://web.archive.org/web/20121221113620/http://informatics.indiana.edu/fil/Papers/ICML.ps>: 0.00036
<https://doi.org/10.1016%2Fs0169-7552%2898%2900110-x>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.126.6094>: 0.00036
<https://www.springer.com/gp/book/9783540233381>: 0.00036
<http://cis.poly.edu/tr/tr-cis-2001-03.pdf>: 0.00036
https://doi.org/10.1007%2F978-3-540-30216-2_14: 0.00036
http://www.scit.wlv.ac.uk/%7Ecml1993/papers/Web_Crawling_Ethics_preprint.doc: 0.00036
<https://doi.org/10.1145%2F958942.958945>: 0.00036
<https://doi.org/10.1145%2F857166.857170>: 0.00036
<https://hdl.handle.net/20.500.11937%2F48288>: 0.00036
<https://web.archive.org/web/20170925203155/https://www.canada.ca/en/revenue->

agency/services/about-canada-revenue-agency-cra/protecting-your-privacy/privacy-impact-assessment/xenon-web-crawling-initiative-privacy-impact-assessment-summary.html: 0.00036
<https://web.archive.org/web/20051001063017/http://vigna.dsi.unimi.it/ftp/papers/ParadoxicalPageRank.pdf>: 0.00036
<https://oa.doria.fi/handle/10024/38506>: 0.00036
<https://web.archive.org/web/20190616133656/https://www.forbes.com/sites/jamesaltucher/2011/03/18/10-unusual-things-about-google-also-the-worst-vc-decision-i-ever-made/>: 0.00036
<https://doi.org/10.1371%2Fjournal.pone.0178458>: 0.00036
<https://api.semanticscholar.org/CorpusID:10284718>: 0.00036
<https://web.archive.org/web/20160325232656/http://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines>: 0.00036
<https://api.semanticscholar.org/CorpusID:1755112>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/26288312>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/28575009>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2003math..7056N>: 0.00036
<https://web.archive.org/web/20200521035518/https://support.google.com/webmasters/answer/66356?hl=en>: 0.00036
<https://doi.org/10.2140%2Finvoke.2012.5.463>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/23382878>: 0.00036
<https://cs.nyu.edu/media/publications/TR2003-846.pdf>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2013PLoS0...854204B>: 0.00036
<https://arxiv.org/abs/cs.GL/0601030>: 0.00036
<https://doi.org/10.1145%2F1255175.1255273>: 0.00036
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.58.9060>: 0.00036
<https://www.ams.org/featurecolumn/archive/pagerank.html>: 0.00036
<http://dl.acm.org/citation.cfm?id=2488433>: 0.00036
<https://web.archive.org/web/20110818093436/http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>: 0.00036
<https://web.archive.org/web/20150402221831/https://managedadmin.com/phoenix-seo-services/google-page-rank-update-coming/>: 0.00036
<https://www.bloomberg.com/news/articles/2015-10-26/google-turning-its-lucrative-web-search-over-to-ai-machines>: 0.00036
<https://doi.org/10.1088%2F1742-5468%2F2008%2F07%2FP07008>: 0.00036
<https://web.archive.org/web/20080603171020/http://dbpubs.stanford.edu:8090/pub/1998-51>: 0.00036
<https://en.wikipedia.org/w/index.php?title=PageRank&oldid=1293484181>: 0.00036
<https://managedadmin.com/phoenix-seo-services/google-page-rank-update-coming>: 0.00036
<https://api.semanticscholar.org/CorpusID:9855132>: 0.00036
<http://searchenginewatch.com/article/2334934/So...-You-Think-SEO-Has-Changed>: 0.00036
http://wwwusers.di.uniroma1.it/~navigli/pubs/ACL_2013_Pilehvar_Jurgens_Navigli.pdf: 0.00036
<https://ui.adsabs.harvard.edu/abs/2008PhRvE..78c6107P>: 0.00036
<https://api.semanticscholar.org/CorpusID:8375649>: 0.00036

<https://ui.adsabs.harvard.edu/abs/2015PLoS..1034794S>: 0.00036
<https://web.archive.org/web/20131004234101/http://groups.yahoo.com/group/sd-2/>: 0.00036
https://web.archive.org/web/20171223215954/http://bradlove.org/papers/love_sloman_1995.pdf: 0.00036
<https://groups.yahoo.com/group/sd-2/>: 0.00036
<https://doi.org/10.1145%2F2488388.2488433>: 0.00036
<https://doi.org/10.1103%2FPhysRevE.78.036107>: 0.00036
<https://web.archive.org/web/20160304112752/http://research.microsoft.com/pubs/66874/qd-pagerank.pdf>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.582.9402>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.148.9274>: 0.00036
<https://www.forbes.com/sites/jamesaltucher/2011/03/18/10-unusual-things-about-google-also-the-worst-vc-decision-i-ever-made/>: 0.00036
<https://api.semanticscholar.org/CorpusID:257232659>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2023JSMTE2023i3401W>: 0.00036
<https://web.archive.org/web/20090126204112/http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/28076982>: 0.00036
<https://api.semanticscholar.org/CorpusID:26880621>: 0.00036
<https://web.archive.org/web/20120120002301/http://www.rankdex.com/about.html>: 0.00036
<https://doi.org/10.1016%2Fj.ipl.2015.02.015>: 0.00036
<https://doi.org/10.1093%2Fbioinformatics%2Fbtq680>: 0.00036
<https://doi.org/10.1371%2Fjournal.pone.0054204>: 0.00036
<https://doi.org/10.1017%2Fs1049096507070771>: 0.00036
<https://api.semanticscholar.org/CorpusID:6012229>: 0.00036
http://www.people.fas.harvard.edu/~gillum/rankings_paper.pdf: 0.00036
<https://hdl.handle.net/10026.1%2F9713>: 0.00036
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5456068>: 0.00036
<https://web.archive.org/web/20151015185034/http://www.google.com/patents/US628599>: 0.00036
https://web.archive.org/web/20131001004443/http://wwwusers.di.uniroma1.it/~navigli/pubs/ACL_2013_Pilehvar_Jurgens_Navigli.pdf: 0.00036
<https://doi.org/10.1137%2F140976649>: 0.00036
<https://citeserx.ist.psu.edu/viewdoc/summary?doi=10.1.1.99.3942>: 0.00036
<https://web.archive.org/web/20190627071550/https://www.nytimes.com/2006/09/17/business/yourmoney/17baidu.html>: 0.00036
https://web.archive.org/web/20150213104340/http://www.people.fas.harvard.edu/~gillum/rankings_paper.pdf: 0.00036
<https://api.semanticscholar.org/CorpusID:3115544>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/21149343>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2008JSMTE..07..008J>: 0.00036
<https://developers.google.com/search/blog/2019/09/evolving-nofollow-new-ways-to-identify>: 0.00036
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3558500>: 0.00036
<https://pubmed.ncbi.nlm.nih.gov/18851105>: 0.00036
<https://web.archive.org/web/20181223110045/https://searchengineland.com/google->

hummingbird-172816: 0.00036
<https://doi.org/10.1371/journal.pone.0134794>: 0.00036
<https://support.google.com/places/bin/answer.py?hl=en&answer=7091>: 0.00036
<https://hdl.handle.net/10338.dmlcz/101787>: 0.00036
<https://doi.org/10.1016/j.tcs.2014.04.003>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2006SJSC..27.1181F>: 0.00036
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4545754>: 0.00036
<https://searchengineland.com/google-hummingbird-172816>: 0.00036
<https://www.google.com/patents?hl=en&lr=&vid=USPAT5920859&id=x04ZAAAAEBAJ&oi=fnd&dq=yanhong+li&printsec=abstract#v=onepage&q=yanhong%20li&f=false>: 0.00036
<https://doi.org/10.1080/13658810802022822>: 0.00036
<https://search.worldcat.org/issn/1944-4184>: 0.00036
<https://api.semanticscholar.org/CorpusID:1436859>: 0.00036
<https://arxiv.org/archive/math.H0>: 0.00036
<https://doi.org/10.1088/1742-5468/21/01/014>: 0.00036
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.5930>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2017PLoS0..1278458B>: 0.00036
<https://search.worldcat.org/issn/1932-6203>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2006cs..1030B>: 0.00036
<https://arxiv.org/archive/cs.IR>: 0.00036
<https://www.nytimes.com/2006/09/17/business/yourmoney/17baidu.html>: 0.00036
<https://en.wikiquote.org/wiki/Special:Search/PageRank>: 0.00036
<https://api.semanticscholar.org/CorpusID:118605727>: 0.00036
<https://web.archive.org/web/20140331043853/http://searchenginewatch.com/article/2334934/So...-You-Think-SEO-Has-Changed>: 0.00036
<https://www.searchenginewatch.com/2011/07/25/google-directory-has-been-shut-down/>: 0.00036
https://web.archive.org/web/20130308094445/http://www.forbes.com/forbes/2009/100/technology-baidu-robin-li-man-whos-beating-google_2.html: 0.00036
<https://doi.org/10.1002/Fnet.20442>: 0.00036
<https://www.google.com/support/forum/p/Webmasters/thread?tid=6a1d6250e26e9e48&hl=en>: 0.00036
<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.18.5264>: 0.00036
<https://web.archive.org/web/20120526230407/http://support.google.com/places/bin/answer.py?hl=en&answer=7091>: 0.00036
<http://research.microsoft.com/pubs/66874/qd-pagerank.pdf>: 0.00036
<https://ui.adsabs.harvard.edu/abs/2009IJGIS..23..823J>: 0.00036
<https://web.archive.org/web/20140706215017/http://techcrunch.com/2012/05/08/swift-type-launch/>: 0.00036
https://doi.org/10.1007/F978-3-540-30216-2_10: 0.00036
<https://api.semanticscholar.org/CorpusID:207205045>: 0.00036
<https://arxiv.org/archive/math.C0>: 0.00036
<https://doi.org/10.1142/FS0129065717500137>: 0.00036
<https://web.archive.org/web/20111205225726/http://www.google.com/patents?hl=en&lr=&vid=USPAT5920859&id=x04ZAAAAEBAJ&oi=fnd&dq=yanhong+li&printsec=abstract#v=onepage&q=yanhong%20li&f=false>: 0.00036
https://www.podomatic.com/podcasts/gwientjes/episodes/2015-09-05T06_05_37-07_00: 0.00036

<https://support.google.com/webmasters/answer/66356?hl=en>: 0.00036

[]: