

snap-papers-web-scrap

April 3, 2024

1 Research Papers Web Scraping from ACM Digital Library

1.1 Team Snap Papers

This notebook explains the webscraping of research papers from the ACM Library depending on the members of Humanities & Engineering Special Interest Group

Import the necessary libraries

```
[1]: import requests
from bs4 import BeautifulSoup
import re
import pandas as pd
import time
```

Define the Profile ID of the author:

```
[2]: profile_id = '81321495239'
```

The response is get through the method requests.get

```
[3]: headers = {'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/
↳537.36 (KHTML, like Gecko) Chrome/91.0.4472.114 Safari/537.36'}
url = 'https://dl.acm.org/profile/{}/publications?
↳Role=author&startPage=0&pageSize=50'.format(profile_id)

response = requests.get(url, headers=headers)
```

Generate the Paper Structure object

```
[4]: paper_doc = BeautifulSoup(response.text, 'html.parser')
```

Get all the information

```
[5]: paper_names = []
for paper in paper_doc.find_all("li", {"class" : "issue-item-container"}):
    paper_type = ''
    paper_title = ''
    author_section = []
    abstract = ''
```

```

details = ''
citations = '0'
metrics = '0'
authors_list = []
authors_list_string = ''
url = ''

if paper.find('div', {'class' : 'issue-heading'}) is not None:
    paper_type = paper.find('div', {'class' : 'issue-heading'}).text
    paper_title = paper.find('h5', {'class' : 'issue-item_title'}).text
    author_section = paper.find("ul", {"title" : "list of authors"})
    abstract = paper.find("div", {"class" : "issue-item_abstract"}).find("p").
↪text
    if paper.find("span", {"class" : "citation"}) is not None:
        citations = paper.find("span", {"class" : "citation"}).find("span").text
    if paper.find("span", {"class" : "metric"}) is not None:
        metrics = paper.find("span", {"class" : "metric"}).find("span").text
    details = paper.find("div", {"class" : "issue-item_detail'}).text
    url = "https://dl.acm.org"+paper.find("h5", {"class" :
↪"issue-item_title"}).find("a").get("href")

print(paper_type)
print(paper_title)
for author in author_section.find_all("a"):
    authors_list.append(author.get("title"))
authors_list_string = ' | '.join(authors_list)
print(authors_list_string)
print(details)
print(abstract)
print(citations)
print(metrics)
print(url)

print("-----")

```

Doctoral Theses

Knowledge Management Using SpiCE

Timothy Joseph Maciag | Hepting, Daryl | Arbuthnott, Katherine | Delbaere,
Marjorie

The idea of Knowledge Management (KM) is continually evolving. A traditional and popular idea of KM is one that emphasizes the activity of transforming data to information, and information to knowledge. Another popular idea of KM emphasizes the ...

0

0

<https://dl.acm.org/doi/book/10.5555/AAI28140953>

Supporting Sustainable Decision-Making: Evaluation of Previous Support Tools
with New Designs

Timothy Maciag

Revision with unchanged content. The quality of the natural environment has become one of the primary concerns in present society. However, very little has been done to illuminate the various connections between our household purchases and the effect ...

0

0

<https://dl.acm.org/doi/book/10.5555/2378550>

Article

Web-Based Support of Crop Selection for Climate Adaptation

Daryl H. Hepting | Timothy Maciag | Harvey Hill

HICSS '12: Proceedings of the 2012 45th Hawaii International Conference on System Sciences•January 2012, pp 1227-1236•

<https://doi.org/10.1109/HICSS.2012.640>

Farmers must now consider climate adaptation amongst other variables when they select crops for the coming year. A changing climate means traditional crop choices may not perform well. Yet, it may be difficult to trust recommendations about new crop ...

0

0

<https://dl.acm.org/doi/10.1109/HICSS.2012.640>

Article

Constructing Collaborative Online Communities for Visualizing Spimes

Timothy Maciag | Daryl H. Hepting

WI-IAT '10: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03•August 2010, pp 436-439• <https://doi.org/10.1109/WI-IAT.2010.297>

The digital age has brought about new platforms for collaboration which have provided interesting and effective ways of enabling people to engage in a wide variety of socially-driven activities. One only needs to observe the many free/libre open source ...

0

0

<https://dl.acm.org/doi/10.1109/WI-IAT.2010.297>

Article

Classification of facial photograph sorting performance based on verbal descriptions

Daryl H. Hepting | Richard Spring | Timothy Maciag | Katherine Arbuthnott | Dominik Ślęzak

RSCTC'10: Proceedings of the 7th international conference on Rough sets and

current trends in computing•June 2010, pp 570-579

Eyewitness identification remains an important element in judicial proceedings. It is very convincing, yet it is not very accurate. To better understand eyewitness identification, we began by examining how people understand similarity. This paper ...

1

0

<https://dl.acm.org/doi/10.5555/1876210.1876282>

Article

A Rough Sets Approach for Personalized Support of Face Recognition

Daryl H. Hepting | Timothy Maciag | Richard Spring | Katherine Arbuthnott | Dominik Ślęzak

RSFDGrC '09: Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing•December 2009, pp 201-208•

https://doi.org/10.1007/978-3-642-10646-0_24

The activity of facial recognition is routine for most people; yet describing the process of recognition, or describing a face to be recognized reveals a great deal of complexity inherent in the activity. Eyewitness identification remains an important ...

2

0

https://dl.acm.org/doi/10.1007/978-3-642-10646-0_24

review-article

Creating synergy between usability courses and open source software projects

Daryl H. Hepting | Lijuan Peng | Timothy J. Maciag | David Gerhard | Brien Maguire

ACM SIGCSE Bulletin, Volume 40, Issue 2•June 2008, pp 120-123 •

<https://doi.org/10.1145/1383602.1383649>

In this paper, we discuss our experience in offering a usability course with projects taken from an active open source software development project. We describe what was done in the class inside the larger context of the usability of open source ...

12

234

<https://dl.acm.org/doi/10.1145/1383602.1383649>

Article

Evaluation of a Dominance-Based Rough Set Approach to Interface Design

Timothy Maciag | Daryl H. Hepting | Robert J. Hilderman | Dominik Slezak

FBIT '07: Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies•October 2007, pp 409-416•

<https://doi.org/10.1109/FBIT.2007.54>

This paper explores refinements to methods used in a procedure being developed by the authors to personalize user interfaces for online shopping support tools. In the authors' original procedure, classical methods in rough set theory are used in ...

1

0

<https://dl.acm.org/doi/10.1109/FBIT.2007.54>

Article

Mining associations for interface design

Timothy Maciag | Daryl H. Hepting | Dominik Ślezak | Robert J. Hilderman

RSKT'07: Proceedings of the 2nd international conference on Rough sets and knowledge technology•May 2007, pp 109-117

Consumer research has indicated that consumers use compensatory and non-compensatory decision strategies when formulating their purchasing decisions. Compensatory decision-making strategies are used when the consumer fully rationalizes their decision ...

0

3

<https://dl.acm.org/doi/10.5555/1767771.1767787>

Article

Consumer Modelling in Support of Interface Design

Timothy J. Maciag | Dominik Slezak | Daryl H. Hepting

ICHIT '06: Proceedings of the 2006 International Conference on Hybrid Information Technology - Volume 02•November 2006, pp 153-160

There is significant interest in developing new methods to design more effective user interfaces for decision support tools in online shopping environments. Many online companies have already begun to provide their consumers with enhanced user interface ...

1

0

<https://dl.acm.org/doi/10.5555/1193209.1193513>

Prepare the dictionary of information

```
[6]: paper_repos_dict = {
    'Paper Title' : [],
    'Publication Type' : [],
    'Authors' : [],
    'Abstract' : [],
    'Publication' : [],
    'Citations' : [],
    'Metrics' : [],
    'URL' : [] }

# adding information in repository
def add_in_paper_repo(paper_title, paper_type, authors_list_string, abstract,
    ↪details, citations, metrics, url):
    paper_repos_dict['Paper Title'] = paper_title
    paper_repos_dict['Publication Type'] = paper_type
```

```

paper_repos_dict['Authors'] = authors_list_string
paper_repos_dict['Abstract'] = abstract
paper_repos_dict['Publication'] = details
paper_repos_dict['Citations'] = citations
paper_repos_dict['Metrics'] = metrics
paper_repos_dict['URL'] = url

return paper_repos_dict

```

Scrap the data and generating CSV file

```

[7]: paper_names = []
data = []
for paper in paper_doc.find_all("li", {"class" : "issue-item-container"}):
    paper_type = ''
    paper_title = ''
    author_section = []
    abstract = ''
    details = ''
    citations = '0'
    metrics = '0'
    authors_list = []
    authors_list_string = ''
    url = ''

    if paper.find('div', {'class' : 'issue-heading'}) is not None:
        paper_type = paper.find('div', {'class' : 'issue-heading'}).text
        paper_title = paper.find('h5', {'class' : 'issue-item_title'}).text
        author_section = paper.find("ul", {"title" : "list of authors"})
        abstract = paper.find("div", {"class" : "issue-item_abstract"}).find("p").
↪text
        if paper.find("span", {"class" : "citation"}) is not None:
            citations = paper.find("span", {"class" : "citation"}).find("span").text
        if paper.find("span", {"class" : "metric"}) is not None:
            metrics = paper.find("span", {"class" : "metric"}).find("span").text
        details = paper.find("div", {"class" : "issue-item_detail"}).text
        url = "https://dl.acm.org"+paper.find("h5", {"class" :
↪"issue-item_title"}).find("a").get("href")

    for author in author_section.find_all("a"):
        authors_list.append(author.get("title"))
    authors_list_string = ' | '.join(authors_list)

    print(add_in_paper_repo(paper_title, paper_type, authors_list_string,
↪abstract, details, citations, metrics, url))

    print("-----")

```

```

data.append(add_in_paper_repo(paper_title, paper_type, authors_list_string,
↪abstract, details, citations, metrics, url).copy())

df = pd.DataFrame(data)
df.to_csv("./data.csv", sep='#', encoding='utf-8')

```

```

{'Paper Title': 'Knowledge Management Using SpiCE', 'Publication Type':
'Doctoral Theses', 'Authors': 'Timothy Joseph Maciag | Hepting, Daryl |
Arbuthnott, Katherine | Delbaere, Marjorie', 'Abstract': 'The idea of Knowledge
Management (KM) is continually evolving. A traditional and popular idea of KM is
one that emphasizes the activity of transforming data to information, and
information to knowledge. Another popular idea of KM emphasizes the ...',
'Publication': '', 'Citations': '0', 'Metrics': '0', 'URL':
'https://dl.acm.org/doi/book/10.5555/AAI28140953'}
-----

```

```

{'Paper Title': 'Supporting Sustainable Decision-Making: Evaluation of Previous
Support Tools with New Designs', 'Publication Type': '', 'Authors': 'Timothy
Maciag', 'Abstract': 'Revision with unchanged content. The quality of the
natural environment has become one of the primary concerns in present society.
However, very little has been done to illuminate the various connections between
our household purchases and the effect ...', 'Publication': '', 'Citations':
'0', 'Metrics': '0', 'URL': 'https://dl.acm.org/doi/book/10.5555/2378550'}
-----

```

```

{'Paper Title': 'Web-Based Support of Crop Selection for Climate Adaptation',
'Publication Type': 'Article', 'Authors': 'Daryl H. Hepting | Timothy Maciag |
Harvey Hill', 'Abstract': 'Farmers must now consider climate adaptation amongst
other variables when they select crops for the coming year. A changing climate
means traditional crop choices may not perform well. Yet, it may be difficult to
trust recommendations about new crop ...', 'Publication': "HICSS '12:
Proceedings of the 2012 45th Hawaii International Conference on System
Sciences•January 2012, pp 1227-1236• https://doi.org/10.1109/HICSS.2012.640",
'Citations': '0', 'Metrics': '0', 'URL':
'https://dl.acm.org/doi/10.1109/HICSS.2012.640'}
-----

```

```

{'Paper Title': 'Constructing Collaborative Online Communities for Visualizing
Spimes', 'Publication Type': 'Article', 'Authors': 'Timothy Maciag | Daryl H.
Hepting', 'Abstract': 'The digital age has brought about new platforms for
collaboration which have provided interesting and effective ways of enabling
people to engage in a wide variety of socially-driven activities. One only needs
to observe the many free/libre open source ...', 'Publication': "WI-IAT '10:
Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web
Intelligence and Intelligent Agent Technology - Volume 03•August 2010, pp
436-439• https://doi.org/10.1109/WI-IAT.2010.297", 'Citations': '0', 'Metrics':
'0', 'URL': 'https://dl.acm.org/doi/10.1109/WI-IAT.2010.297'}
-----

```

```

{'Paper Title': 'Classification of facial photograph sorting performance based

```

on verbal descriptions', 'Publication Type': 'Article', 'Authors': 'Daryl H. Hepting | Richard Spring | Timothy Maciag | Katherine Arbuthnott | Dominik Ślęzak', 'Abstract': 'Eyewitness identification remains an important element in judicial proceedings. It is very convincing, yet it is not very accurate. To better understand eyewitness identification, we began by examining how people understand similarity. This paper ...', 'Publication': 'RSCTC'10: Proceedings of the 7th international conference on Rough sets and current trends in computing•June 2010, pp 570-579', 'Citations': '1', 'Metrics': '0', 'URL': 'https://dl.acm.org/doi/10.5555/1876210.1876282'}

{'Paper Title': 'A Rough Sets Approach for Personalized Support of Face Recognition', 'Publication Type': 'Article', 'Authors': 'Daryl H. Hepting | Timothy Maciag | Richard Spring | Katherine Arbuthnott | Dominik Ślęzak', 'Abstract': 'The activity of facial recognition is routine for most people; yet describing the process of recognition, or describing a face to be recognized reveals a great deal of complexity inherent in the activity. Eyewitness identification remains an important ...', 'Publication': 'RSFDGrC '09: Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing•December 2009, pp 201-208• https://doi.org/10.1007/978-3-642-10646-0_24', 'Citations': '2', 'Metrics': '0', 'URL': 'https://dl.acm.org/doi/10.1007/978-3-642-10646-0_24'}

{'Paper Title': 'Creating synergy between usability courses and open source software projects', 'Publication Type': 'review-article', 'Authors': 'Daryl H. Hepting | Lijuan Peng | Timothy J. Maciag | David Gerhard | Brien Maguire', 'Abstract': 'In this paper, we discuss our experience in offering a usability course with projects taken from an active open source software development project. We describe what was done in the class inside the larger context of the usability of open source ...', 'Publication': 'ACM SIGCSE Bulletin, Volume 40, Issue 2•June 2008, pp 120-123 • https://doi.org/10.1145/1383602.1383649', 'Citations': '12', 'Metrics': '234', 'URL': 'https://dl.acm.org/doi/10.1145/1383602.1383649'}

{'Paper Title': 'Evaluation of a Dominance-Based Rough Set Approach to Interface Design', 'Publication Type': 'Article', 'Authors': 'Timothy Maciag | Daryl H. Hepting | Robert J. Hilderman | Dominik Slezak', 'Abstract': 'This paper explores refinements to methods used in a procedure being developed by the authors to personalize user interfaces for online shopping support tools. In the authors' original procedure, classical methods in rough set theory are used in ...', 'Publication': 'FBIT '07: Proceedings of the 2007 Frontiers in the Convergence of Bioscience and Information Technologies•October 2007, pp 409-416• https://doi.org/10.1109/FBIT.2007.54', 'Citations': '1', 'Metrics': '0', 'URL': 'https://dl.acm.org/doi/10.1109/FBIT.2007.54'}

{'Paper Title': 'Mining associations for interface design', 'Publication Type': 'Article', 'Authors': 'Timothy Maciag | Daryl H. Hepting | Dominik Ślęzak | Robert J. Hilderman', 'Abstract': 'Consumer research has indicated that consumers use compensatory and non-compensatory decision strategies when


```
formulating their purchasing decisions. Compensatory decision-making strategies
are used when the consumer fully rationalizes their decision ...',
'Publication': "RSKT'07: Proceedings of the 2nd international conference on
Rough sets and knowledge technology•May 2007, pp 109-117", 'Citations': '0',
'Metrics': '3', 'URL': 'https://dl.acm.org/doi/10.5555/1767771.1767787'}
-----
```

```
{'Paper Title': 'Consumer Modelling in Support of Interface Design',
'Publication Type': 'Article', 'Authors': 'Timothy J. Maciag | Dominik Slezak |
Daryl H. Hepting', 'Abstract': 'There is significant interest in developing new
methods to design more effective user interfaces for decision support tools in
online shopping environments. Many online companies have already begun to
provide their consumers with enhanced user interface ...', 'Publication': "ICHIT
'06: Proceedings of the 2006 International Conference on Hybrid Information
Technology - Volume 02•November 2006, pp 153-160", 'Citations': '1', 'Metrics':
'0', 'URL': 'https://dl.acm.org/doi/10.5555/1193209.1193513'}
-----
```

The reference of the previous code was an article made by Nandini Saini in Medium

Saini, N. (2021). Scraping Information of Research Papers on Google Scholar using Python. Medium. Web Article. <https://medium.com/@nandinisaini021/scraping-publications-of-aerial-image-research-papers-on-google-scholar-using-python-a0dee9744728>