

```
In [ ]: from bs4 import BeautifulSoup
import requests
from urllib.request import urljoin
from urllib.request import urlparse
```

```
In [ ]: links_intern = set()
links_extern = set()
input_url=input("Enter Link:")
depth=1
```

BFS Crawling

```
In [ ]: def level_crawler(input_url):
    temp_urls=set()
    current_url_domain=urlparse(input_url).netloc
    beautiful_soup_object = BeautifulSoup(
requests.get(input_url).content, "html.parser")
    for anchor in beautiful_soup_object.findAll("a"):
        href = anchor.attrs.get("href")
        if(href!="" or href!=None):
            href=urljoin(input_url, href)
            href_parsed=urlparse(href)
            href=href_parsed.scheme
            href+="://"
            href+=href_parsed.netloc
            href+=href_parsed.path
            final_parsed_href=urlparse(href)
            is_valid=bool(final_parsed_href.scheme) and bool(
final_parsed_href.netloc)
            if is_valid:
                if current_url_domain not in href and href not in links_extern:
                    print("External - {}".format(href))
                    links_extern.add(href)
                if current_url_domain in href and href not in links_intern:
                    print("Internal - {}".format(href))
                    links_intern.add(href)
                    temp_urls.add(href)
    return temp_urls
```

```
In [ ]: if(depth == 0):
    print("Intern - {}".format(input_url))
elif(depth == 1):
    level_crawler(input_url)
else:
    #We have used a BFS approach
    # considering The structure as a tree.
    # It uses queue based approach to traverse links upto a particular depth
    queue = []
    queue.append(input_url)
    for j in range(depth):
        for count in range(len(queue)):
            url = queue.pop(0)
            urls = level_crawler(url)
            for i in urls:
                queue.append(i)
```

Internal - <https://www.vit.ac.in>
External - <https://viteee.vit.ac.in/>
External - <https://admissions.vit.ac.in/ugnriapplication>
External - <https://admissions.vit.ac.in/pgapplication/>
External - <http://chennai.vit.ac.in/>
External - <https://vitap.ac.in/>
External - <https://vitbhopal.ac.in/>
Internal - <https://www.vit.ac.in/>
External - <https://vit.ac.in>
External - <https://vit.ac.in/about-vit>
External - <https://vit.ac.in/about/vision-mission>
External - <https://vit.ac.in/vit-milestones>
External - <https://vit.ac.in/about/leadership>
External - <https://vit.ac.in/governance>
External - <https://vit.ac.in/about/administrative-offices>
External - <https://vit.ac.in/about/infrastructure>
External - <https://vit.ac.in/about/sustainability>
External - <https://vit.ac.in/true-green>
External - <https://vit.ac.in/about/community-outreach>
External - <https://vit.ac.in/about/communityradio>
External - <https://vit.ac.in/all-news-archieved>
External - <https://vit.ac.in/all-events>
External - <https://vit.ac.in/national-institutional-ranking-framework-nirf>
External - <https://vit.ac.in/mhrdugcaicte>
External - <http://careers.vit.ac.in/>
External - <https://vit.ac.in/about/news-letter>
External - <https://vit.ac.in/academics/home>
External - <https://vit.ac.in/sites/default/files/academic/Academic-Regulations.pdf>
External - <https://vit.ac.in/programmes-offered-1>
External - <https://vit.ac.in/programmes-offered-2021-22>
External - <https://vit.ac.in/programmes-offered-2020-21>
External - <https://vit.ac.in/schools>
External - <https://vit.ac.in/academics/ffcs>
External - <https://vit.ac.in/academics/library>
External - <https://vit.ac.in/academics-feedback>
External - <https://vit.ac.in/admissions/overview>
External - <https://vit.ac.in/admissions/programmes-offered>
External - <https://vit.ac.in/all-courses/ug>
External - <https://vit.ac.in/all-courses/pg>
External - <https://vit.ac.in/admissions/research>
External - <https://vit.ac.in/admissions/international>
External - <https://vit.ac.in/stars-support-advancement-rural-students-0>
External - <https://vit.ac.in/placements/overview>
External - <https://vit.ac.in/career-development-centre>
External - <https://vit.ac.in/placements/superdreamoffers>
External - <https://vit.ac.in/placements/dreamoffers>
External - <https://vit.ac.in/placements/internship>
External - <https://vit.ac.in/placements/statistics>
External - <https://vit.ac.in/placements/pat-Office>
External - <https://vit.ac.in/placements/contact-us>
External - <https://vit.ac.in/InternationalRelations>
External - <https://vit.ac.in/internationalrelations/itp>
External - [https://vit.ac.in/academics-more/Semester Abroad Program \(SAP\)](https://vit.ac.in/academics-more/Semester Abroad Program (SAP))
External - <https://vit.ac.in/internationalrelations/partneruniversities>
External - <https://vit.ac.in/admissions/international/overview>
External - <https://vit.ac.in/academics-more/Contact us>
External - <https://vit.ac.in/research>
External - <https://vit.ac.in/research/academic>
External - <https://vit.ac.in/research/sponsored-research>
External - <https://vit.ac.in/iprcell>
External - <https://vit.ac.in/research/centers-list>
External - <https://vit.ac.in/campuslife/overview>
External - <https://vit.ac.in/campuslife/fests>
External - <https://vit.ac.in/campuslife/studentswelfare>
External - <https://vit.ac.in/campus-category/Newsletter>
External - <https://vit.ac.in/campus-category/Clubs>
External - <https://vit.ac.in/campus-category/Chapters>
External - <https://vit.ac.in/campus-category/Campus-Events>
External - <https://vit.ac.in/campus-category/Counselling-Division>

External - <https://vit.ac.in/campus-category/grievancecell>
External - <https://vit.ac.in/campus-category/Programme Representatives>
External - <https://vit.ac.in/campus-category/Student-Council>
External - <https://vit.ac.in/campus/Teams/PROJECTS-LAB>
External - <https://vit.ac.in/campuslife/sports>
External - <https://vit.ac.in/campuslife/hostels>
External - <https://vit.ac.in/campuslife/healthservices>
External - <https://vit.ac.in/campuslife/otheramenities>
External - <https://vit.ac.in/detailview/green-vit>
External - <https://vit.ac.in/innovation-and-startup-policy>
External - <https://vit.ac.in/institutions-innovation-council>
External - <https://vit.ac.in/vitol/>
External - [javascript://void\(0\)](javascript://void(0))
External - <https://vit.ac.in/academics/coe>
External - <https://vit.ac.in/transcripts-alumni>
External - <https://vit.applydirect.org/student/>
External - <https://applications.wes.org/required-documents/Search/Result>
External - <https://www.wes.org/advisor-blog/send-electronic-academic-records/>
External - <https://vit.ac.in/files/e-SANAD-Notification.pdf>
External - <https://web.vit.ac.in/CertificateVerification/login>
External - <https://vit.ac.in/coe-email-contacts>
External - <https://vit.ac.in/centers/asc>
External - <http://www.vittbi.com/>
External - <https://vtop.vit.ac.in/>
External - <https://vit.ac.in/guest-house>
External - <https://vit.ac.in/redressal>
External - <https://vit.ac.in/hotels-in-vellore>
External - <https://mail.google.com/>
External - <http://intranet.vit.ac.in>
External - https://vit.ac.in/sites/default/files/SBST_Freshers_App.rar
External - <https://vit.ac.in/anti-ragging-committee>
External - <https://vit.ac.in/capability-enhancement-scheme>
External - https://vit.ac.in/sites/default/files/Industrial_Visit_Form.doc
External - <https://vit.ac.in/internal-complaints-committee>
External - https://vit.ac.in/alumni_progression
External - <https://vit.ac.in/academics/transcripts>
External - <https://vit.ac.in/instruction>
External - <https://www.campusinteraction.com/>
External - <https://vtop.vit.ac.in/vtop/initialProcess>
External - <https://vit.ac.in/alumni-events>
External - <https://www.vitaa.org/page/Distinguished-Alumni-Award-Winners.dz>
External - <https://vit.ac.in/sites/default/files/vitaa/e-Tracks-2020/index.html>
External - <https://vit.ac.in/detailview/alumni-photo-gallery>
External - <http://www.vitaa.org/>
External - <https://vit.ac.in/alumni-office-contact>
External - <https://campustour.vit.ac.in/>
Internal - <https://www.vit.ac.in/> <https://admissions.vit.ac.in/pgapplication/>
Internal - <https://www.vit.ac.in/> <https://vit.ac.in/vit-rank/>
Internal - <https://www.vit.ac.in/> QS-Rank-2022
Internal - <https://www.vit.ac.in/> RANKING
External - <https://vit.ac.in/ranking-and-recognition>
External - <https://www.facebook.com/VITUniversity/>
External - https://twitter.com/vit_univ
External - <https://www.linkedin.com/school/vellore-institute-of-technology/>
External - https://www.instagram.com/vellore_vit/
External - <https://www.youtube.com/c/VITUniversityVellore>
External - <https://vit.ac.in/school-advanced-sciences-sas/vit-mathematical-meet-2023>
External - <https://vit.ac.in/school-electronics-engineering-sense/three-day-workshop-emcemi-techniques-and-standards>
External - <https://vit.ac.in/general/one-day-workshop-non-destructive-techniques-concrete-structures>
External - <https://vit.ac.in/school-mechanical-engineering-smec/international-conference-materials-manufacturing-and-modelling>
External - <https://vit.ac.in/school-electronics-engineering-sense/2nd-international-conference-vision-towards-emerging-trends-0>
External - <https://web.vit.ac.in/phd>
External - <https://admissions.vit.ac.in/freshersportal/>
External - <http://info.vit.ac.in/phd-vlr-dec-2022/index.asp>
External - <http://info.vit.ac.in/phd-che-dec-2022/index.asp>

External - <https://results.vit.ac.in/APDEC2022/>
 External - <https://results.vit.ac.in/BPDEC2022/>
 External - <https://vit.ac.in/our-ambassadors>
 External - <https://admissions.vit.ac.in/onlinewithdraw2022>
 External - <https://vit.ac.in/schools-centres-list-research-guides-2022>
 External - <https://vit.ac.in/Ariia-2021/>
 Internal - <https://www.vit.ac.in/national-institutional-ranking-framework-nirf>
 External - <https://vit.ac.in/events-photo-gallery>
 External - <https://vit.ac.in/scholarship>
 External - <https://www.vidyalakshmi.co.in/Students/>
 Internal - <https://www.vit.ac.in/ranking>
 External - <https://vit.ac.in/vit-rank>
 External - <https://vit.ac.in/ariia-ranking>
 External - <https://vit.ac.in/vit-346th-best-institution-world-engineering-and-technology>
 External - <https://vit.ac.in/vit-rank/>
 Internal - <https://www.vit.ac.in/academics/InternationalRelations>
 Internal - <https://www.vit.ac.in/placements/statistics>
 External - <https://vit.ac.in/vit-university-sets-record-limca-book-records>
 Internal - <https://www.vit.ac.in/admission-overview>
 Internal - <https://www.vit.ac.in/all-news>
 External - <https://vit.ac.in/youth-red-cross-blood-donation-camp-0>
 External - <https://vit.ac.in/2-vit-students-get-rs-102-crore-salary-offer-0>
 External - <https://vit.ac.in/viteee-2022-exam-results-news-clips-0>
 External - <https://vit.ac.in/vit-engineering-entrance-exam-viteee-news>
 Internal - <https://www.vit.ac.in/sih2022-achievement-0>
 Internal - <https://www.vit.ac.in/prof-mookambeswaran-vijayalakshmi-vit-world-renowned-scientist>
 Internal - <https://www.vit.ac.in/vit-alumnus-recognized-best-performing-ias-officer>
 Internal - <https://www.vit.ac.in/vit-sae-team-rotorfpv>
 Internal - <https://www.vit.ac.in/5-teams-vit-qualify-finals-nasscom---ibm-all-india-hackathon>
 Internal - <https://www.vit.ac.in/virtual-track-competitions-communications-award>
 Internal - <https://www.vit.ac.in/team-roverx-creation-labs-vit-vellore-has-achieved-new-milestone-year-scoring-9036-its-system>
 Internal - <https://www.vit.ac.in/3d-printing-play-major-role-mitigating-spread-covid-19-say-researchers-vit>
 External - <https://vit.ac.in/galleries>
 External - <http://campustour.vit.ac.in/>
 External - <https://vit.ac.in/video>
 External - <https://vit.ac.in/campus-hostel/hostels>
 External - <http://info.vit.ac.in/guesthouse/>
 External - <https://vit.ac.in/vit-privacy-policy>
 External - <https://vit.ac.in/sitemap/>
 External - <https://vit.ac.in/academics/iqac>
 External - <http://www.mhrdnats.gov.in/>
 External - <https://vit.ac.in/sites/default/files/Student-Code-of-Conduct.pdf>
 External - <http://intranet.vit.ac.in/>
 External - <https://peopleorbit.vit.ac.in/>
 External - <https://webmail.vit.ac.in/>
 External - <https://careers.vit.ac.in/>
 External - <https://www.vitaa.org/>
 External - <https://vit.ac.in/contactus>
 External - <https://vit.ac.in/guesthouse/>
 External - <https://vit.ac.in/covid-19-initiatives>
 External - <https://www.facebook.com/vituniversity/>
 External - <https://campustour.vit.ac.in>
 External - <https://vit.ac.in/tender-notifications>

DFS Crawling

In []:

```

def dfs(base, path, visited, max_depth=3, depth=0):
    if depth < max_depth:
        try:
            soup = BeautifulSoup(requests.get(base + path).text, "html.parser")
            for link in soup.find_all("a"):
                href = link.get("href")

```

```

    if href not in visited:
        visited.add(href)
        print(f"at depth {depth}: {href}")
        if href.startswith("http"):
            dfs(href, "", visited,max_depth-1, depth + 1)
        else:
            dfs(base, href,visited,max_depth-1, depth + 1)
except:
    pass

```

In []:

```

link="https://www.vit.ac.in"
dfs(link, "", set([link]))

```

```

at depth 0: #main-content
at depth 1: https://viteee.vit.ac.in/?utm_source=V71TI1557&utm_campaign=viteee2023
at depth 1: https://admissions.vit.ac.in/ugnriapplication
at depth 1: https://admissions.vit.ac.in/pgapplication/?utm_source=V71TI1557&utm_campaign=mba2023
at depth 1: http://chennai.vit.ac.in/
at depth 1: https://vitap.ac.in/
at depth 1: https://vitbhopal.ac.in/
at depth 1: /
at depth 1: #
at depth 1: https://vit.ac.in
at depth 1: https://vit.ac.in/about-vit
at depth 1: https://vit.ac.in/about/vision-mission
at depth 1: https://vit.ac.in/vit-milestones
at depth 1: https://vit.ac.in/about/leadership
at depth 1: https://vit.ac.in/governance
at depth 1: https://vit.ac.in/about/administrative-offices
at depth 1: https://vit.ac.in/about/infrastructure
at depth 1: https://vit.ac.in/about/sustainability
at depth 1: https://vit.ac.in/true-green
at depth 1: https://vit.ac.in/about/community-outreach
at depth 1: https://vit.ac.in/about/communityradio
at depth 1: https://vit.ac.in/all-news-archieved
at depth 1: https://vit.ac.in/all-events
at depth 1: https://vit.ac.in/national-institutional-ranking-framework-nirf
at depth 1: https://vit.ac.in/mhrdugcaicte
at depth 1: http://careers.vit.ac.in/
at depth 1: https://vit.ac.in/about/news-letter
at depth 1: https://vit.ac.in/academics/home
at depth 1: https://vit.ac.in/sites/default/files/academic/Academic-Regulations.pdf
at depth 1: https://vit.ac.in/programmes-offered-1
at depth 1: https://vit.ac.in/programmes-offered-2021-22
at depth 1: https://vit.ac.in/programmes-offered-2020-21
at depth 1: https://vit.ac.in/schools
at depth 1: https://vit.ac.in/academics/ffcs
at depth 1: https://vit.ac.in/academics/library
at depth 1: https://vit.ac.in/academics-feedback
at depth 1: https://vit.ac.in/admissions/overview
at depth 1: https://vit.ac.in/admissions/programmes-offered
at depth 1: https://vit.ac.in/all-courses/ug
at depth 1: https://vit.ac.in/all-courses/pg
at depth 1: https://vit.ac.in/admissions/research
at depth 1: https://vit.ac.in/admissions/international
at depth 1: https://vit.ac.in/stars-support-advancement-rural-students-0
at depth 1: https://vit.ac.in/placements/overview
at depth 1: https://vit.ac.in/career-development-centre
at depth 1: https://vit.ac.in/placements/superdreamoffers
at depth 1: https://vit.ac.in/placements/dreamoffers
at depth 1: https://vit.ac.in/placements/internship
at depth 1: https://vit.ac.in/placements/statistics
at depth 1: https://vit.ac.in/placements/pat-Office
at depth 1: https://vit.ac.in/placements/contact-us
at depth 1: https://vit.ac.in/InternationalRelations
at depth 1: https://vit.ac.in/internationalrelations/itp

```

at depth 1: [https://vit.ac.in/academics-more/Semester Abroad Program \(SAP\)](https://vit.ac.in/academics-more/Semester%20Abroad%20Program%20(SAP))
at depth 1: <https://vit.ac.in/internationalrelations/partneruniversities>
at depth 1: <https://vit.ac.in/admissions/international/overview>
at depth 1: [https://vit.ac.in/academics-more/Contact us](https://vit.ac.in/academics-more/Contact%20us)
at depth 1: <https://vit.ac.in/research>
at depth 1: <https://vit.ac.in/research/academic>
at depth 1: <https://vit.ac.in/research/sponsored-research>
at depth 1: <https://vit.ac.in/iprcell>
at depth 1: <https://vit.ac.in/research/centers-list>
at depth 1: <https://vit.ac.in/campuslife/overview>
at depth 1: <https://vit.ac.in/campuslife/fests>
at depth 1: <https://vit.ac.in/campuslife/studentwelfare>
at depth 1: <https://vit.ac.in/campus-category/Newsletter>
at depth 1: <https://vit.ac.in/campus-category/Clubs>
at depth 1: <https://vit.ac.in/campus-category/Chapters>
at depth 1: <https://vit.ac.in/campus-category/Campus-Events>
at depth 1: <https://vit.ac.in/campus-category/Counselling-Division>
at depth 1: <https://vit.ac.in/campus-category/grievancecell>
at depth 1: [https://vit.ac.in/campus-category/Programme Representatives](https://vit.ac.in/campus-category/Programme%20Representatives)
at depth 1: <https://vit.ac.in/campus-category/Student-Council>
at depth 1: <https://vit.ac.in/campus/Teams/PROJECTS-LAB>
at depth 1: <https://vit.ac.in/campuslife/sports>
at depth 1: <https://vit.ac.in/campuslife/hostels>
at depth 1: <https://vit.ac.in/campuslife/healthservices>
at depth 1: <https://vit.ac.in/campuslife/otheramenities>
at depth 1: <https://vit.ac.in/detailview/green-vit>
at depth 1: <https://vit.ac.in/innovation-and-startup-policy>
at depth 1: <https://vit.ac.in/institutions-innovation-council>
at depth 1: <https://vit.ac.in/vitol/>
at depth 1: javascript:void(0)
at depth 1: <https://vit.ac.in/academics/coe>
at depth 1: <https://vit.ac.in/transcripts-alumni>
at depth 1: <https://vit.applydirect.org/student/#/app/request>
at depth 1: <https://applications.wes.org/required-documents/Search/Result?cId=2&eId=1&iNm=Vellore%20Institute%20of%20Technology%20%28VIT%29&cNm=abc&iSec=University%20Education&iSub=Bachelor%27s%20Degree%20Programs&comp=1&isEca=False&lan=0>
at depth 1: <https://www.wes.org/advisor-blog/send-electronic-academic-records/>
at depth 1: <https://vit.ac.in/files/e-SANAD-Notification.pdf>
at depth 1: <https://web.vit.ac.in/CertificateVerification/login>
at depth 1: <https://vit.ac.in/coe-email-contacts>
at depth 1: <https://vit.ac.in/centers/asc>
at depth 1: <http://www.vittbi.com/#/>
at depth 1: <https://vtop.vit.ac.in/>
at depth 1: <https://vit.ac.in/guest-house>
at depth 1: <https://vit.ac.in/redressal>
at depth 1: <https://vit.ac.in/hotels-in-vellore>
at depth 1: <https://mail.google.com/>
at depth 1: <http://intranet.vit.ac.in>
at depth 1: https://vit.ac.in/sites/default/files/SBST_Freshers_App.rar
at depth 1: <https://vit.ac.in/anti-ragging-committee>
at depth 1: <https://vit.ac.in/capability-enhancement-scheme>
at depth 1: https://vit.ac.in/sites/default/files/Industrial_Visit_Form.doc
at depth 1: <https://vit.ac.in/internal-complaints-committee>
at depth 1: https://vit.ac.in/alumni_progression
at depth 1: <https://vit.ac.in/academics/transcripts>
at depth 1: <https://vit.ac.in/instruction>
at depth 1: <https://www.campusinteraction.com/>
at depth 1: <https://vtop.vit.ac.in/vtop/initialProcess>
at depth 1: <https://vit.ac.in/alumni-events>
at depth 1: <https://www.vitaa.org/page/Distinguished-Alumni-Award-Winners.dz>
at depth 1: <https://vit.ac.in/sites/default/files/vitaa/e-Tracks-2020/index.html>
at depth 1: <https://vit.ac.in/detailview/alumni-photo-gallery>
at depth 1: <http://www.vitaa.org/>
at depth 1: <https://vit.ac.in/alumni-office-contact>
at depth 1: <https://campustour.vit.ac.in/>
at depth 1: https://admissions.vit.ac.in/pgapplication/?utm_source=V71TI1557&utm_campaign=mba2023
at depth 1: <https://vit.ac.in/vit-rank/>
at depth 1: QS-Rank-2022

at depth 1: RANKING
at depth 1: https://viteee.vit.ac.in/?utm_source=V71TI1557&utm_campaign=viteee2023
at depth 1: <https://admissions.vit.ac.in/pgapplication/>
at depth 1: <https://vit.ac.in/ranking-and-recognition>
at depth 1: <https://www.facebook.com/VITUniversity/>
at depth 1: https://twitter.com/vit_univ
at depth 1: <https://www.linkedin.com/school/vellore-institute-of-technology/>
at depth 1: https://www.instagram.com/vellore_vit/
at depth 1: <https://www.youtube.com/c/VITUniversityVellore>
at depth 1: <https://vit.ac.in/school-advanced-sciences-sas/vit-mathematical-meet-2023>
at depth 1: <https://vit.ac.in/school-electronics-engineering-sense/three-day-workshop-emcemi-techniques-and-standards>
at depth 1: <https://vit.ac.in/general/one-day-workshop-non-destructive-techniques-concrete-structures>
at depth 1: <https://vit.ac.in/school-mechanical-engineering-smec/international-conference-materials-manufacturing-and-modelling>
at depth 1: <https://vit.ac.in/school-electronics-engineering-sense/2nd-international-conference-vision-towards-emerging-trends-0>
at depth 1: <https://web.vit.ac.in/phd>
at depth 1: <https://admissions.vit.ac.in/freshersportal/>
at depth 1: <http://info.vit.ac.in/phd-vlr-dec-2022/index.asp>
at depth 1: <http://info.vit.ac.in/phd-che-dec-2022/index.asp>
at depth 1: <https://results.vit.ac.in/APDEC2022/>
at depth 1: <https://results.vit.ac.in/BPDEC2022/>
at depth 1: <https://vit.ac.in/our-ambassadors>
at depth 1: <https://admissions.vit.ac.in/onlinewithdraw2022>
at depth 1: <https://vit.ac.in/schools-centres-list-research-guides-2022>
at depth 1: <https://vit.ac.in/Ariia-2021/>
at depth 1: [/national-institutional-ranking-framework-nirf](https://national-institutional-ranking-framework-nirf)
at depth 1: <https://vit.ac.in/events-photo-gallery>
at depth 1: <https://vit.ac.in/scholarship>
at depth 1: <https://www.vidyalakshmi.co.in/Students/>
at depth 1: [/ranking](https://vit.ac.in/ranking)
at depth 1: <https://vit.ac.in/vit-rank>
at depth 1: <https://vit.ac.in/ariia-ranking>
at depth 1: <https://vit.ac.in/vit-346th-best-institution-world-engineering-and-technology>
at depth 1: <https://vit.ac.in/vit-rank/>
at depth 1: [/academics/InternationalRelations](https://vit.ac.in/academics/InternationalRelations)
at depth 1: [placements/statistics](https://vit.ac.in/placements/statistics)
at depth 1: <https://vit.ac.in/vit-university-sets-record-limca-book-records>
at depth 1: [/admission-overview](https://vit.ac.in/admission-overview)
at depth 1: [/all-news](https://vit.ac.in/all-news)
at depth 1: <https://vit.ac.in/youth-red-cross-blood-donation-camp-0>
at depth 1: <https://vit.ac.in/2-vit-students-get-rs-102-crore-salary-offer-0>
at depth 1: <https://vit.ac.in/viteee-2022-exam-results-news-clips-0>
at depth 1: <https://vit.ac.in/vit-engineering-entrance-exam-viteee-news>
at depth 1: [sih2022-achievement-0](https://vit.ac.in/sih2022-achievement-0)
at depth 1: [prof-mookambeswaran-vijayalakshmi-vit-world-renowned-scientist](https://vit.ac.in/prof-mookambeswaran-vijayalakshmi-vit-world-renowned-scientist)
at depth 1: [vit-alumnus-recognized-best-performing-ias-officer](https://vit.ac.in/vit-alumnus-recognized-best-performing-ias-officer)
at depth 1: [vit-sae-team-rotorfpv](https://vit.ac.in/vit-sae-team-rotorfpv)
at depth 1: [5-teams-vit-qualify-finals-nasscom---ibm-all-india-hackathon](https://vit.ac.in/5-teams-vit-qualify-finals-nasscom---ibm-all-india-hackathon)
at depth 1: [virtual-track-competitions-communications-award](https://vit.ac.in/virtual-track-competitions-communications-award)
at depth 1: [team-roverx-creation-labs-vit-vellore-has-achieved-new-milestone-year-scoring-9036-its-system](https://vit.ac.in/team-roverx-creation-labs-vit-vellore-has-achieved-new-milestone-year-scoring-9036-its-system)
at depth 1: [3d-printing-play-major-role-mitigating-spread-covid-19-say-researchers-vit](https://vit.ac.in/3d-printing-play-major-role-mitigating-spread-covid-19-say-researchers-vit)
at depth 1: <https://vit.ac.in/galleries>
at depth 1: <http://campustour.vit.ac.in/>
at depth 1: <https://vit.ac.in/video>
at depth 1: <https://vit.ac.in/campus-hostel/hostels>
at depth 1: <http://info.vit.ac.in/guesthouse/>
at depth 1: <https://vit.ac.in/vit-privacy-policy>
at depth 1: <https://vit.ac.in/sitemap/>
at depth 1: <https://vit.ac.in/academics/iqac>
at depth 1: <http://www.mhrdnats.gov.in/>
at depth 1: <https://vit.ac.in/sites/default/files/Student-Code-of-Conduct.pdf>
at depth 1: <http://intranet.vit.ac.in/>

```
at depth 1: https://peopleorbit.vit.ac.in/
at depth 1: https://webmail.vit.ac.in/
at depth 1: https://careers.vit.ac.in/
at depth 1: https://www.vitaa.org/
at depth 1: https://vit.ac.in/contactus
at depth 1: https://vit.ac.in/guesthouse/
at depth 1: https://vit.ac.in/covid-19-initiatives
at depth 1: https://www.facebook.com/vituniversity/
at depth 1: https://campustour.vit.ac.in
at depth 1: https://vit.ac.in/tender-notifications
```




Twitter Scraping Data using Twint Library

```
In [ ]: import twint
import nest_asyncio
nest_asyncio.apply()
```

```
In [ ]: c = twint.Config()
c.Search = "FIFA Qatar 2022"
c.Limit = 10
twint.run.Search(c)
```


[illegible]

eration for Football Associations (FIFA) has said that Lionel #Messi and Kylian #Mba ppe top the list of nominees for the Best Player Award at the "The Best FIFA Footb all Awards" slated for February 27, 2023. #FIFA #Qatar2022 <https://t.co/9wbCeihWMH>

1607030470138822658 2022-12-25 20:38:15 +0530 <aboubakrmounir> 4K #Wallpapers L
 FIFA World Cup .  FIFA World Cup Qatar 2022 Game Changers (v1.0). Opti mized for iOS 16.  Find full quality download links in this thread. i Behance: h <https://t.co/jztpQLptMD> <https://t.co/nGLoY0mCJT>

1607030219759845376 2022-12-25 20:37:16 +0530 <SoyRosmy1> FIFA investiga cómo chef t urco Salt Bae ingresó al campo en final de Qatar 2022 <https://t.co/2Cqkgc9asI> #Fe lizNavidadVenezuela

1607029408375754752 2022-12-25 20:34:02 +0530 <PatadaIndie> Kylian Mbappe habla sobr e la 'obsesión' de la Copa del Mundo antes de los cuartos de final de Inglaterra #ar gentina #catar #fifa #francia #inglaterra #mundial #qatar #qatar2022 #WorldCup <http://t.co/VDLAAWEDsD> <https://t.co/VDLAAWEDsD>

1607028524052799489 2022-12-25 20:30:31 +0530 <EfectoCocuyo> #SepaQue La FIFA invest iga cómo el cocinero Salt Bae, cuyo nombre real es Nusret Gokce, obtuvo junto a otra s personas "acceso indebido" en el campo tras la final de Qatar 2022. Más detalles aquí  <https://t.co/hDJ8GUNQ4S>

1607028403244244992 2022-12-25 20:30:03 +0530 <voz_hispana> La FIFA no detectó casos de amaños de partidos durante Qatar 2022 <https://t.co/i5ayAiMmmm>

1607027950817509376 2022-12-25 20:28:15 +0530 <hattabHolding> The FIFA World Cup fin als were incredible! It's been quite a journey since we first met the courageous fin alists and caught up on their most memorable experiences. We'd like to congratulate Argentina on winning this year's FIFA World Cup. #AlHattabHolding #Qatar2022 #20قطر 22

Robot Exclusion Principle

In []:

```
import urllib.robotparser as urobot
import urllib.request
from bs4 import BeautifulSoup

url = "https://vit.ac.in/"
rp = urobot.RobotFileParser()
rp.set_url(url + "/robots.txt")
rp.read()
if rp.can_fetch("*", url):
    print()
else:
    print("cannot scrap")
```

None