

OXPath: A language for scalable data extraction, automating web

VLDB Journal

22, 47-72

DOI: [10.1007/s00778-012-0286-6](https://doi.org/10.1007/s00778-012-0286-6)

Citation Report

| # | ARTICLE | IF | CITATIONS |
|----|---|-----|-----------|
| 1 | Strigil. , 2013, , . | | 3 |
| 2 | Parallel Approach and Platform for Large-Scale WEB Data Extraction. , 2013, , . | | 2 |
| 3 | Web Information Systems Engineering â€œ WISE 2013. Lecture Notes in Computer Science, 2013, , . | 1.3 | 0 |
| 4 | Effective web scraping with XPath. , 2013, , . | | 11 |
| 5 | A methodology for social BI. , 2014, , . | | 15 |
| 6 | Ducky. , 2014, , . | | 4 |
| 7 | Multi-feature and DAG-Based Multi-tree Matching Algorithm for Automatic Web Data Mining. , 2014, , . | | 3 |
| 8 | Search in the universe of big networks and data. IEEE Network, 2014, 28, 20-25. | 6.9 | 16 |
| 9 | DIADEM. Proceedings of the VLDB Endowment, 2014, 7, 1845-1856. | 3.8 | 43 |
| 10 | The Augmented Web. ACM Transactions on the Web, 2015, 9, 1-30. | 2.5 | 27 |
| 11 | A Scalable Approach to Harvest Modern Weblogs. International Journal on Artificial Intelligence Tools, 2015, 24, 1540005. | 1.0 | 0 |
| 12 | Crawling images with web browser support. , 2015, , . | | 2 |
| 13 | A deep web query interface discovery method. , 2015, , . | | 2 |
| 14 | AutoRM: An effective approach for automatic Web data record mining. Knowledge-Based Systems, 2015, 89, 314-331. | 7.1 | 12 |
| 15 | Complexities of practical web automation. , 2015, , . | | 1 |
| 16 | WADaR. Proceedings of the VLDB Endowment, 2015, 8, 1996-1999. | 3.8 | 18 |
| 17 | CWrap: web wrapping using context variables. International Journal of Knowledge and Web Intelligence, 2016, 5, 304. | 0.2 | 0 |
| 18 | Dynamic Integration for Deep Web Search Results. , 2016, , . | | 1 |

| # | ARTICLE | IF | CITATIONS |
|----|--|------|-----------|
| 19 | Social Business Intelligence in Action. Lecture Notes in Computer Science, 2016, , 33-48. | 1.3 | 6 |
| 20 | P <scp>ea</scp> CE-Ful Web Event Extraction and Processing as Bitemporal Mutable Events. ACM Transactions on the Web, 2016, 10, 1-47. | 2.5 | 1 |
| 21 | A survey of methods for the extraction of information from Web resources. Programming and Computer Software, 2016, 42, 279-291. | 0.9 | 15 |
| 22 | Web News Extraction via Tag Path Feature Fusion Using DS Theory. Journal of Computer Science and Technology, 2016, 31, 661-672. | 1.5 | 5 |
| 23 | Survey on challenges of Question Answering in the Semantic Web. Semantic Web, 2017, 8, 895-920. | 1.9 | 151 |
| 24 | OXPath-Based Data Acquisition for dblp. , 2017, , . | | 7 |
| 25 | Data context informed data wrangling. , 2017, , . | | 11 |
| 26 | Advances in Information Systems Development. Lecture Notes in Information Systems and Organisation, 2018, , . | 0.6 | 3 |
| 27 | Browserless Web Data Extraction. , 2018, , . | | 9 |
| 28 | Stable web scraping: an approach based on neighbour zone and path similarity of page elements. International Journal of Web Engineering and Technology, 2018, 13, 301. | 0.2 | 3 |
| 30 | Deep Web crawling: a survey. World Wide Web, 2019, 22, 1577-1610. | 4.0 | 24 |
| 31 | VADA: an architecture for end user informed data preparation. Journal of Big Data, 2019, 6, . | 11.0 | 16 |
| 32 | Dynamap. , 2019, , . | | 4 |
| 33 | A Crawler Architecture for Harvesting the Clear, Social, and Dark Web for IoT-Related Cyber-Threat Intelligence. , 2019, , . | | 25 |
| 34 | Design and Implementation of Engineering Standard Database System Based on Data Mining. , 2019, , . | | 0 |
| 35 | A novel approach for Web page modeling in personal information extraction. World Wide Web, 2019, 22, 603-620. | 4.0 | 4 |
| 36 | Leopard â€” A baseline approach to attribute prediction and validation for knowledge graph population. Web Semantics, 2019, 55, 102-107. | 2.9 | 4 |
| 37 | DCADE: divide and conquer alignment with dynamic encoding for full page data extraction. Applied Intelligence, 2020, 50, 271-295. | 5.3 | 2 |

| # | ARTICLE | IF | CITATIONS |
|----|--|-----|-----------|
| 38 | Feedback driven improvement of data preparation pipelines. Information Systems, 2020, 92, 101480. | 3.6 | 9 |
| 39 | Robust Web Data Extraction Based on Weighted Path-layer Similarity. Journal of Computer Information Systems, 2022, 62, 536-546. | 2.9 | 1 |
| 40 | inTIME: A Machine Learning-Based Framework for Gathering and Leveraging Web Data to Cyber-Threat Intelligence. Electronics (Switzerland), 2021, 10, 818. | 3.1 | 36 |
| 41 | IHWC: intelligent hidden web crawler for harvesting data in urban domains. Complex & Intelligent Systems, 0, , 1. | 6.5 | 1 |
| 43 | Bitemporal Complex Event Processing of Web Event Advertisements. Lecture Notes in Computer Science, 2013, , 333-346. | 1.3 | 5 |
| 44 | Ringer: web automation by demonstration. , 2016, , . | | 30 |
| 45 | Ringer: web automation by demonstration. ACM SIGPLAN Notices, 2016, 51, 748-764. | 0.2 | 6 |
| 46 | Research on Deep Web Query Interface Clustering Based on Hadoop. Journal of Software, 2014, 9, . | 0.6 | 2 |
| 47 | Predicting Economic Indicators from Web Text Using Sentiment Composition. International Journal of Computer and Communication Engineering, 2014, 3, 109-115. | 0.2 | 26 |
| 48 | NEXIR: A Novel Web Extraction Rule Language toward a Three-Stage Web Data Extraction Model. Lecture Notes in Computer Science, 2013, , 29-42. | 1.3 | 3 |
| 49 | PeaCE-Ful Web Event Extraction and Processing. Lecture Notes in Computer Science, 2013, , 523-526. | 1.3 | 0 |
| 51 | Heterogeneous Web Data Extraction Algorithm Based On Modified Hidden Conditional Random Fields. Journal of Networks, 2014, 9, . | 0.4 | 0 |
| 52 | CWrap: web wrapping using context variables. International Journal of Knowledge and Web Intelligence, 2016, 5, 304. | 0.2 | 0 |
| 53 | UniQue: An Approach for Unified and Efficient Querying of Heterogeneous Web Data Sources. , 2016, , . | | 0 |
| 54 | Enriching Existing Test Collections with XPath. Lecture Notes in Computer Science, 2017, , 152-158. | 1.3 | 1 |
| 55 | A MODEL FOR AUTOMATED MATCHING BETWEEN JOB MARKET DEMAND AND UNIVERSITY CURRICULA OFFER. SEEU Review, 2017, 12, 188-217. | 0.8 | 0 |
| 56 | User-Friendly and Extensible Web Data Extraction. Lecture Notes in Information Systems and Organisation, 2018, , 225-241. | 0.6 | 0 |
| 57 | Crawling Chinese-Myanmar Parallel Corpus: Automatic Collection, Screening and Cleaning Corpus. IOP Conference Series: Materials Science and Engineering, 0, 646, 012046. | 0.6 | 0 |

| # | ARTICLE | IF | CITATIONS |
|----|---|-----|-----------|
| 58 | A browserless architecture for extracting web prices. , 2020, , . | | 0 |
| 59 | Data science with Vadalog: Knowledge Graphs with machine learning and reasoning in practice. Future Generation Computer Systems, 2022, 129, 407-422. | 7.5 | 11 |
| 60 | Modified Kleene Star Algorithm Using Max-Plus Algebra and Its Application in the Railroad Scheduling Graphical User Interface. Computation, 2023, 11, 11. | 2.0 | 2 |
| 61 | Scraping Data from Web Pages Using SPARQL Queries. Lecture Notes in Computer Science, 2023, , 293-300. | 1.3 | 0 |
| 62 | When Automatic Filtering Comes to the Rescue: Pre-Computing Company Competitor Pairs in Owler. , 2023, 1, 1-23. | | 0 |
| 63 | Manipulation Mask Generator: High-Quality Image Manipulation Mask Generation Method Based on Modified Total Variation Noise Reduction. , 2023, , . | | 0 |