

- Online First
- Current Issue
- Archive
- Most Downloaded
- Most Accessed
- Most Cited



Elementary Education Online

Current Issue : 2023, Vol: 22, Issue: 1

ISSN : 1305-3515

ISSUES : 6 per year

PUBLISHER : Elementary Education Online

Elementary Education Online (EEO) aims to provide researchers with scholarly discourses, theories, research methods, and findings. Therefore, the journal accepts manuscripts related to all aspects of Elementary Education. It is also keen to help make connections among researchers.

Elementary Education Online (EEO) is intended to contribute to the field with academic research outcomes and disseminate knowledge about elementary education, including theories, quality of instruction, instructional planning, design and technology from a disciplinary and/or interdisciplinary holistic approach, but it is not limited with these subjects.

Elementary Education Online (EEO) aims at:

- Disseminating research findings, either in qualitative or quantitative inquiry;
- Proposing new discussion directions and issues through literature reviews
- Emphasizing creative solutions to the existing problems and issues in the field.
- Contributing to professional development by discussing research findings in a systematic way.

Dear Authors,

Please note that journal has only two email id's i.e
editor@ilkogretim-online.org , editor.eeo.online@gmail.com

Please submit your manuscripts to these emails only.
 Kindly beware of fake email id's and scammers.

Thank you for your cooperation,
 Team EEO.

[← Previous Issue](#)

EEO. Year: 2023, Volume: 22

Research Article

1. **The Importance Of Blended Learning In Teaching The Second Language To Non-Native Speakers In Palestinian Universities**
Enes Khlaf Fahed Khlaf
 EEO. 2023; 22(1): 1-11
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.01
2. **Ecological and Environmental Issues: A Study of Religious Perspective**
Firdose Ahmad Mir, Dr. Mohd Arif Hussain Bhat
 EEO. 2023; 22(1): 12-19
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.02
3. **Marital Adjustment In The Light Of Some Variables**
Dr. Keltoum Aicha Saji, Dr. Kheira Kouider
 EEO. 2023; 22(1): 20-29
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.03
4. **Psychological Distress, Work Satisfaction And Turnover Intention Among Front Line Nurses Working During Covid-19 In Various Hospitals And Isolation Centres Of M.P**
Prof. Neetu Bhadorria, Ms. Abhilasha Francis, Ms. Anju Singh, Mr. Ramesh Singh, Mr. Abhishek Singh, Ms. Arvi
 EEO. 2023; 22(1): 30-37
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.04
5. **Islamic Conquests In Transoxiana**
Dr. Azar Lila, Dr. Rafaa Reda, Dr. Guebli Abdelah
 EEO. 2023; 22(1): 38-49
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.05
6. **The Role Of Radio Stations In The Diffusion Of Nazi Propaganda In The Maghreb During The Second World War 1939-1945**
Nadia Zerrouk
 EEO. 2023; 22(1): 50-52
 » Abstract » PDF » doi: 10.17051/ilkonline.2023.01.06

eJM Submit Article

[AUTHOR LOGIN](#)

[REVIEWER LOGIN](#)

Indexed & Abstracted

Indexing

İlköğretim Online (İÖÖ) - Elementary Education Online (EEO) is indexed in:

- [Scopus](#) (2013)
- [Dimensions](#) (2017)
- [DOAJ](#) (2019)
- [IndexCopernicus](#) (2010)
- [AERA](#) (2013)
- [UDL-EDGE](#) (2018)

ABOUT JOURNAL

- [About Publisher](#)
- [Contact Us](#)
- [Editorial & Peer Review Process](#)

POLICIES

- [Privacy Policy](#)
- [Editorial Policy](#)
- [Author's Rights and Obligations](#)
- [Conflict of Interest Policy](#)
- [Plagiarism Policy](#)
- [Advertising Policy](#)

STATEMENTS

- [Terms of Use](#)
- [Author's Rights and Obligations](#)
- [Publication Ethics and Publication Malpractice Statement](#)
- [Protection of Research Participants \(Statement On Human And Animal Rights\)](#)
- [License Information](#)



- Online First
- Current Issue
- Archive
- Most Downloaded
- Most Accessed
- Most Cited

2023

» [Volume 22, Issue 1](#)
pp. 1-50

2022

- » [Volume 21, Issue 4](#)
pp. 1-90
- » [Volume 21, Issue 3](#)
pp. 1-317
- » [Volume 21, Issue 2](#)
pp. 1-354
- » [Volume 21, Issue 1](#)
pp. 1-542

2021

- » [Volume 20, Issue 6](#)
pp. 1-4660
- » [Volume 20, Issue 5](#)
pp. 1-8549
- » [Volume 20, Issue 4](#)
pp. 1-4701
- » [Volume 20, Issue 3](#)
pp. 1-4518
- » [Volume 20, Issue 2](#)
pp. 1-3774
- » [Volume 20, Issue 1](#)
pp. 553-8366

2020

- » [Volume 20, Issue 3](#)
pp. 4099-4099
- » [Volume 19, Issue 4](#)
pp. 1-8214
- » [Volume 19, Issue 3](#)
pp. 1-5179
- » [Volume 19, Issue 2](#)
pp. 1-2299
- » [Volume 19, Issue 1](#)
pp. 299-1144

2019

- » [Volume 18, Issue 4](#)
pp. 16-2428
- » [Volume 18, Issue 3](#)
pp. 8-1514
- » [Volume 18, Issue 2](#)
pp. 3-1034
- » [Volume 18, Issue 1](#)
pp. 1-610

2018

- » [Volume 17, Issue 4](#)
pp. 1-2314
- » [Volume 17, Issue 3](#)
pp. 1-1863
- » [Volume 17, Issue 2](#)
pp. 1-1165
- » [Volume 17, Issue 1](#)
pp. 1-1437

2017

- » [Volume 16, Issue 4](#)
pp. 1-1947
- » [Volume 16, Issue 3](#)
pp. 900-1343
- » [Volume 16, Issue 2](#)
pp. 394-887
- » [Volume 16, Issue 1](#)
pp. 1-401

2016

- » [Volume 15, Issue 4](#)
pp. 1-1436
- » [Volume 15, Issue 3](#)
pp. 723-1066
- » [Volume 15, Issue 2](#)



[AUTHOR LOGIN](#)

[REVIEWER LOGIN](#)

**Indexed
&
Abstracted**

Indexing

İÖE Elementary Education Online (İÖE - Elementary Education Online (EEO)) is indexed in:

- [Scopus](#) (2013)
- [Dimensions](#) (2017)
- [DOAJ](#) (2019)
- [IndexCopernicus](#) (2010)
- [AERA](#) (2013)
- [UDL-EDGE](#) (2018)

- pp. 1-723
» **Volume 15, Issue 1**
pp. 1-286

2015

- » **Volume 14, Issue 4**
pp. 1159-1450
» **Volume 14, Issue 3**
pp. 1-1241
» **Volume 14, Issue 2**
pp. 364-787
» **Volume 14, Issue 1**
pp. 1-459

2014

- » **Volume 13, Issue 4**
pp. 1135-1542
» **Volume 13, Issue 3**
pp. 726-1122
» **Volume 13, Issue 2**
pp. 2-710
» **Volume 13, Issue 1**
pp. 1-319
» **Volume 1, Issue 0**
pp. -

2013

- » **Volume 12, Issue 4**
pp. 926-1212
» **Volume 12, Issue 3**
pp. 1-912
» **Volume 12, Issue 2**
pp. 1-614
» **Volume 12, Issue 1**
pp. 1-308

2012

- » **Volume 11, Issue 4**
pp. 856-1121
» **Volume 11, Issue 3**
pp. 702-1093
» **Volume 11, Issue 2**
pp. 1-689
» **Volume 11, Issue 1**
pp. 1-629

2011

- » **Volume 10, Issue 3**
pp. 1-1208
» **Volume 10, Issue 2**
pp. 620-1138
» **Volume 10, Issue 1**
pp. 1-378
» **Volume 1, Issue 1**
pp. 61-61

2010

- » **Volume 9, Issue 3**
pp. 1-1080
» **Volume 9, Issue 2**
pp. 1-816
» **Volume 9, Issue 1**
pp. 1-424

2009

- » **Volume 8, Issue 3**
pp. 1-978
» **Volume 8, Issue 2**
pp. 1-605
» **Volume 8, Issue 1**
pp. 1-268

2008

- » **Volume 7, Issue 3**
pp. 1-814
» **Volume 7, Issue 2**
pp. 232-537
» **Volume 7, Issue 1**
pp. 1-203

2007

- » **Volume 6, Issue 3**
pp. 1-480
» **Volume 6, Issue 2**
pp. 1-313

» [Volume 6, Issue 1](#)
pp. 1-180

2006

» [Volume 5, Issue 2](#)
pp. 1-126
» [Volume 5, Issue 1](#)
pp. 82-85

2005

» [Volume 4, Issue 2](#)
pp. 1-55
» [Volume 4, Issue 1](#)
pp. 1-73

2004

» [Volume 3, Issue 2](#)
pp. 1-67
» [Volume 3, Issue 1](#)
pp. 2-28

2003

» [Volume 2, Issue 2](#)
pp. 2-64
» [Volume 2, Issue 1](#)
pp. 1-54

2002

» [Volume 1, Issue 2](#)
pp. 33-64
» [Volume 1, Issue 1](#)
pp. 2-29

No issue...

ABOUT JOURNAL

- [About Publisher](#)
- [Contact Us](#)
- [Editorial & Peer Review Process](#)

POLICIES

- [Privacy Policy](#)
- [Editorial Policy](#)
- [Author's Rights and Obligations](#)
- [Conflict of Interest Policy](#)
- [Plagiarism Policy](#)
- [Advertising Policy](#)

STATEMENTS

- [Terms of Use](#)
- [Author's Rights and Obligations](#)
- [Publication Ethics and Publication Malpractice Statement](#)
- [Protection of Research Participants \(Statement On Human And Animal Rights\)](#)
- [License Information](#)



- Online First
- Current Issue
- Archive
- Most Downloaded
- Most Accessed
- Most Cited

About Publisher

EEO an International Journal is established in the year 2002. People from various avenues of Education profession, who have come together in a single platform to redefine the structure of education profession in the country. EEO is peer reviewed online journal. The motto behind the journal is to help students, researchers and scientist worldwide to benefit from the high quality peer reviewed articles and to their high performing works in the entire arena of Education, Childhood Education. All information collected, as well as related requests, will be handled carefully and efficiently as possible in accordance with EEO standards for integrity and objectivity.



[AUTHOR LOGIN](#)

[REVIEWER LOGIN](#)

**Indexed
&
Abstracted**

Indexing

Elementary Education Online (EEO) is indexed in:

- Scopus (2013)
- Dimensions (2017)
- DOAJ (2019)
- IndexCopernicus (2010)
- AERA (2013)
- UDL-EDGE (2018)

ABOUT JOURNAL

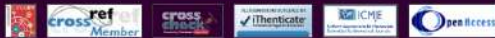
- [About Publisher](#)
- [Contact Us](#)
- [Editorial & Peer Review Process](#)

POLICIES

- [Privacy Policy](#)
- [Editorial Policy](#)
- [Author's Rights and Obligations](#)
- [Conflict of Interest Policy](#)
- [Plagiarism Policy](#)
- [Advertising Policy](#)

STATEMENTS

- [Terms of Use](#)
- [Author's Rights and Obligations](#)
- [Publication Ethics and Publication Malpractice Statement](#)
- [Protection of Research Participants \(Statement On Human And Animal Rights\)](#)
- [License Information](#)



- Online First
- Current Issue
- Archive
- Most Downloaded
- Most Accessed
- Most Cited

Editorial & Peer Review Process

Peer Review Policy

İİÖ Elementary Education Online (EEO) - Elementary Education Online (EEO) operates a double-blind review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final.

İİÖ Elementary Education Online (EEO) - Elementary Education Online (EEO) uses double-blind review, which means the identities of the authors are concealed from the reviewers, and vice versa. To facilitate this, please include the following separately:

Title page (with author details): This should include the title, authors' names, affiliations, acknowledgements and any Declaration of Interest statement, and a complete address for the corresponding author including an e-mail address.

Blinded manuscript (no author details): The main body of the paper (including the references, figures, tables and any acknowledgements) should not include any identifying information, such as the authors' names or affiliations.

Manuscript Submission

İİÖ Elementary Education Online (EEO) - ELEMENTARY EDUCATION ONLINE (EEO) publishes papers in English. Authors are requested to prepare their papers using the Journal's [template](#) and to submit their papers and other required files through <http://ilkogretim-online.org>. Submitted files must not include any information about authors, institutions, or organizations. Submitted papers must include both the Turkish version and the English version of the study in a single file. Manuscripts from authors of other nations will be evaluated on the English version during the review process and the Turkish version will be required after acceptance.

Editorial Review

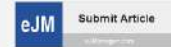
All manuscripts submitted for publication are received by the editorial office. Manuscripts will be checked for plagiarism and examined all files including cover letter, text files, graphs, charts and figures. If the files meet the technical and formatting requirements of the İİÖ Elementary Education Online (EEO), authors are sent an acknowledgement of the manuscript along with a Manuscript ID number. Authors are requested to quote the manuscript ID number in all correspondences with the journal. If the manuscript files do not meet the technical requirements of İİÖ Elementary Education Online (EEO), the manuscript are declined, and authors are sent detailed instructions about the changes required in the manuscript. This process lasts approximately a week.

Following formatting check, the manuscripts are passed on the Editor for a first evaluation. The editor reviews the manuscript and returns the manuscript along with the editorial comments. The editorial process gives the editors the authority to reject any manuscript if it does not meet the scientific publishing standards or is out of scope of the journal. Then, the manuscripts are sent to two or three reviewers. The manuscript is usually sent to two to three expert peer reviewers (double-blind). In this process the peer reviewer's identities are kept confidential and not revealed to the authors. In the same manner the author's identities are not revealed to the peer reviewers. This ensures a fair and unbiased review of every manuscript. This phase lasts approximately two weeks. Each reviewer has one month to complete the review and 15 additional days are given if requested. If the review is not completed within this time, a new reviewer is appointed. The review process lasts approximately two months.

Decisions are announced to the authors following the review process. If revisions are advised by the reviewers, authors are given one month to revise their manuscripts. Revised manuscripts are sent to the reviewers for a second check. This step lasts approximately one month.

If the authors feel that they are not satisfied with the editorial decision about publication of the manuscript, authors may appeal the decision.

Upon the reviewers' final decision, the editors check the submission process and give their final decision on the submission. This step lasts around two weeks. Accepted manuscripts go through redaction process first. The manuscripts are revised grammar, spelling, and style checking. This step lasts around one month. Submissions completing all the publication phases are given DOI numbers and published online upon the consent of the authors. Publication duration varies between 3 and 12 months.



AUTHOR LOGIN

REVIEWER LOGIN

Indexed
&
Abstracted

Indexing

İİÖ Elementary Education Online (EEO) is indexed in:

- Scopus (2013)
- Dimensions (2017)
- DOAJ (2019)
- IndexCopernicus (2010)
- AERA (2013)
- UDL-EDGE (2018)

ABOUT JOURNAL

- About Publisher
- Contact Us
- Editorial & Peer Review Process

POLICIES

- Privacy Policy
- Editorial Policy
- Author's Rights and Obligations
- Conflict of Interest Policy
- Plagiarism Policy
- Advertising Policy

STATEMENTS

- Terms of Use
- Author's Rights and Obligations
- Publication Ethics and Publication Malpractice Statement
- Protection of Research Participants (Statement On Human And Animal Rights)
- License Information



- Online First
- Current Issue
- Archive
- Most Downloaded
- Most Accessed
- Most Cited

« Previous Issue

Next Issue »

EEO, Year: 2021, Volume: 20, Issue: 3

Research Article

1. **How is the Awareness of Javanese Language Phonology of Elementary School Teacher Candidates?: Descriptive Qualitative Study**
Endang Sri Maruti, Bambang Yulianto, Suhartono, Nur Samsiyah
EEO, 2021; 20(3): 1-11
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.01
2. **Elements to be established by brand owners seeking redress in court for trademark infringement on social media sites**
Dr Anon Younes, Dr Hasan Mustafa, Dr Ahmed kamel
EEO, 2021; 20(3): 12-20
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.02
3. **“Context-Based Analysis in Interpreting Meaning and Identifying Sentence Structure in Fitzgerald’s The Great Gatsby Using Utterance Force Techniques”**
Hussein Hameed Ma’Ayauf, Abdul-Haq Abdul-Kareem Abdullah Al-Sahlani
EEO, 2021; 20(3): 21-30
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.03
4. **A Corpus-Based Analysis of the Most Frequent Idiomatic Expressions: The Semantic Differences Between British and American Idioms in L1D**
Salih Mahdi Adai Al-Mamoory, Abdul-Haq Abdul-Kareem Abdullah Al-Sahlani
EEO, 2021; 20(3): 31-42
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.04
5. **“Towards a Stylistic Account of Semantic/Pragmatic Deviation in Robert Browning’s the Last Ride Together & A Grammarian’s funeral Poem”**
Qasim Abbas Dhayef, Abdul-Haq Abdul-Kareem Al-Sahlani
EEO, 2021; 20(3): 43-52
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.05
6. **The Effectiveness Strategies for Students with Reading Challenges: New Frontiers**
Ayman Alnani, Siddiq Ahmed
EEO, 2021; 20(3): 73-82
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.07
7. **Psychological Effects of Quarantine During Covid-19: Students Attitude Towards Their Academics**
Prof. Dr. Abdul Waheed Mughal, Dr. Khawar Kaleem Raza, Dr. Obaid Ullah
EEO, 2021; 20(3): 94-103
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.09
8. **The Linguistic Functions and Educational Implications of Silence in Pinter’s The Dumb Waiter**
Asst. Prof. Dr. Wiaam A. Albayati, Asst. Prof. Dr. Aghar Hussain Ubied Abbas
EEO, 2021; 20(3): 104-110
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.10
9. **The Epidemic Prevention Approach in the Honest Narrations: A Contemporary Jurisprudential Study**
Nahlah J. Abdulhassan, Nada Suhail Abdul Hussein
EEO, 2021; 20(3): 111-117
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.11
10. **Science Fiction between Virtual Space Opera’s Adventure and the Alienation of Future Cities in Movie Poster Design**
Ahmed Osama Kahloun
EEO, 2021; 20(3): 118-127
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.12
11. **Relationship between Participation Motivation and Psychological Well-being of Marine Leisure Sports Participants**
Young Hoon Kwon, Chun-Ho Yang
EEO, 2021; 20(3): 119-1126
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.123
12. **Comparative Analysis of Banking Financial Performance: Evidence From Indonesia**
Indra Prasetyo, Woro Utari, Hadi Susanto, Rusdiyanto, Wildi Hidayat
EEO, 2021; 20(3): 152-164
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.15
13. **Meeting the Students’ Expectations: Evaluating the Implementation of English Language Teaching Curriculum**
Disavetriah Ahmad, Mardiana, Kaharuddin,
EEO, 2021; 20(3): 165-176
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.16
14. **The Effect of Social Economy Internet Shopping Mall Characteristics on Customer Reliability and Loyalty: Focusing on Ethical Purchase Intention and the Moderating Effect of the Residential Area**
Eun-Jeong Kim, Yen-Yoo You
EEO, 2021; 20(3): 187-1174
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.129
15. **The Effect Of Transformational Leadership, Job Satisfaction, And Organizational Commitments Toward Organizational Citizenship Behavior Of Indonesian Senior High School Teachers**
M. Shofiyuddin, Arif Budibharja, Casrameko, Nur Khoimah, Anand Syarifudin, Hadiyah, Ahmad Musyafak, Nasitaurul Janah, Ratna Wijayanti, Agus Paruwanto
EEO, 2021; 20(3): 177-185
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.17
16. **Identifying Cognitive Obstacle on the Topic of Definite Integral among Prospective Teacher**
Karrinah, Toto Nuzantra, Sudirman, Tjeng Darstel
EEO, 2021; 20(3): 186-192
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.18
17. **Training teachers to integrate technology: Drawing parallels between fluent (L2) and foreign language (EFL) users of English**
Mohammad Inran Khan
EEO, 2021; 20(3): 193-207
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.19
18. **Teacher Professional Development in Sindh Through Nacte: Analytical Study**
Ms Nusrat Ali, Dr Kamal Haider
EEO, 2021; 20(3): 208-215
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.20
19. **Impacts of multimedia gloss on Learning Reading Comprehension in ESL context**
Dr Habibullah Pathan, Sahib Khatoon, Dr Shamaila Aijaz Memon, Dr Muhammad Azeem Raheem, Dr Farina Dar, Faiza Mubshiq
EEO, 2021; 20(3): 216-225
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.22
20. **Implementation of Assessment for Learning and The Need for Teachers Refreshing Trainings.**
Dr. Shahzadi Seema, Dr. Wilayat Bibi, Dr. Waqar Uin Nisa Faizi
EEO, 2021; 20(3): 226-234
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.23
21. **A Journey of a Thousand Miles Begins with A Quantum Step: The Importance of Quantum Leadership to Promote Lifelong Learning in Organisations**
Soh Wan Geok, Mohamad Bin Bilal Ali
EEO, 2021; 20(3): 235-247
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.24
22. **Application of Dialectical Behavior Therapy in patients with opioid dependence- a pilot study**
Komal Sanoheti, Dr. Nandha Kumara Pajam S
EEO, 2021; 20(3): 248-261
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.25
23. **Impact of Interactive Pedagogies on Students’ Academic Achievement in Mathematics at Elementary School Level in Quetta City, Balochistan.**
Dr. Aha Asghar, Dr. Raza Gul, Dr. Mubshah Malik, Prof. Dr. Muhammad Yousof Sharjeel, Dr. Maroof Bin Rauf
EEO, 2021; 20(3): 262-270
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.26



AUTHOR LOGIN

REVIEWER LOGIN

Indexed & Abstracted

Indexing

Å°IköÄYretim Online (IOO) - Elementary Education Online (EEO) is indexed in:

- Scopus (2013)
- Dimensions (2017)
- DOAJ (2019)
- IndexCopernicus (2010)
- AERA (2013)
- UDL-EDGE (2018)

24. **Impact of E Learning on Students Performance and Expenditures Borne by Students and Virtual Degree Awarding Institutes in Pakistan**
Noem Akhtar, Nazia Aslam, Syed Qaiser Hussain
 EEO. 2021; 20(3): 271-283
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.27
25. **Teaching Vocabulary through CAVI: An Action Research Project**
Rifa Anjum, Dr Shumaila Memon, Dr Habibullah Pathan, Dania Wasi, Dr-Muhammad Arslan Raheem, Fatma Mushtaq
 EEO. 2021; 20(3): 284-297
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.28
26. **Rhetorical Analysis of Abu Al-Talib Al-Mutanabbi Selected Poems: The Case of Metaphor**
Badi' Ahmad Hasan Alazam
 EEO. 2021; 20(3): 298-303
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.29
27. **The degree to which faculty members at Imam Abdul Rahman bin Faisal University possess educational and social values from the students' viewpoint and its relationship to some variables.**
Sa'adiah Raneh, Darussahib, Ahmad Malmoud Elkhatim, Ammar Saad Al-Sha'ar, Sabrin Abdelaty Labib Abdelaty, Yumna Ahmed Atoum, Hana Youssef Falrakh, Hussein Mohamed Atoom, Lubna Abdullah abass almahdi
 EEO. 2021; 20(3): 304-316
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.30
28. **Studying the Effects of Human Resource Information System on Corporate Performance**
Laili T. Khris, Dyanan S. Sulawati, Aloned Alafandi, Norah Yusuf Alsaeed
 EEO. 2021; 20(3): 316-326
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.31
29. **The Innovative Role of the Islamic World to Counter the Phenomenon of Western "Islamophobia" during the Post-Corona Era**
Muhammad Umar Rizq Abbasi, Dr Hafiz Haris Saleem, Tariq Ejaz, Dr Sana Ullah, Dr Hafiz Muhammad Mushtaq, Dr Mohsin Akhbar
 EEO. 2021; 20(3): 326-340
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.32
30. **Effect of Corporate Governance, Information Asymmetry, And Firm Size on Earning Management**
Nurainun Bangun, S.E., Ak., M.M., CA., Dr. Rio Monoarfa, SE, M. ST Ak
 EEO. 2021; 20(3): 351-365
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.34

Original Article

31. **Integration of Local Wisdom Values of Seren Taun Traditional Ceremony in Social Science Lesson as Learning Model of Ethnopedagogy in SMPN 1 Nagrak District of Nagrak Regency of Sukabumi**
Endang Komara, Yudikusyadi, Neneng Meliyani
 EEO. 2021; 20(3): 366-375
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.35
32. **Integration of Local Wisdom Values of Seren Taun Traditional Ceremony in Social Science Lesson as Learning Model of Ethnopedagogy in SMPN 1 Nagrak District of Nagrak Regency of Sukabumi**
Endang Komara, Yudikusyadi, Neneng Meliyani
 EEO. 2021; 20(3): 366-375
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.35
33. **The Deliberation Of Pancasila Democracy Perspective In The Indonesian Constitutional System**
Megawati Megawati, Aboeri Absari
 EEO. 2021; 20(3): 385-393
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.37

Research Article

34. **Polygamy Marriage Law Without Court Permits**
Mustaqim, Agus Satory,
 EEO. 2021; 20(3): 394-401
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.38
35. **Politic and Election in Papua Indonesia**
Budi Setyanto, Jansal Wicoko, Mohammad Jamin, Isharyanto, Co-Promotor
 EEO. 2021; 20(3): 402-406
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.39
36. **Institution of The Judicial Commission of The Republic of Indonesia As an Ethical Court for Judges**
Jaja Ahmad Jayus, Taufiqurrahman Syahuri, Wahyu Nugroho
 EEO. 2021; 20(3): 407-416
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.40
37. **Environmental Settings in The Indonesian Constitution Perspective**
Taufiqurrahman Syahuri, Wahyu Nugroho, Ibnu Sina Chandranegara, Wensira Yunalidi
 EEO. 2021; 20(3): 417-423
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.41
38. **Analysis Study of Factors Affecting Students 'Digital Literacy Competency**
Ratna Sari Dewi, Fahrurrozi, Usawatun Hasanah, Muhammad Zuhri Dj
 EEO. 2021; 20(3): 424-431
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.42
-
39. **Implementation of Web Scrapping on News Sites Using the Supervised Learning Method**
Dedy Rahman Prehanto, Aries Dwi Indriyanti, I Gusti Lanang Eka Prismano, Gunanjar Setyo Permadi, Edwin Hari Agus Prasetyo,
 EEO. 2021; 20(3): 432-441
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.43
-
40. **The conditions of Mejaz in the Islamic era in Yemeni sources.**
Ghalib Mahmoud Hamza, Nasur Abid Albaker Mahmoud
 EEO. 2021; 20(3): 442-449
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.44
41. **Semantic employment of color in Syrian drama: Bab Al-hara movie as a model**
Dr. Adrian Kazem Ananias,
 EEO. 2021; 20(3): 450-461
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.45
42. **The Identity Problem in the Diaspora Movies**
Dr. Amman Ibrahim Mohammed Al-Yasri
 EEO. 2021; 20(3): 462-471
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.46
43. **Leonid Brezhnev and His Orientations Towards France**
Dr. Tariq Najm Abdul Wahid, Dr. Khalaf Obaid Hamood
 EEO. 2021; 20(3): 472-480
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.47
44. **Revitalization of Education for Children in Indonesian Families During the Covid-19 Pandemic**
Ahmad Muradi, M. Syahran Jatlani, Sufiyati HS, Lemal Suardi Wiekko, Hasni Noor
 EEO. 2021; 20(3): 481-490
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.48
45. **Acceptance Model Of E-Flipped Learning for Students In Jordan Schools**
Samar Mahmoud Mohammad Akho Zhayeh,
 EEO. 2021; 20(3): 491-500
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.49
46. **Open Access Publication Status of Educational Technology with The Bibliometrics Approach**
Fatkhul Anam, Andre Sapfianto, Mega Firdaus, Nur Asihah, Agung Purwono,
 EEO. 2021; 20(3): 501-509
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.50
47. **Influence of Website Quality, Innovation, and Brand Trust on Customer Satisfaction of Embedded Website in Game Application**
Hendhigo Wahyu, Legowo Nila
 EEO. 2021; 20(3): 510-520
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.51
48. **Reconstruction of Andragogical Principles in Porang Farming Community**
Rofik Jalal Rosyanafi, Djoko Saryono, Abdul Latief Bustami, Hardika,
 EEO. 2021; 20(3): 521-535
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.52
49. **The Influence of Soft Skills and Hard Skills on the Competence of Vocational High School Graduate Students in the Field of Building Construction**
Soeparno, Luthfiyah Nurfaela, Eko Hariadi, Supari, Suparji, Agus Wijayono,
 EEO. 2021; 20(3): 536-541
 » Abstract » PDF » doi: 10.17051/ikonline.2021.03.53
50. **Contextual Teaching And Learning (Ct) Learning Model Using The Module For Futsal Courses, Department Of Sports Trainee Education, Faculty Of Sports Science, Surabaya State University**

Mr. S Mahesh Reddy , Dr. Vijay Prakash Singh , Dr. Sudhā N Shaik
EEO. 2021; 20(3): 4474-4485
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.461

439. Reviewing Need And Application Of Artificial Intelligence In Business During Present Times
Sivakrishna Kulkarni , Dr. Harsh Lohtya
EEO. 2021; 20(3): 4486-4496
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.462

440. Evaluating Degree Of Anxiety Among Chess Players At Three Different Stages
Mr. Govindan Kutty M S, Dr. Minakshi Pathak
EEO. 2021; 20(3): 4496-4503
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.463

441. Exploring Recent Applications Of Conducting Polymers
Mr. Rajavarapu Sridhara Reddy , Dr. Sanjay Rathore
EEO. 2021; 20(3): 4504-4510
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.464

442. An Attempt To Examine Antibacterial Activity Of Different Organic Solvents Extract Of Marine Seaweeds
Mr. Dnyaneshwar Sahawke , Dr. Pushpendra Sharma
EEO. 2021; 20(3): 4511-4517
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.465

443. Identifying The Level Of Employability Skill Readiness Among Businessâ€™ Student
Syed Ahmed Moin Quadri , Dr. Rajesh Sharma
EEO. 2021; 20(3): 4518-4525
» Abstract » PDF » doi: 10.17051/ikonline.2021.03.466

ABOUT JOURNAL

- [About Publisher](#)
- [Contact Us](#)
- [Editorial & Peer Review Process](#)

POLICIES

- [Privacy Policy](#)
- [Editorial Policy](#)
- [Author's Rights and Obligations](#)
- [Conflict of Interest Policy](#)
- [Plagiarism Policy](#)
- [Advertising Policy](#)

STATEMENTS

- [Terms of Use](#)
- [Author's Rights and Obligations](#)
- [Publication Ethics and Publication Malpractice Statement](#)
- [Protection of Research Participants \(Statement On Human And Animal Rights\)](#)
- [License Information](#)





Implementation of Web Scraping on News Sites Using the Supervised Learning Method

Dedy Rahman Prehanto, *University of Surabaya, Dept. of Information System – Indonesia*

Aries Dwi Indriyanti, *University of Surabaya, Dept. of Information System – Indonesia*

I Gusti Lanang Eka Prisma, *University of Surabaya, Dept. of Information Technology – Indonesia*

Ginanjar Setyo Permadi, *University of Hasyim Asy'ari, Dept. of Informatics Management – Indonesia*

Edwin Hari Agus Prastyo, *University of Hasyim Asy'ari, Dept. of Information Technology – Indonesia*

Abstract. Indonesia is one of the highest internet users in the world, including in the penetration of information on the internet, online news media. But in general news sites not only display news information, but most sites also display other information such as advertisements and also forms of navigation that interfere with news site readers and interfere with reader's comfort, from these problems this study aims to implement web scraping techniques with supervised learning methods and analyzing the form of DOM tree and XPath news sites. The supervised learning approach method is the method used in this study, which is one of the methods of machine learning. By combining these web scraping techniques with supervised learning, the aim is to be able to implement and optimize web scraping techniques to gather news information from various sites. To do basic web scraping namely knowing DOM patterns, XPath structure as a data model or selector at each site. The results of research in the form of a web scrap application that can retrieve news site content without copy paste and the data is stored in a database and displayed to the user application form for the reader without any ads and navigation that disturb the reader.

Keywords: web scraping, supervised learning, XPath, DOM tree.

Received: 05.12.2020

Accepted: 10.01.2021

Published: 05.02.2021

INTRODUCTION

Indonesia is one of the highest internet users in the world. Based on the infographic survey agency for penetration & behavior of Indonesian internet users, it reaches 68% of the 264.18 million Indonesian population, many users penetrate information via the internet. Information is the most important thing in life, sources of information can be through news sites(Permadi et al., 2018). A news site is a website that provides various information from various news sources to be displayed to users(Permadi et al., 2019). However, in general, sites do not only display information from outside sources, sometimes they also display information on their own website such as advertisements and navigation that disturbs news site readers and disturbs the convenience of the readers, and also some people may still collect data on the website with copying one by one on the website, but if the website that you manage is a large site with thousands of data, of course the work takes a long time, so to solve this problem the researchers conducted one of the web content extraction techniques using machine learning(Dedy Rahman Prehanto et al., 2019).

In classification, supervised learning is carried out by conducting training (training) to form a model(D. R. Prehanto et al., 2019). Classifier (classification algorithm) will form a model that adapts according to the features that exist in the data(Mashuri et al., 2019). The purpose of applying the supervised learning method is to implement and optimize web scraping techniques, as well as make it easier for users to process information from the results of web scraping techniques used to gather news information from various sites(Xia et al., 2017). To do web scraping, the basic thing is to know the DOM tree structure on news sites(Iacus, 2015).

The researcher analyzed the DOM tree and Xpath patterns which were later used as materials in the training data(Niculescu-Mizil & Caruana, 2005). DOM tree is a standard html document structure to form a web page, while Xpath is a query language from the representation of the DOM resulting in a particular node(Hardt et al., 2016). Rizaldi's research states that the use of Xpath is better than the CSS selector, therefore this study uses Xpatht.

So this study takes the title Implementation of Web Scraping on News Sites Using the Supervised Learning Method. Below are some definitions related to this research.

METHODOLOGY

Web scraping

Web scraping is a process of data collecting through humans using a web browser, to handle the web that does not provide an API, web scraping requests data in HTML tags and then parses that data to extract the required information(Nylen & Wallisch, 2017). In the implementation, the use of web scraping includes programming techniques and technologies, such as data analysis and information security. Web scraping retrieves HTML data from a domain name, parses that data for target information, optionally stores target information, and moves to another page to repeat the scraping process(S.C.M. de S Sirisuriya, 2015).



Figure 1

Web Scraping

Web scraping describes the use of a program for extracting data from HTML files that exist on a particular site on the internet. Web scraping is the term used to describe the data extraction process on electronic files using a computer program(Ulbricht, 2020).

The next research was carried out by Rizaldi, explaining the crawling technique using Scrapy and Xpath, in this research XPath managed to meet the expected target. This is indicated by the production of a news corpus that has been divided into 3 news categories namely entertainment, community and culinary news(Khalil & Fakir, 2017).

Web scraping has several stages, the first step is to make scraping templates, explore the site navigation, automate navigation and extract information and save it(Mitchell, 2015).

Illustrating the steps for performing web scraping, starting with determining which sites will be used as scraping objects, then analyzing the page structure of the site to create a template(Aries Dwi Indriyanti et al., 2019). Then carry out exploration and navigation, the results of this process are then automated for extraction(A. D. Indriyanti et al., 2019). The final step is to save the scraping that has been done. The process of extracting information before it is saved can be done by utilizing the XPath method to find the desired information with a pattern according to a predefined template(Kistofer et al., 2019).



Figure 2

Web scraping steps

Content Extraction using Machine Learning

Content extraction using Machine Learning which is part of the branch of artificial intelligence. This includes the construction and study of training data systems as an introductory model, the essence of machine learning is related with representation and generalization. Generalization is a trait that the system will perform well on unseen data (Mutasim et al., 2018).

Supervised Learning

Supervised learning is a method for classifying each object in the data into several classes (Iacus, 2015). In supervised learning, each object in a data has features, namely the features that exist in each object. Each object in a data has the same number of features (Huber & Stuckenschmidt, 2020). Features are used as input to define a class on an object. In supervised learning, the class of each object is known (Huber & Stuckenschmidt, 2020). Therefore, the problem faced in supervised learning is how to map the objects into the right class using the features that each object has (Grandvalet & Bengio, 2005).

Classification in supervised learning is done by conducting training (training) to form a model. Classifier (classification algorithm) will form a model that adapts according to the features that exist in the data (Laine & Aila, 2017). The resulting model can be in the form of a tree, rule, or a function that can predict a class based on the features the data has (You et al., 2018)

Python

Python is a desktop or web-based programming language, a popular programming language used by many developers (Chollet, 2018). Python is one of the popular programming languages used by many developers. A survey according to the programming language website via www.tiobe.com, Python was ranked 4th in 2019 (GCM et al., 2016). Python can also be used for enterprises. Python is included in the programming language level, including high level language (Buduma & Locascio, 2017). Python is a programming language that can be used to build applications, be it desktop-based, web-based or mobile-based.

In python, there is a module used for the data extraction process so that the results obtained are faster and more practical. Python modules used in this study are (PF et al., 2016):

- a. Urllib is a module for handling request urls, the url you want to extract will be handled by urllib.
- b. BeautifulSoup is a module for parsing HTML. HTML parsing is a technique for separating text from HTML code tags on a website page to produce specific data

RESULTS AND DISCUSSION

In this study, using processing data derived from online news sites totaling 100 pages of news sites. List of news sites that extract web content in the table.

Table 1. list of news sites

| No | Name Web | Url |
|----|-------------------|-----------------------------------------------------------------------------------------|
| 1 | Media Indonesia | https://mediaindonesia.com |
| 2 | Kompas | https://www.kompas.com |
| 3 | Bisnis Indonesia | https://teknologi.bisnis.com |
| 4 | Pikiran Rakyat | https://www.pikiran-rakyat.com |
| 5 | Cek & Ricek | https://ceknricek.com |
| 6 | Siwalima | https://siwalimanews.com |
| 7 | Waspada | http://waspada.co.id |
| 8 | Analisa | https://analisadaily.com |
| 9 | Tribun Timur | https://makassar.tribunnews.com |
| 10 | Kedaulatan Rakyat | https://krjogja.com |
| 11 | Harian Jogja | https://jogjapolitan.harianjogja.com |
| 12 | Suara Merdeka | https://www.suaramerdeka.com |
| 13 | Solo Pos | https://www.solopos.com |
| 14 | Koran Sindo | https://jateng.sindonews.com |
| 15 | Sindo Weekly | https://sumeks.co |
| 16 | Sumatera Ekspres | http://www.sindoweekly.com |

- d. Admins can manage the target page of the news site to be scraped or extracted from the headlines and content.
- e. Users can read scraping results containing news headlines and news content in the newsscra mobile web application.
- f. Users can read news according to the categories provided and search for news according to the headline.

The results of the analysis of the DOM tree structure and site XPath

In retrieval of news content two XPath sections are taken the first XPath the title and the second XPath the contents, the XPath laying titles before content on the website has a different structure forms so that this analysis is carried out.

a. Media Indonesia

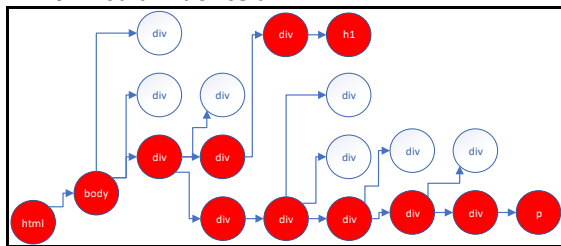
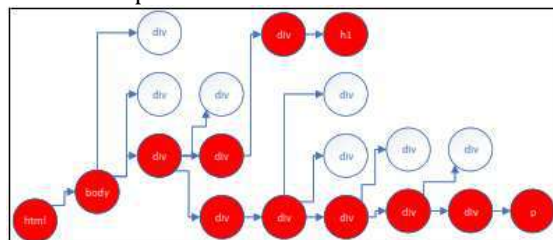


Figure 5
Dom tree media Indonesia

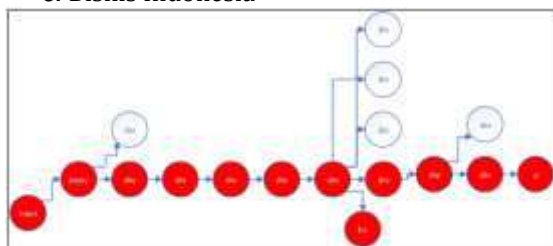
| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[3]/div[2]/div/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

b. Kompas



| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[3]/div[2]/div/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

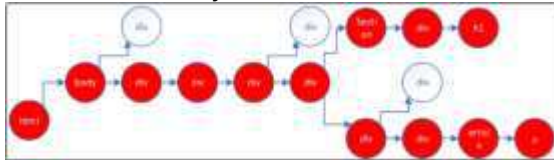
c. Bisnis Indonesia



| No | Target | XPath |
|----|--------|-------|
|----|--------|-------|

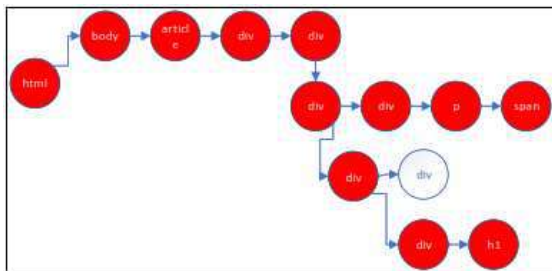
| | | |
|---|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[2]/div/div/div[1]/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

d. Pikiran Rakyat



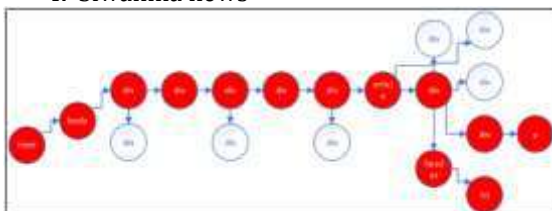
| No | Target | XPath |
|----|----------|----------------------------------------------------------|
| 1 | Headline | /html/body/div[2]/div/div/div[2]/section/div/h1 |
| 2 | Content | /html/body/div[2]/div/div/div[2]/div[1]/div[2]/article/p |

e. Cek & Ricek



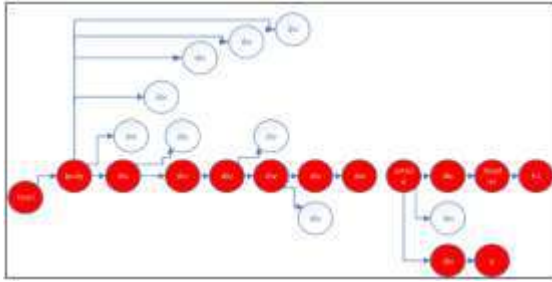
| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/article/div/div[1]/div[1]/div[2]/h1 |
| 2 | Content | /html/body/article/div/div[1]/div[1]/div/p/span/text() |

f. Siwalima news



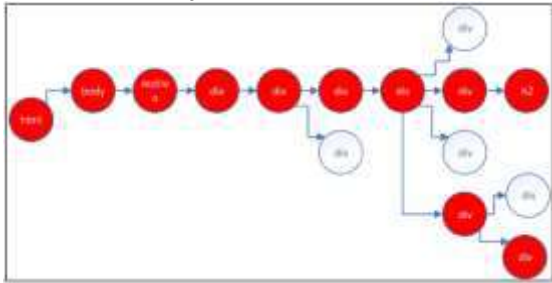
| No | Target | XPath |
|----|----------|------------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/div/div/div[1]/div[1]/article/div[2]/header/h1 |
| 2 | Content | /html/body/div[1]/div/div/div[1]/div[1]/article/div[2]/div[3]/p |

g. Waspada



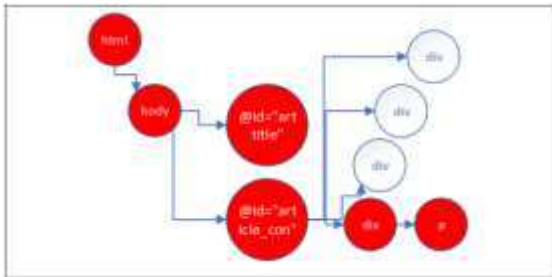
| No | Target | XPath |
|----|----------|-------------------------------------------------------------------------|
| 1 | Headline | /html/body/div[6]/div[2]/div/div[2]/div[1]/div/article/div[1]/header/h1 |
| 2 | Content | /html/body/div[6]/div[2]/div/div[2]/div[1]/div/article/div[3]/p |

h. Analisisaday



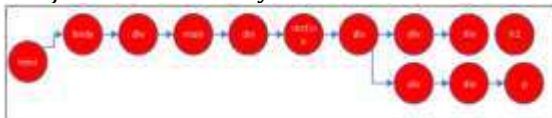
| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

i. Makasar Trimbum



| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

j. Kedaulatan Rakyat



| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

This target data page contains a list of site links that will be scrapped or retrieved from the news content through the XPath selector (training data) that has been provided previously. The admin clicks the button, the process will perform the html feature extraction and analyze the XPath matched with the XPath selector or the previous XPath model that has been provided if the same then the headline content and the news content will be saved if the XPath selector model is not the same as the news site that will be scraping the food data can retrieve the news content and the system will display an error such as Figure.



Figure 6. Target Site

This user page displays all selected news contents by pressing the read-further button and news search pages based on news categories as shown in the Figure 7.

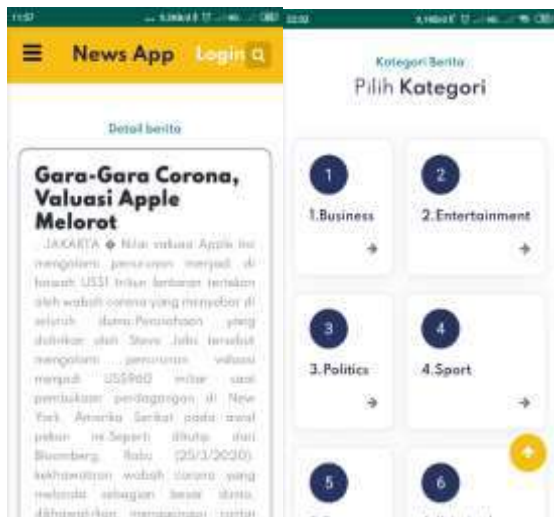


Figure 6. read the news and choose a category

CONCLUSION

Based on the results and previous discussion, the results of this study are in the form of DOM tree and XPath pattern analysis of the news sites studied, and implemented in the form of a web-based scrap news web application using the python programming language with the flask framework. The results obtained are knowing the pattern of DOM and XPath news sites. The XPath pattern always differs in a news site among others. The results of the extraction data using XPath were more complete. By using this supervised learning method, the more data on the XPath training model the better this application will be.

So when conducting web scraping, the site is no longer analyzed because it's already in the XPath model data.

REFERENCES

- Buduma, N., & Locascio, N. (2017). Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. / Nikhil Buduma ; with contributions by Nicholas Locascio. In Designing next-generation machine intelligence algorithms.
- Chollet, F. (2018). Deep Learning with Python. In Manning.
- GCM, G., Laterra, P., Aparicio, V., & Costa, J. (2016). Glyphosate retention in grassland riparian areas is reduced by the invasion of exotic trees. *Phyton*. <https://doi.org/10.32604/phyton.2016.85.108>
- Grandvalet, Y., & Bengio, Y. (2005). Semi-supervised learning by entropy minimization. *Advances in Neural Information Processing Systems*.
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*.
- Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2020.02.005>.
- Iacus, S. M. (2015). Automated Data Collection with R - A Practical Guide to Web Scraping and Text Mining. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v068.b03>.
- Indriyanti, A. D., Prehanto, D. R., Prisma, I. G. L. E. P., Soeryanto, Sujatmiko, B., & Fikandda, J. (2019). Simple Additive Weighting algorithm to aid administrator decision making of the underprivileged scholarship. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1402/6/066070>.
- Indriyanti, Aries Dwi, Prehanto, D. R., Permadi, G. S., Mashuri, C., & Vitadiar, T. Z. (2019). Using Fuzzy Time Series (FTS) and Linear Programming for Production Planning and Planting Pattern Scheduling Red Onion. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523007>
- .Khalil, S., & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*. <https://doi.org/10.1016/j.softx.2017.04.004>.
- Kistofer, T., Permadi, G. S., & Vitadiar, T. Z. (2019). Development of Digital System Learning Media Using Digital Learning System. <https://doi.org/10.2991/assehr.k.191217.030>.
- Laine, S., & Aila, T. (2017). Temporal ensembling for semi-supervised learning. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.
- Mashuri, C., Mujianto, A. H., Sucipto, H., Arsam, R. Y., & Permadi, G. S. (2019). Production Time Optimization using Campbell Dudek Smith (CDS) Algorithm for Production Scheduling. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523009>.
- Mitchell, R. (2015). Web Scraping with Python Collecting Data From The Modern Web. In O'Reilly.
- Mutasim, A. K., Tipu, R. S., Bashar, M. R., Islam, M. K., & Amin, M. A. (2018). Computational intelligence for pattern recognition in EEG signals. In *Studies in Computational Intelligence*. https://doi.org/10.1007/978-3-319-89629-8_11.
- Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*. <https://doi.org/10.1145/1102351.1102430>.
- Nylen, E. L., & Wallisch, P. (2017). Web Scraping. In *Neural Data Science*. <https://doi.org/10.1016/b978-0-12-804043-0.00010-6>.
- Permadi, G. S., Adi, K., & Gernowo, R. (2018). Application Mail Tracking Using RSA Algorithm As Security Data and HOT-Fit a Model for Evaluation System. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/20183111007>.
- Permadi, G. S., Vitadiar, T. Z., Kistofer, T., & Mujianto, A. H. (2019). The Decision Making Trial and Evaluation Laboratory (Dematel) and Analytic Network Process (ANP) for Learning Material Evaluation System. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523011>.
- PF, C., Scarpassa, J., Pretto-Giordano, L., Otaguiri, E., Yamada-Ogatta, S., Nakazato, G., Perugini, M., Moreira, I., & Vilas-Balãs, G. (2016). Antibacterial activity of avocado extracts (*Persea americana* Mill.) against *Streptococcus agalactiae*. *Phyton*. <https://doi.org/10.32604/phyton.2016.85.218>.
- Prehanto, D. R., Indriyanti, A. D., Nuryana, K. D., Soeryanto, S., & Mubarak, A. S. (2019). Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1402/6/066069>.
- Prehanto, Dedy Rahman, Indriyanti, A. D., Mashuri, C., & Permadi, G. S. (2019). Soil Moisture Prediction using Fuzzy Time Series and Moisture sensor Technology on Shallot Farming. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523002>.

- S.C.M. de S Sirisuriya. (2015). A Comparative Study on Web Scraping. 8th International Research Conference KDU.
- Ulbricht, L. (2020). Scraping the demos. Digitalization, web scraping and the democratic project. Democratization. <https://doi.org/10.1080/13510347.2020.1714595>.
- Xia, Y., Qin, T., Chen, W., Bian, J., Yu, N., & Liu, T. Y. (2017). Dual supervised learning. 34th International Conference on Machine Learning, ICML 2017.
- You, Z., Raich, R., Fern, X. Z., & Kim, J. (2018). Weakly Supervised Dictionary Learning. IEEE Transactions on Signal Processing. <https://doi.org/10.1109/TSP.2018.2807422>.



Implementation of Web Scraping on News Sites Using the Supervised Learning Method

Dedy Rahman Prehanto, *University of Surabaya, Dept. of Information System – Indonesia*

Aries Dwi Indriyanti, *University of Surabaya, Dept. of Information System – Indonesia*

I Gusti Lanang Eka Prisma, *University of Surabaya, Dept. of Information Technology – Indonesia*

Ginanjari Setyo Permadi, *University of Hasyim Asy'ari, Dept. of Informatics Management – Indonesia*

Edwin Hari Agus Prastyo, *University of Hasyim Asy'ari, Dept. of Information Technology – Indonesia*

Abstract. Indonesia is one of the highest internet users in the world, including in the penetration of information on the internet, online news media. But in general news sites not only display news information, but most sites also display other information such as advertisements and also forms of navigation that interfere with news site readers and interfere with reader's comfort, from these problems this study aims to implement web scraping techniques with supervised learning methods and analyzing the form of DOM tree and XPath news sites. The supervised learning approach method is the method used in this study, which is one of the methods of machine learning. By combining these web scraping techniques with supervised learning, the aim is to be able to implement and optimize web scraping techniques to gather news information from various sites. To do basic web scraping namely knowing DOM patterns, XPath structure as a data model or selector at each site. The results of research in the form of a web scrap application that can retrieve news site content without copy paste and the data is stored in a database and displayed to the user application form for the reader without any ads and navigation that disturb the reader.

Keywords: web scraping, supervised learning, XPath, DOM tree.

Received: 05.12.2020

Accepted: 10.01.2021

Published: 05.02.2021

INTRODUCTION

Indonesia is one of the highest internet users in the world. Based on the infographic survey agency for penetration & behavior of Indonesian internet users, it reaches 68% of the 264.18 million Indonesian population, many users penetrate information via the internet. Information is the most important thing in life, sources of information can be through news sites(Permadi et al., 2018). A news site is a website that provides various information from various news sources to be displayed to users(Permadi et al., 2019). However, in general, sites do not only display information from outside sources, sometimes they also display information on their own website such as advertisements and navigation that disturbs news site readers and disturbs the convenience of the readers, and also some people may still collect data on the website with copying one by one on the website, but if the website that you manage is a large site with thousands of data, of course the work takes a long time, so to solve this problem the researchers conducted one of the web content extraction techniques using machine learning(Dedy Rahman Prehanto et al., 2019).

In classification, supervised learning is carried out by conducting training (training) to form a model(D. R. Prehanto et al., 2019). Classifier (classification algorithm) will form a model that adapts according to the features that exist in the data(Mashuri et al., 2019). The purpose of applying the supervised learning method is to implement and optimize web scraping techniques, as well as make it easier for users to process information from the results of web scraping techniques used to gather news information from various sites(Xia et al., 2017). To do web scraping, the basic thing is to know the DOM tree structure on news sites(Iacus, 2015).

The researcher analyzed the DOM tree and Xpath patterns which were later used as materials in the training data(Niculescu-Mizil & Caruana, 2005). DOM tree is a standard html document structure to form a web page, while Xpath is a query language from the representation of the DOM resulting in a particular node(Hardt et al., 2016). Rizaldi's research states that the use of Xpath is better than the CSS selector, therefore this study uses Xpath.

So this study takes the title Implementation of Web Scraping on News Sites Using the Supervised Learning Method. Below are some definitions related to this research.

METHODOLOGY

Web scraping

Web scraping is a process of data collecting through humans using a web browser, to handle the web that does not provide an API, web scraping requests data in HTML tags and then parses that data to extract the required information(Nylen & Wallisch, 2017). In the implementation, the use of web scraping includes programming techniques and technologies, such as data analysis and information security. Web scraping retrieves HTML data from a domain name, parses that data for target information, optionally stores target information, and moves to another page to repeat the scraping process(S.C.M. de S Sirisuriya, 2015).



Figure 1

Web Scraping

Web scraping describes the use of a program for extracting data from HTML files that exist on a particular site on the internet. Web scraping is the term used to describe the data extraction process on electronic files using a computer program(Ulbricht, 2020).

The next research was carried out by Rizaldi, explaining the crawling technique using Scrapy and Xpath, in this research XPath managed to meet the expected target. This is indicated by the production of a news corpus that has been divided into 3 news categories namely entertainment, community and culinary news(Khalil & Fakir, 2017).

Web scraping has several stages, the first step is to make scraping templates, explore the site navigation, automate navigation and extract information and save it(Mitchell, 2015).

Illustrating the steps for performing web scraping, starting with determining which sites will be used as scraping objects, then analyzing the page structure of the site to create a template(Aries Dwi Indriyanti et al., 2019). Then carry out exploration and navigation, the results of this process are then automated for extraction(A. D. Indriyanti et al., 2019). The final step is to save the scraping that has been done. The process of extracting information before it is saved can be done by utilizing the XPath method to find the desired information with a pattern according to a predefined template(Kistofer et al., 2019).



Figure 2

Web scraping steps

Content Extraction using Machine Learning

Content extraction using Machine Learning which is part of the branch of artificial intelligence. This includes the construction and study of training data systems as an introductory model, the essence of machine learning is related with representation and generalization. Generalization is a trait that the system will perform well on unseen data (Mutasim et al., 2018).

Supervised Learning

Supervised learning is a method for classifying each object in the data into several classes (Iacus, 2015). In supervised learning, each object in a data has features, namely the features that exist in each object. Each object in a data has the same number of features (Huber & Stuckenschmidt, 2020). Features are used as input to define a class on an object. In supervised learning, the class of each object is known (Huber & Stuckenschmidt, 2020). Therefore, the problem faced in supervised learning is how to map the objects into the right class using the features that each object has (Grandvalet & Bengio, 2005).

Classification in supervised learning is done by conducting training (training) to form a model. Classifier (classification algorithm) will form a model that adapts according to the features that exist in the data (Laine & Aila, 2017). The resulting model can be in the form of a tree, rule, or a function that can predict a class based on the features the data has (You et al., 2018)

Python

Python is a desktop or web-based programming language, a popular programming language used by many developers (Chollet, 2018). Python is one of the popular programming languages used by many developers. A survey according to the programming language website via www.tiobe.com, Python was ranked 4th in 2019 (GCM et al., 2016). Python can also be used for enterprises. Python is included in the programming language level, including high level language (Buduma & Locascio, 2017). Python is a programming language that can be used to build applications, be it desktop-based, web-based or mobile-based.

In python, there is a module used for the data extraction process so that the results obtained are faster and more practical. Python modules used in this study are (PF et al., 2016):

- a. Urllib is a module for handling request urls, the url you want to extract will be handled by urllib.
- b. BeautifulSoup is a module for parsing HTML. HTML parsing is a technique for separating text from HTML code tags on a website page to produce specific data

RESULTS AND DISCUSSION

In this study, using processing data derived from online news sites totaling 100 pages of news sites. List of news sites that extract web content in the table.

Table 1. list of news sites

| No | Name Web | Url |
|----|-------------------|-----------------------------------------------------------------------------------------|
| 1 | Media Indonesia | https://mediaindonesia.com |
| 2 | Kompas | https://www.kompas.com |
| 3 | Bisnis Indonesia | https://teknologi.bisnis.com |
| 4 | Pikiran Rakyat | https://www.pikiran-rakyat.com |
| 5 | Cek & Ricek | https://ceknricek.com |
| 6 | Siwalima | https://siwalimanews.com |
| 7 | Waspada | http://waspada.co.id |
| 8 | Analisa | https://analisadaily.com |
| 9 | Tribun Timur | https://makassar.tribunnews.com |
| 10 | Kedaulatan Rakyat | https://krjogja.com |
| 11 | Harian Jogja | https://jogjapolitan.harianjogja.com |
| 12 | Suara Merdeka | https://www.suaramerdeka.com |
| 13 | Solo Pos | https://www.solopos.com |
| 14 | Koran Sindo | https://jateng.sindonews.com |
| 15 | Sindo Weekly | https://sumeks.co |
| 16 | Sumatera Ekspres | http://www.sindoweekly.com |

| | | |
|----|-------------------|---------------------------------------------------------------------------------|
| 17 | Radar Palembang | http://www.radar-palembang.com |
| 18 | Tribul Sumsel | https://sumsel.tribunnews.com |
| 19 | Palempang Ekspres | https://palembang.tribunnews.com |
| 20 | Republika | https://republika.co.id |
| 21 | Antara | https://www.antaranews.com |
| 22 | Okezone | https://www.okezone.com/ |
| 23 | merdeka | https://www.merdeka.com/ |
| 24 | detik | https://www.detik.com/ |
| 25 | liputan | https://www.liputan6.com/ |

From all the list of sites in the research object, only certain data will be extracted and saved into several variables. The first variable is a link that contains a link to a web page that contains news. second is the headline, which contains news headlines, next is the content of news content, and the fourth is the XPath Selector variable.

DOM tree and Xpath tree structure analysis

In this analysis manual tag classification and after observing the html structure and tags, an example of an HTML document modeled as a tree is shown in Figure.

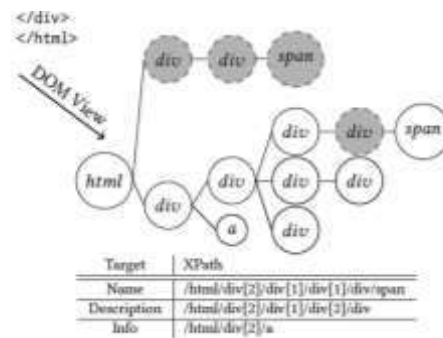


Figure 3 . HTML document and model tree

In this observation, the researcher conducted modeling from thehtml document for each website page and the tree is modeled to get the desired tags in the 70 news website web pages. As an example of Figure 4 scraping hml tags.

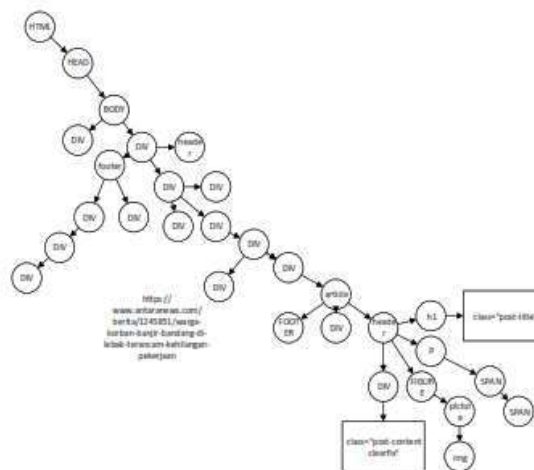


Figure 4 . DOM tree structure

The results of research on the implementation of web scraping techniques on news sites using the supervised learning method are as follows.

- Knowing the DOM tree and XPath patterns contained in a news site.
- Admins can manage news sites and manage news categories.
- Admins can manage the news site's XPath model selector.

- d. Admins can manage the target page of the news site to be scraped or extracted from the headlines and content.
- e. Users can read scraping results containing news headlines and news content in the newsscra mobile web application.
- f. Users can read news according to the categories provided and search for news according to the headline.

The results of the analysis of the DOM tree structure and site XPath

In retrieval of news content two XPath sections are taken the first XPath the title and the second XPath the contents, the XPath laying titles before content on the website has a different structure forms so that this analysis is carried out.

a. Media Indonesia

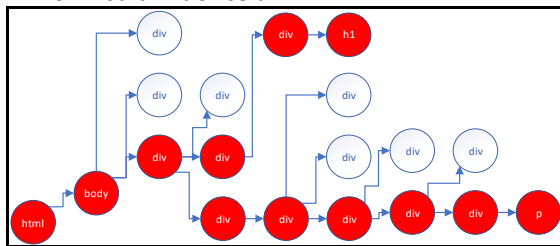
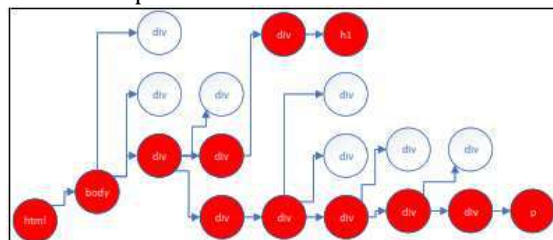


Figure 5
Dom tree media Indonesia

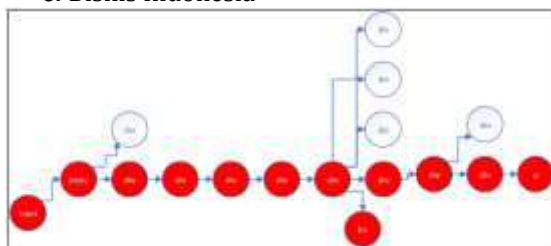
| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[3]/div[2]/div/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

b. Kompas



| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[3]/div[2]/div/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

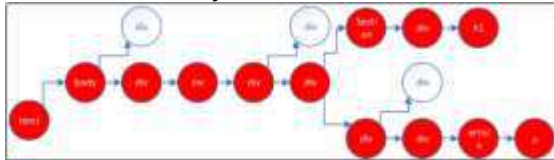
c. Bisnis Indonesia



| No | Target | XPath |
|----|--------|-------|
|----|--------|-------|

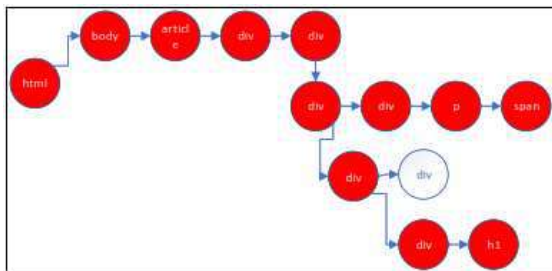
| | | |
|---|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/div[2]/div/div/div[1]/h1 |
| 2 | Content | /html/body/div[3]/div[3]/div[1]/div[3]/div[2]/div[2]/p |

d. Pikiran Rakyat



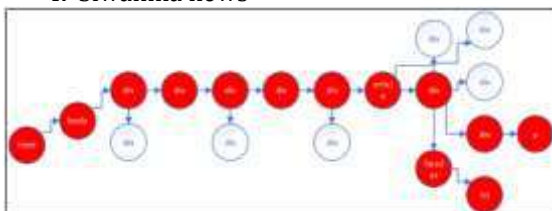
| No | Target | XPath |
|----|----------|----------------------------------------------------------|
| 1 | Headline | /html/body/div[2]/div/div/div[2]/section/div/h1 |
| 2 | Content | /html/body/div[2]/div/div/div[2]/div[1]/div[2]/article/p |

e. Cek & Ricek



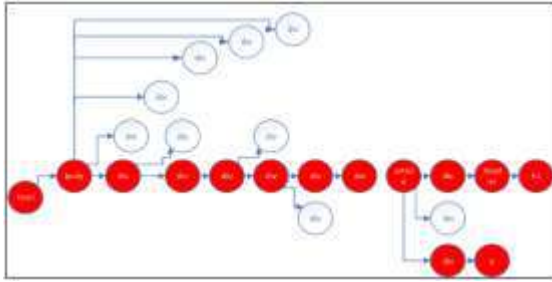
| No | Target | XPath |
|----|----------|--------------------------------------------------------|
| 1 | Headline | /html/body/article/div/div[1]/div[1]/div[2]/h1 |
| 2 | Content | /html/body/article/div/div[1]/div[1]/div/p/span/text() |

f. Siwalima news



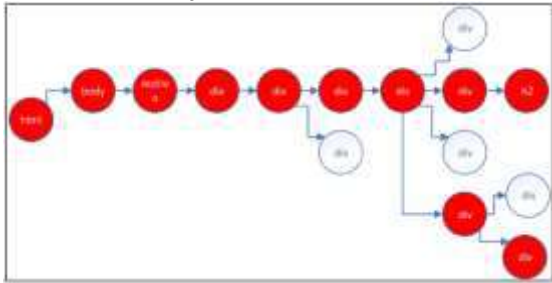
| No | Target | XPath |
|----|----------|------------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/div/div/div[1]/div[1]/article/div[2]/header/h1 |
| 2 | Content | /html/body/div[1]/div/div/div[1]/div[1]/article/div[2]/div[3]/p |

g. Waspada



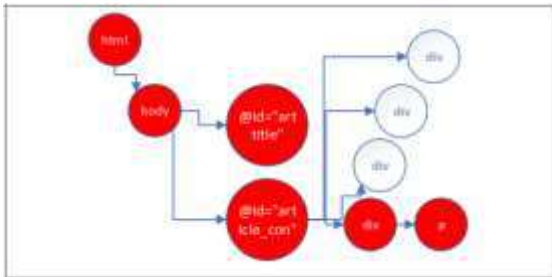
| No | Target | XPath |
|----|----------|-------------------------------------------------------------------------|
| 1 | Headline | /html/body/div[6]/div[2]/div/div[2]/div[1]/div/article/div[1]/header/h1 |
| 2 | Content | /html/body/div[6]/div[2]/div/div[2]/div[1]/div/article/div[3]/p |

h. Analisisaday



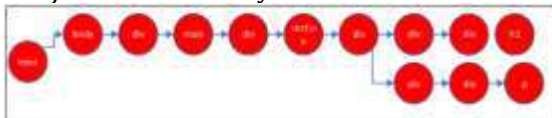
| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

i. Makasar Trimbum



| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

j. Kedaulatan Rakyat



| No | Target | XPath |
|----|----------|------------------------------------------------------------|
| 1 | Headline | /html/body/div[1]/main/div/section/div/div[1]/div[1]/h1 |
| 2 | Content | /html/body/div[1]/main/div/section/div/div[1]/div[2]/div/p |

This target data page contains a list of site links that will be scrapped or retrieved from the news content through the XPath selector (training data) that has been provided previously. The admin clicks the button, the process will perform the html feature extraction and analyze the XPath matched with the XPath selector or the previous XPath model that has been provided if the same then the headline content and the news content will be saved if the XPath selector model is not the same as the news site that will be scraping the food data can retrieve the news content and the system will display an error such as Figure.



Figure 6. Target Site

This user page displays all selected news contents by pressing the read-further button and news search pages based on news categories as shown in the Figure 7.

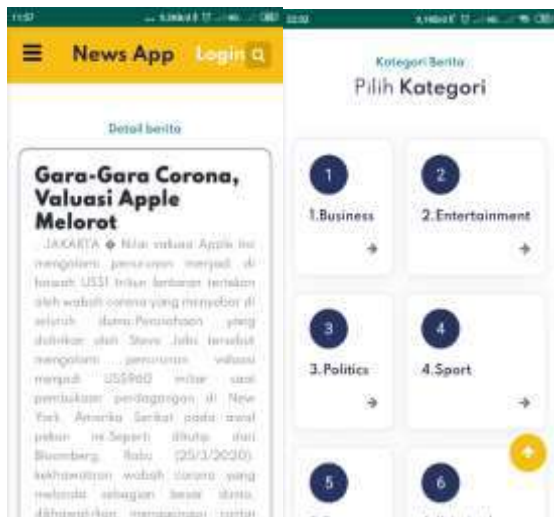


Figure 6. read the news and choose a category

CONCLUSION

Based on the results and previous discussion, the results of this study are in the form of DOM tree and XPath pattern analysis of the news sites studied, and implemented in the form of a web-based scrap news web application using the python programming language with the flask framework. The results obtained are knowing the pattern of DOM and XPath news sites. The XPath pattern always differs in a news site among others. The results of the extraction data using XPath were more complete. By using this supervised learning method, the more data on the XPath training model the better this application will be.

So when conducting web scraping, the site is no longer analyzed because it's already in the XPath model data.

REFERENCES

- Buduma, N., & Locascio, N. (2017). Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. / Nikhil Buduma ; with contributions by Nicholas Locascio. In Designing next-generation machine intelligence algorithms.
- Chollet, F. (2018). Deep Learning with Python. In Manning.
- GCM, G., Laterra, P., Aparicio, V., & Costa, J. (2016). Glyphosate retention in grassland riparian areas is reduced by the invasion of exotic trees. *Phyton*. <https://doi.org/10.32604/phyton.2016.85.108>
- Grandvalet, Y., & Bengio, Y. (2005). Semi-supervised learning by entropy minimization. *Advances in Neural Information Processing Systems*.
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*.
- Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2020.02.005>.
- Iacus, S. M. (2015). Automated Data Collection with R - A Practical Guide to Web Scraping and Text Mining. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v068.b03>.
- Indriyanti, A. D., Prehanto, D. R., Prisma, I. G. L. E. P., Soeryanto, Sujatmiko, B., & Fikandda, J. (2019). Simple Additive Weighting algorithm to aid administrator decision making of the underprivileged scholarship. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1402/6/066070>.
- Indriyanti, Aries Dwi, Prehanto, D. R., Permadi, G. S., Mashuri, C., & Vitadiar, T. Z. (2019). Using Fuzzy Time Series (FTS) and Linear Programming for Production Planning and Planting Pattern Scheduling Red Onion. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523007>
- .Khalil, S., & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*. <https://doi.org/10.1016/j.softx.2017.04.004>.
- Kistofer, T., Permadi, G. S., & Vitadiar, T. Z. (2019). Development of Digital System Learning Media Using Digital Learning System. <https://doi.org/10.2991/assehr.k.191217.030>.
- Laine, S., & Aila, T. (2017). Temporal ensembling for semi-supervised learning. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*.
- Mashuri, C., Mujianto, A. H., Sucipto, H., Arsam, R. Y., & Permadi, G. S. (2019). Production Time Optimization using Campbell Dudek Smith (CDS) Algorithm for Production Scheduling. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523009>.
- Mitchell, R. (2015). Web Scraping with Python Collecting Data From The Modern Web. In O'Reilly.
- Mutasim, A. K., Tipu, R. S., Bashar, M. R., Islam, M. K., & Amin, M. A. (2018). Computational intelligence for pattern recognition in EEG signals. In *Studies in Computational Intelligence*. https://doi.org/10.1007/978-3-319-89629-8_11.
- Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*. <https://doi.org/10.1145/1102351.1102430>.
- Nylen, E. L., & Wallisch, P. (2017). Web Scraping. In *Neural Data Science*. <https://doi.org/10.1016/b978-0-12-804043-0.00010-6>.
- Permadi, G. S., Adi, K., & Gernowo, R. (2018). Application Mail Tracking Using RSA Algorithm As Security Data and HOT-Fit a Model for Evaluation System. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/20183111007>.
- Permadi, G. S., Vitadiar, T. Z., Kistofer, T., & Mujianto, A. H. (2019). The Decision Making Trial and Evaluation Laboratory (Dematel) and Analytic Network Process (ANP) for Learning Material Evaluation System. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523011>.
- PF, C., Scarpassa, J., Pretto-Giordano, L., Otaguiri, E., Yamada-Ogatta, S., Nakazato, G., Perugini, M., Moreira, I., & Vilas-Balãs, G. (2016). Antibacterial activity of avocado extracts (*Persea americana* Mill.) against *Streptococcus agalactiae*. *Phyton*. <https://doi.org/10.32604/phyton.2016.85.218>.
- Prehanto, D. R., Indriyanti, A. D., Nuryana, K. D., Soeryanto, S., & Mubarak, A. S. (2019). Use of Naïve Bayes classifier algorithm to detect customers' interests in buying internet token. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1402/6/066069>.
- Prehanto, Dedy Rahman, Indriyanti, A. D., Mashuri, C., & Permadi, G. S. (2019). Soil Moisture Prediction using Fuzzy Time Series and Moisture sensor Technology on Shallot Farming. *E3S Web of Conferences*. <https://doi.org/10.1051/e3sconf/201912523002>.

- S.C.M. de S Sirisuriya. (2015). A Comparative Study on Web Scraping. 8th International Research Conference KDU.
- Ulbricht, L. (2020). Scraping the demos. Digitalization, web scraping and the democratic project. Democratization. <https://doi.org/10.1080/13510347.2020.1714595>.
- Xia, Y., Qin, T., Chen, W., Bian, J., Yu, N., & Liu, T. Y. (2017). Dual supervised learning. 34th International Conference on Machine Learning, ICML 2017.
- You, Z., Raich, R., Fern, X. Z., & Kim, J. (2018). Weakly Supervised Dictionary Learning. IEEE Transactions on Signal Processing. <https://doi.org/10.1109/TSP.2018.2807422>.