

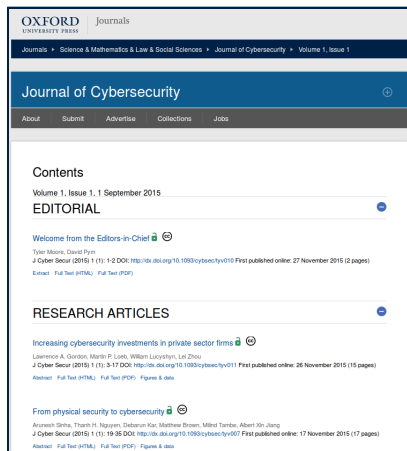
Smart Harvesting with XPath

Mandy Neumann
TH Köln
mandy.neumann@th-koeln.de

Christopher Michels
University of Trier
michelsc@uni-trier.de

February 22, 2018

Harvesting Bibliographic Data



OXFORD UNIVERSITY PRESS Journals

Journals > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1



Journal of Cybersecurity

About Submit Advertise Collections Jobs

Contents



Volume 1, Issue 1, 1 September 2015

EDITORIAL



Welcome from the Editors-in-Chief  

Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: <http://dx.doi.org/10.1093/cysec/hyv010> First published online: 27 November 2015 (2 pages)
Extract Full Text (HTML) Full Text (PDF)

RESEARCH ARTICLES

Increasing cybersecurity investments in private sector firms  

Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: <http://dx.doi.org/10.1093/cysec/hyv011> First published online: 26 November 2015 (15 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data

From physical security to cybersecurity  

Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang
J Cyber Secur (2015) 1 (1): 19-35 DOI: <http://dx.doi.org/10.1093/cysec/hyv007> First published online: 17 November 2015 (17 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data



home | browse | search | about



computer science bibliography

search dblp

Journal of Cybersecurity, Volume 1

> Home > Journals > Journal of Cybersecurity  modern  Trier 2

Volume 1, Number 1, September 2015

Editorial

  Tyler Moore, David J. Pym:
Welcome from the Editors-in-Chief. 1-2

Research Articles

  Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou:
Increasing cybersecurity investments in private sector firms. 3-17

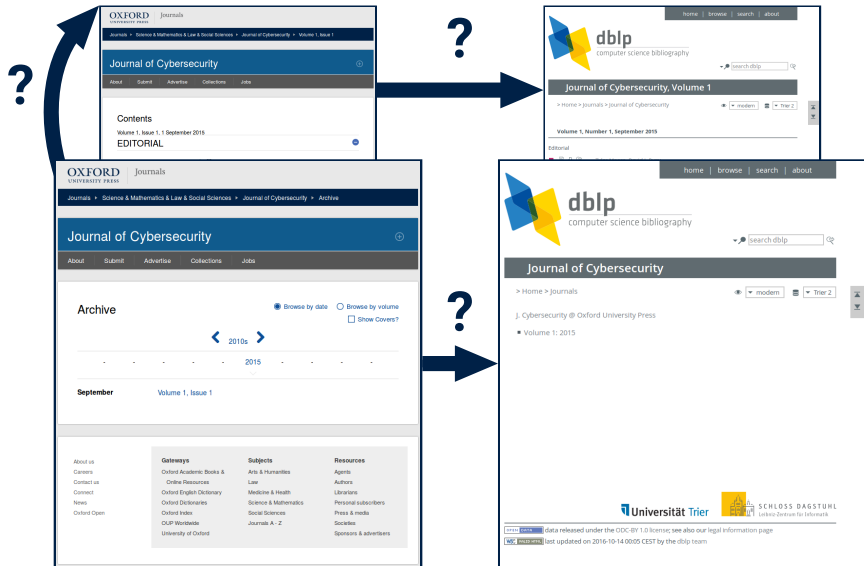
  Arunesh Sinha, Thanh Hong Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang:
From physical security to cybersecurity. 19-35

  Tristan Caulfield, Andrew Fielder:
Optimizing time allocation for network defence. 37-51

  Jon R. Lindsay:
Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. 53-67

  Harold Abelson, Ross J. Anderson, Steven M. Bellovin, Josh Benaloh, Matt Blaze, Whitfield Diffie, John Gilmore, Matthew Green, Susan Landau, Peter G. Neumann, Ronald L. Rivest, Jeffrey I. Schiller, Bruce Schneier, Michael A. Specter,

Accessing Bibliographic Data



Accessing Bibliographic Data



Google Sign In

Alerts More ▾

Google Scholar

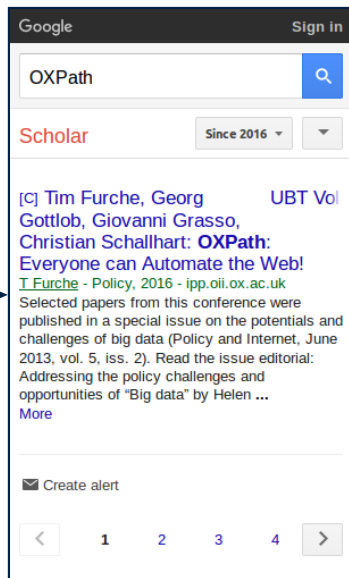
XPath

Articles Case law

Stand on the shoulders of giants

About Google Scholar Privacy Terms

Go to Google Scholar



Google Sign In

XPath

Scholar Since 2016 ▾ ▾

[C] Tim Furche, Georg UBT Vol
Gottlob, Giovanni Grasso,
Christian Schallhart: **XPath:**
Everyone can Automate the Web!
[T Furche - Policy, 2016 - ipp.oii.ox.ac.uk](#)
Selected papers from this conference were published in a special issue on the potentials and challenges of big data (Policy and Internet, June 2013, vol. 5, iss. 2). Read the issue editorial: Addressing the policy challenges and opportunities of "Big data" by Helen ...
[More](#)

Create alert

< 1 2 3 4 >

Project Profile: Smart Harvesting II

Partners:

- dblp, GESIS, TH Köln

Motivation:

- extract bibliographic data with XPath
- facilitate maintenance of scientific literature databases

Solution:

- provide working environment and tools to use XPath

XPath:

- simple, declarative language for web data extraction

Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 XPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration

Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 XPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration

Sources of Raw Digital Data

OXFORD JOURNALS

THE COMPUTER JOURNAL

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Science & Mathematics > Computer Journal > Volume 59 Issue 9

Table of Contents

Volume 59 Issue 9 September 2016

For checked items
 view abstracts download to citation manager

Section C

ORIGINAL ARTICLES

Arambam Neelima and Kh Manglem Singh
Perceptual Hash Function based on Scale-Invariant Feature Transform and Singular Value Decomposition
The Computer Journal (2016) 59 (9): 1275-1283 doi:10.1093/comjnl/btw019
» Abstract » Full Text (HTML) » Full Text (PDF)

Wei Ni
Minimized Error Propagation Location Method Based on Error Estimation
The Computer Journal (2016) 59 (9): 1282-1288 doi:10.1093/comjnl/btw021
» Abstract » Full Text (HTML) » Full Text (PDF)

D. Thernozhi and Chandrase Aravindan
Paraphrase Identification by Using Clause-Based Similarity Features and Machine Translation Metrics
The Computer Journal (2016) 59 (9): 1289-1302 doi:10.1093/comjnl/btw023
» Abstract » Full Text (HTML) » Full Text (PDF)

Alok Kumar Singh Kushwaha and Rajeev Srivastava
Maritime Object Segmentation Using Dynamic Background Modeling and Shadow Suppression
The Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/btw021
» Abstract » Full Text (HTML) » Full Text (PDF)

« Previous | Next »

This Issue
September 2016 59 (9)

» Index By Author
» Front Matter (PDF)
» Table of Contents (PDF)
» Back Matter (PDF)

» Section C

» ORIGINAL ARTICLES

Find articles in this issue containing these words:

Advance Access

OXFORD journals
CYBERSECURITY TEAM

Home > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1

Journal of Cybersecurity

Home Editors Advertisement Collections Jobs

Contents

Volume 1, Issue 1.1 September 2015

EDITORIAL

Welcome from the Editors-in-Chief

Tim Moore, David Szym
J Cyber Secur (2015) 1 (1): 1-2 DOI: <https://doi.org/10.1093/cysec/tyv011> First published online: 27 November 2015 (2 pages)
Cover: Full Text (HTML) Full Text (PDF)

RESEARCH ARTICLES

Increasing cybersecurity investments in private sector firms

Lawrence A. Gordon, Mark P. Lewis, William Longstaffe, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: <https://doi.org/10.1093/cysec/tyv011> First published online: 26 November 2015 (15 pages)
Abstract: Full Text (HTML) Full Text (PDF) Figures & data

From physical security to cybersecurity

Arunesh Sinha, Thom H. Higgins, Debaran Kar, Matthew Brown, Mridul Taneja, Albert Yu-Jong
J Cyber Secur (2015) 1 (1): 19-30 DOI: <https://doi.org/10.1093/cysec/tyv007> First published online: 17 November 2015 (17 pages)
Abstract: Full Text (HTML) Full Text (PDF) Figures & data

Sources of Raw Digital Data

OXFORD JOURNALS

THE COMPUTER JOURNAL

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Science & Mathematics > Computer Journal > Volume 59 Issue 9

Table of Contents

Volume 59 Issue 9 September 2016

For checked items
 view abstracts download to citation manager

Section C

ORIGINAL ARTICLES

Arambam Neelima and Kh Manglem Singh
Perceptual Hash Function based on Scale-Invariant Feature Transform and Singular Value Decomposition
The Computer Journal (2016) 59 (9): 1279-1293 doi:10.1093/comjnl/btw019
» Abstract » Full Text (HTML) » Full Text (PDF)

Wei Ni
Minimized Error Propagation Location Method Based on Error Estimation
The Computer Journal (2016) 59 (9): 1292-1298 doi:10.1093/comjnl/btw011
» Abstract » Full Text (HTML) » Full Text (PDF)

D. Theramozi and Chandrabose Aravindan
Paraphrase Identification by Using Clause-Based Similarity Features and Machine Translation Metrics
The Computer Journal (2016) 59 (9): 1289-1302 doi:10.1093/comjnl/btw013
» Abstract » Full Text (HTML) » Full Text (PDF)

Alok Kumar Singh Kushwaha and Rajeev Srivastava
Maritime Object Segmentation Using Dynamic Background Modeling and Shadow Suppression
The Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/btw011
» Abstract » Full Text (HTML) » Full Text (PDF)

» Previous | Next »

This Issue
September 2016 59 (9)



» Index By Author
» Front Matter (PDF)
» Table of Contents (PDF)
» Back Matter (PDF)

» Section C

» ORIGINAL ARTICLES

Find articles in this issue containing these words:

Advance Access

OXFORD journals
UNIVERSITY PRESS

Home > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1

Journal of Cybersecurity

Home Editors Advertisement Collections Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

Welcome from the Editors-in-Chief 

Tim Moore, David Szym
J Cyber Secur (2015) 1 (1): 1-2 DOI: <https://doi.org/10.1093/cysec/tyv011> First published online: 27 November 2015 (2 pages)
Cover: Full Text (HTML), Full Text (PDF)

RESEARCH ARTICLES

Increasing cybersecurity investments in private sector firms 

Lawrence A. Gordon, Mark P. Lusk, William Longstaffe, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: <https://doi.org/10.1093/cysec/tyv011> First published online: 26 November 2015 (15 pages)
Abstract: Full Text (HTML), Full Text (PDF), Figures & data

From physical security to cybersecurity 

Arachch Sathya, Thamm H. Nigam, Dattaraj Kar, Matthew Brown, Mridu Taneja, Albert Yu-Jong
J Cyber Secur (2015) 1 (1): 19-30 DOI: <https://doi.org/10.1093/cysec/tyv011> First published online: 17 November 2015 (17 pages)
Abstract: Full Text (HTML), Full Text (PDF), Figures & data

Sources of Raw Digital Data

The image shows a screenshot of the Oxford Journals website, specifically the 'Journal of Cybersecurity' page. The page displays the journal's title, volume information, and a list of articles under the 'CONTENTS' section. Overlaid on the left side of the page is a code editor window showing the HTML source code for the table of contents. The code is structured as follows:

```
<h3 id="sectionORIGINALARTICLES">
  <span>
    <a class="toc-section-return" href="#content-block"/>
      ORIGINAL ARTICLES
    </span>
</h3>
<ul class="cit-list">
  <li class="cit has-earlier-version from-current-issue toc-cit">
    <div class="cit-form-select">
      <div class="cit-metadata">
        <ul class="cit-first-element cit-auth-list">
          <h4 class="cit-title-group">Perceptual Hash Function based on
            Scale-Invariant Feature Transform and Singular Value Decomposition
          </h4>
          <cite>
            <abbr class="site-title" title="The Computer
              Journal">The Computer Journal</abbr>
            <span class="cit-print-date">
              <span class="cit-vol">59 </span>
              <span class="cit-issue">
                <span class="cit-pages">
                  <span class="cit-ahead-of-print-date">
                    <span class="cit-doi">
</div>
</div>
</li>
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
<li class="cit has-earlier-version from-current-issue toc-cit">
</ul>
```

The right side of the image shows the actual rendered content of the website, including the 'Contents' section with the volume and issue information, and a list of articles such as 'Increasing cybersecurity investments in private sector firms' and 'From physical security to cybersecurity'.

Sources of Raw Digital Data

```
OXFORD JOURNALS
OXFORD
UNIVERSITY PRESS
Journals
Journals > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1
Journal of Cybersecurity
About Submit Advertise Collections Jobs
Contents
Volume 1, Issue 1, 1 September 2015
EDITORIAL
Welcome from the Editors-in-Chief
Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: http://dx.doi.org/10.1093/cybersec/tyv010 First published online: 27 November 2015 (2 pages)
Extract Full Text (HTML) Full Text (PDF)
RESEARCH ARTICLES
Increasing cybersecurity investments in private sector firms
Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: http://dx.doi.org/10.1093/cybersec/tyv011 First published online: 26 November 2015 (15 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data
From physical security to cybersecurity
Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang
J Cyber Secur (2015) 1 (1): 19-35 DOI: http://dx.doi.org/10.1093/cybersec/tyv007 First published online: 17 November 2015 (17 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data
```

OXFORD UNIVERSITY PRESS Journals

Journals > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1

Journal of Cybersecurity

About Submit Advertise Collections Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

Welcome from the Editors-in-Chief 

Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: <http://dx.doi.org/10.1093/cybersec/tyv010> First published online: 27 November 2015 (2 pages)
Extract Full Text (HTML) Full Text (PDF)

RESEARCH ARTICLES

Increasing cybersecurity investments in private sector firms 

Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: <http://dx.doi.org/10.1093/cybersec/tyv011> First published online: 26 November 2015 (15 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data

From physical security to cybersecurity 

Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang
J Cyber Secur (2015) 1 (1): 19-35 DOI: <http://dx.doi.org/10.1093/cybersec/tyv007> First published online: 17 November 2015 (17 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data

Sources of Raw Digital Data

```
OXFORD JOURNALS  
OXFORD UNIVERSITY PRESS  
Journals  
Journals > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1  
Journal of Cybersecurity  
About Submit Advertise Collections Jobs  
Contents  
Volume 1, Issue 1, 1 September 2015  
EDITORIAL  
Welcome from the Editors-in-Chief  
Tyler Moore, David Pym  
J Cyber Secur (2015) 1 (1): 1-2 DOI: http://dx.doi.org/10.1093/cybsec/tyv010 First published online: 27 November 2015 (2 pages)  
RESEARCH ARTICLES  
Increasing cybersecurity investments in private sector firms  
Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou  
J Cyber Secur (2015) 1 (1): 3-17 DOI: http://dx.doi.org/10.1093/cybsec/tyv011 First published online: 26 November 2015 (15 pages)  
From physical security to cybersecurity  
Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang  
J Cyber Secur (2015) 1 (1): 19-35 DOI: http://dx.doi.org/10.1093/cybsec/tyv007 First published online: 17 November 2015 (17 pages)
```

OXFORD UNIVERSITY PRESS | Journals

Journals > Science & Mathematics & Law & Social Sciences > Journal of Cybersecurity > Volume 1, Issue 1

Journal of Cybersecurity

About Submit Advertise Collections Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

Welcome from the Editors-in-Chief

Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: <http://dx.doi.org/10.1093/cybsec/tyv010> First published online: 27 November 2015 (2 pages)
Extract Full Text (HTML) Full Text (PDF)

RESEARCH ARTICLES

Increasing cybersecurity investments in private sector firms

Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, Lei Zhou
J Cyber Secur (2015) 1 (1): 3-17 DOI: <http://dx.doi.org/10.1093/cybsec/tyv011> First published online: 26 November 2015 (15 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data

From physical security to cybersecurity

Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang
J Cyber Secur (2015) 1 (1): 19-35 DOI: <http://dx.doi.org/10.1093/cybsec/tyv007> First published online: 17 November 2015 (17 pages)
Abstract Full Text (HTML) Full Text (PDF) Figures & data

Sources of Raw Digital Data

```

<div id="section1" class="ARTICLES">
  <ul class="list">
    <li class="list-item">
      <div class="toc-section-heading">
        <h2 id="editorial" class="toc-heading">EDITORIAL</h2>
      </div>
      <div class="inner-collapsible-content-wrapper">
        <ul class="toc-section">
          <li class="first last odd toc-item">
            <div class="toc-citation">
              <div id="" class="highwire-article-citation highwire-citation type-highwire-article tooltip-enable highwire_article_citation tooltip-processed" title="Welcome from the Editors-in-Chief" rel="/highwire/article_citation_preview/60529" data-node-nid="60529" data-pisa="cybers;1/1/1" data-pisa-master="cybers;tyv010" data-apath="/cybers/1/1/1.atom">
                <cite class="highwire-cite highwire-cite-highwire-article highwire-citation-jnl-oup-toc-one-line clearfix">
                  <div class="highwire-cite-title-access">
                    <div class="highwire-cite-detail-wrapper">
                      <span class="highwire-cite-authors add-author-link-processed">
                        <span class="highwire-cite-jnl-info">
                          <span class="highwire-cite-doi">
                            <span class="highwire-cite-pub">First published online: 27 November 2015 (2 pages)</span>
                          </div>
                        <span class="highwire-cite-extras">
                          </span>
                        </cite>
                      </div>
                    </div>
                  </div>
                </div>
              </div>
            </li>
          </ul>
        </div>
      </div>
    </li>
  </ul>
</div>
    
```

```

<div class="inner-collapsible-content-heading-wrapper">
  <h2 id="editorial" class="toc-heading">EDITORIAL</h2>
  <span class="inner-content-toggle">
</div>
<div class="inner-collapsible-content-wrapper">
  <ul class="toc-section">
    <li class="first last odd toc-item">
      <div class="toc-citation">
        <div id="" class="highwire-article-citation highwire-citation type-highwire-article tooltip-enable highwire_article_citation tooltip-processed" title="Welcome from the Editors-in-Chief" rel="/highwire/article_citation_preview/60529" data-node-nid="60529" data-pisa="cybers;1/1/1" data-pisa-master="cybers;tyv010" data-apath="/cybers/1/1/1.atom">
          <cite class="highwire-cite highwire-cite-highwire-article highwire-citation-jnl-oup-toc-one-line clearfix">
            <div class="highwire-cite-title-access">
              <div class="highwire-cite-detail-wrapper">
                <span class="highwire-cite-authors add-author-link-processed">
                  <span class="highwire-cite-jnl-info">
                    <span class="highwire-cite-doi">
                      <span class="highwire-cite-pub">First published online: 27 November 2015 (2 pages)</span>
                    </div>
                  <span class="highwire-cite-extras">
                    </span>
                  </cite>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </li>
  </ul>
</div>
    
```

From physical security to cybersecurity

Arunesh Sinha, Thanh H. Nguyen, Debarun Kar, Matthew Brown, Milind Tambe, Albert Xin Jiang
 J Cyber Secur (2015) 1 (1): 19-35 DOI: <http://dx.doi.org/10.1093/cybsec/yyv007> First published online: 17 November 2015 (17 pages)

Abstract Full Text (HTML) Full Text (PDF) Figures & data

Sources of Raw Digital Data

OXFORD
ACADEMIC

THE COMPUTER JOURNAL

bc
The Chartered
Institute
for IT

Select Year 2016 Select Issue 1 December - Volume 59, Issue 12

January - Volume 59, Issue 1
February - Volume 59, Issue 2
March - Volume 59, Issue 3
April - Volume 59, Issue 4
May - Volume 59, Issue 5
June - Volume 59, Issue 6
July - Volume 59, Issue 7
August - Volume 59, Issue 8
September - Volume 59, Issue 9
13 October - Volume 59, Issue 10
3 November - Volume 59, Issue 11
1 December - Volume 59, Issue 12

Volume 59,
1 December 2016
ISSN 0010-4620
EISSN 1460-2067

Issue Navigation

Section A

ORIGINAL ARTICLES

A Transformation For Optimizing String-Matching Algorithms For Long Patterns

Minhaj Ahmad Khan

Abstract View article

Oxford Academic:

- moved to new platform
- Winter 2016 - Spring 2017
- gradually moving individual journals
- 3 content platforms in use at the same time

Sources of Raw Digital Data



EDM 2015
The 8th International Conference
on Educational Data Mining
26-29 June 2015
Madrid - Spain



You are here: Proceedings

Proceedings

This page holds the proceedings for the 8th International Conference on Educational Data Mining. The conference will be held on June 26 -2 9, 2015, in Madrid, Spain.

Organized by the International Educational Data Mining Society (IEDMS).

Comercial Sponsors

Gold



| | |
|------------------------|--|
| EDM2015 | |
| Proceedings | |
| Keynotes | |
| Panels | |
| Workshops & Tutorials | |
| Schedule | |
| Presenter Instructions | |
| Student Information | |
| Important Dates | |

Table of contents

Invited Talks (abstracts)

Behind the Scenes of Duolingo
Luis Von Ahn, Matt Streefer

Personal Knowledge/Learning Graph
George Siemens, Ryan Baker, Dragan Gasevic



EDM16
The 9th Intl. Conf. on
Educational Data Mining
June 29 - July 2, 2016
Raleigh
North Carolina, USA

Proceedings

The page holds the proceedings for the 9th International Conference on Educational Data Mining. The conference will be held on June 29 - July 2, 2016 in Raleigh, North Carolina, USA.

Individual papers

Invited Talks


Workshops & Tutorials

Schedule

Presenter Instructions

Student Information

Important Dates



EDM 2017
THE 10th INTERNATIONAL CONFERENCE
ON EDUCATIONAL DATA MINING
July 08 - 08, 2017
Miami, FL, USA

PROCEEDINGS

This page holds the proceedings for the 10th International Conference on Educational Data Mining. The conference will be held on June 21 - 26, 2017 in Miami, FL, USA.

Main Proceedings

Workshops & Tutorials

Invited Talks

Keynotes

Panels

Workshops & Tutorials

Schedule

Presenter Instructions

Student Information

Important Dates

ACADEMIC SPONSORS

Search box with 'Submit' button

Sources of Raw Digital Data

EDM16

The 9th Intl. Conf. on Educational Data Mining

June 29 – July 2, 2016
Raleigh
North Carolina, USA



Organized by the International Educational Data Mining Society (IEDMS).

Speakers

Keynotes
Industry Panel

Proceedings

Awards
Attendees

Proceedings

This page holds the proceedings for the 9th International Conference on Educational Data Mining. The conference will be held on June 29 - July 2, 2016, in Raleigh, North Carolina, USA.

Individual papers

Invited Talks

Data-Driven Education: Some opportunities and Challenges
Rakesh Agrawal

WSE Ways to Strengthen Inquiry Science Learning
Mercia Linn (presentation)

Enabling people to harness and control EDM for lifelong, life-wide learning
Judy Kay

Sponsors





EDM 2015

The 8th International Conference on Educational Data Mining

26-30 June 2015
#EDM15 - Spain



Organized by the International Educational Data Mining Society (IEDMS).

Proceedings

This page holds the proceedings for the 8th International Conference on Educational Data Mining. The conference will be held on June 26 - 30, 2015, in Madrid, Spain.

Invited Talks (abstracts)

Student Use Success of Challenge
Luigi De Raio, IEDM Director

Personal Knowledge Learning Graph
George Siemens, Ryan Baker, Gregor Stewin

Congressional Speakers





EDM 2014

THE 10th INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING

Madrid, Spain
June 21 - 26, 2014

PROCEEDINGS

This page holds the proceedings for the 10th International Conference on Educational Data Mining. The conference will be held on June 21 - 26, 2014, in Madrid, Spain.

Main Proceedings

Additional Proceedings

SEARCH

ACADEMIC SPONSORS

Sources of Raw Digital Data

The screenshot shows the IEEE Xplore Digital Library interface. At the top, there is a search bar with the text "Enter Search Term" and a "Search" button. Below the search bar are tabs for "Basic Search", "Author Search", "Publication Search", and "Advanced Search". The main content area displays "Browse Journals & Magazines > IEEE Transactions on Image Pro ... Volume 25 Issue 12". The journal title "IEEE Transactions on Image Processing" is prominently displayed. Below the title are navigation tabs: "Popular", "Early Access", "Current Issue", "Past Issues", and "About Journal". A "Submit Your Manuscript" button is also visible. The "Current Issue" tab is selected, showing "Issue 12 • Date Dec. 2016". A "Filter Results" section is active, displaying "Displaying Results 1 - 5 of 5". Under "Filter Results", there is a "Search within results" field and a "Select All Results" checkbox. A list of authors is shown under the "AUTHOR" heading, including Yanwei Pang, Zhanyi Hu, Hanqing Lu, Miao Yu, Xuelong Li, Jianqiang Li, Yue Gao, Jinqiao Wang, and Shuhan Shen. Two publications are listed:

- Dynamic Parallel and Distributed Graph Cuts**
Miao Yu ; Shuhan Shen ; Zhanyi Hu
Publication Year: 2016, Page(s):5511 - 5525
Abstract | PDF (2857 KB)
- Multi-View 3D Object Retrieval With Deep Embedding Network**
Haiyun Guo ; Jinqiao Wang ; Yue Gao ; Jianqiang Li ; Hanqing Lu
Publication Year: 2016, Page(s):5526 - 5537

A "Sponsor" section for the IEEE Signal Processing Society is also visible on the right side of the page.

The screenshot shows a search results page on IEEE Xplore. The search criteria are "2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)". The results are filtered to show "Displaying Results 51 - 75 of 85". A "Poster" section is highlighted, showing a list of publications:

- Big Brother is Watching You - Now in a Doublespangled Way**
Carry Starting; Curtis St. Pierre ; David Reinhardt
Publication Year: 2017, Page(s):1 - 2
Abstract | PDF (507 KB) | HTML

Navigation icons for back, forward, and search are visible at the bottom of the page.

Sources of Raw Digital Data

IEEE.org | IEEE Xplore Digital Library | IEEE-SA | IEEE Spectrum | More Sites Cart (0) | Create Account | Personal Sign In

IEEE Xplore®
Digital Library

Browse ▾ | **My Settings** ▾ | **Get Help** ▾

All ▾ Enter keywords or short phrases (searches metadata only by default)

Search within Publication Advanced Search | Other Search Options ▾

Browse Conference Publications > 2017 ACM/IEEE Joint Conference ...

2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)

19-23 June 2017

Filter Results | **Displaying Results 51 - 75 of 85** | Show: **25** ▾

Search within results:

Select All Results | | | |

SUBJECT CATEGORY

- Technical Paper (36)
- Poster (30)
- Tutorial (3)
- Workshop (3)
- Panel (1)

Poster

- Big Brother Is Watching You - Now in a Doubleplusgood Way**
Corey Sterling ; Carlin St. Pierre ; David Bainbridge
Publication Year: 2017, Page(s):1 - 2
[Abstract](#) | [PDF \(507 KB\)](#) | [HTML](#)

IEEE Xplore®
Digital Library

IEEE Transactions on Image Processing

Issue 12 • Date Dec. 2016

Filter Results | **Displaying Results 1 - 5 of 5**

Search within results:

AUTHOR

- Binak Peng (3)
- Zhang Peng (3)
- Hengyuan Li (3)
- Mao Xu (3)
- Fuhong Li (3)
- Jieping Li (3)
- Yan Sun (3)
- Jinyan Wang (3)
- Yueshan Wang (3)

Dynamic Parallel and Distributed Graph Cut
Ma X; Chakrabarti S; Zhang P
Publication Year: 2016, Page(s):5511 - 5525
[Abstract](#) | [PDF \(2857 KB\)](#)

Multi-View 3D Object Retrieval With Deep Embedding Network
Moghan G; Jinjie Wang ; Yan Gao ; Jiansong Han
Publication Year: 2016, Page(s):5526 - 5537

Sources of Raw Digital Data



ACL Anthology A Digital Archive of Research Papers in Computational Linguistics

Search the Anthology | via Google | via Searchbench @ DFKI | via AAN @ UMich | via Saffron @ Insight |

The ACL Anthology currently hosts over 42,000 papers on the study of computational linguistics and natural language processing. [Subscribe to the mailing list](#) to receive announcements and updates to the Anthology.

The [current version of the ACL Anthology](#) will replace this legacy version of the Anthology as the default version starting some time in 2017. Please start using this new service.

Do you love the Anthology? Not an ACL member yet? Please [join as an ACL member](#) to help keep the Anthology open for all to use.

NEW Sep 2017: The [Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing \(EMNLP 2017\)](#), as well as its 14 associated workshops and conferences are now all available in the ACL Anthology. Also, the [Proceedings of the Linguistic Resources for Automatic Natural Language Generation \(LIRA@NLG9\)](#) the [Proceedings of the Workshop on Computational Creativity in Natural Language Generation \(CC-NLG 2017\)](#), the [Proceedings of the 13th International Conference on Finite State Methods and Natural Language Processing \(FSM/NLP 2017\)](#) and [Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms \(TAG+13\)](#) are all now available on the ACL Anthology.

If you wish to submit your presentations or posters of your papers to be archived, please do by [emailing us a copy at this link](#).

ACL events

CL: [Intro](#) [FS](#) [MT&CL](#) [74](#) [79](#) [80](#) [81](#) [82](#) [83](#) [84](#) [85](#) [86](#) [87](#) [88](#) [89](#) [90](#) [91](#) [92](#) [93](#) [94](#) [95](#) [96](#) [97](#) [98](#) [99](#) [00](#) [01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#)

TACL: [13](#) [14](#) [15](#) [16](#) [17](#)

ACL: [Intro](#) [79](#) [80](#) [81](#) [82](#) [83](#) [84](#) [85](#) [86](#) [87](#) [88](#) [89](#) [90](#) [91](#) [92](#) [93](#) [94](#) [95](#) [96](#) [97](#) [98](#) [99](#) [00](#) [01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#)

EACL: [Intro](#) [83](#) [85](#) [87](#) [89](#) [91](#) [93](#) [95](#) [97](#) [99](#) [01](#) [03](#) [06](#) [09](#) [12](#) [14](#) [17](#)

NAACL: [Intro](#) [00](#) [01](#) [03](#) [04](#) [06](#) [07](#) [09](#) [10](#) [11](#) [12](#) [13](#) [15](#) [16](#)

EMNLP: [96](#) [97](#) [98](#) [99](#) [00](#) [01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [NEW](#) [17](#)

CoNLL: [97](#) [98](#) [99](#) [00](#) [01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#)

*Sem/
SemEval: [98](#) [01](#) [04](#) [07](#) [10](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#)

ANLP: [Intro](#) [83](#) [88](#) [92](#) [94](#) [97](#) [00](#)

Workshops: [77](#) [79](#) [81](#) [83](#) [85](#) [87](#) [89](#) [91](#) [93](#) [94](#) [95](#) [96](#) [97](#) [98](#) [99](#) [00](#) [01](#) [02](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [WORK](#) [17](#)

SIGs: ANN BIOMED DAT DIAL FSM GEN HAN HUM LEX MEDIA MOL MI NLP PARSE MORPHON SEM SLAV SEMITIC SLXPAT UR WAC

Other Events

COLING: [65](#) [67](#) [69](#) [73](#) [80](#) [82](#) [84](#) [86](#) [88](#) [90](#) [92](#) [94](#) [96](#) [98](#) [00](#) [02](#) [04](#) [06](#) [08](#) [10](#) [12](#) [14](#) [16](#)

HLT: [86](#) [89](#) [90](#) [91](#) [92](#) [93](#) [94](#) [01](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [15](#) [16](#)

IJCNLP: [05](#) [08](#) [09](#) [11](#) [13](#) [15](#)

LREC: [00](#) [02](#) [04](#) [06](#) [08](#) [10](#) [12](#) [14](#)

ALTA [Intro](#) [03](#) [04](#) [05](#) [06](#) [07](#) [08](#) [09](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#)

[16](#)

RANLP [09](#) [11](#) [13](#) [15](#)

The screenshot shows a search results page for 'EMNLP 2017'. It includes a table with columns for 'Year', 'ACL Anthology ID', and 'Title'. The table lists several papers from the 2017 conference, such as 'A Neural Language Model (EMNLP 2017)' and 'A Neural Language Model (EMNLP 2017)'. The table is partially obscured by a grid overlay.

Problem of Data Heterogeneity

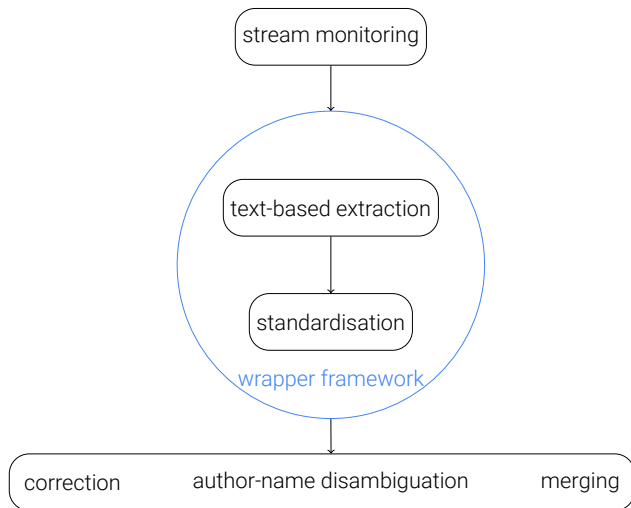
Sources of raw bibliographic data vary largely in quality and format, e.g.:

- website layouts
- change

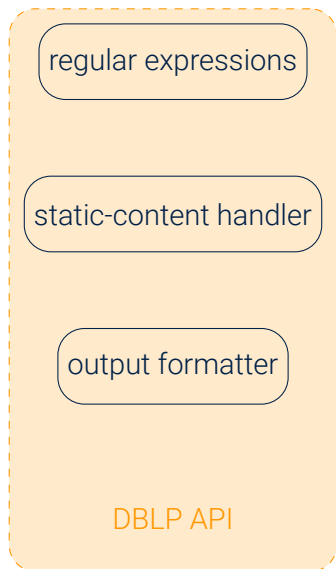
Problem of Data Integration

- question of feasibility: automated vs. manual harvesting
- expensive maintenance

DBLP as a Case Example



DBLP as a Case Example



Used across several steps, e.g.:

- publisher-key validation
- retrieving lists of issues
- retrieving tables of content
- retrieving records:

```
<tr[^>]*>. *?</tr>
```

DBLP as a Case Example

regular expressions

static-content handler

output formatter

DBLP API

XPath expressions

controllable browser

output-handler interface

XPath API

DBLP Author-Name Disambiguation

Disambiguating author names while inserting new raw publication record:

- simple, global search for matching names (entire database)
- reduce candidate set
- search for additional candidates, weighted additive factors:
 - co-author graph (similar names in vicinity)

Example: Co-Author Graphs

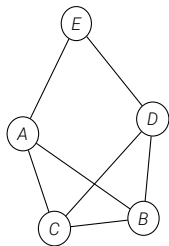


Figure 1: Co-Author Graph for dblp

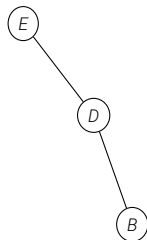


Figure 2: Partial graph for new record

Example: Co-Author Graphs

Disambiguating author names while inserting new raw publication record:

- simple, global search for matching names (entire database)
- reduce candidate set
- search for additional candidates, weighted additive factors:
 - co-author graph (similar names in vicinity)
 - stream activity (similar journal, conference)
 - publication year activity
 - publication topic
- rating of candidate set for each author of the new publication
- manual post-processing of ranked list of candidates with dblpi

Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 XPath**
- 4 Monitoring
- 5 Examples
- 6 Demonstration

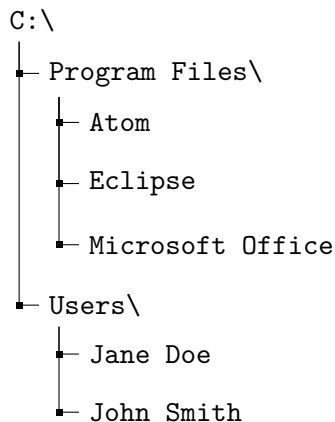
What Is OXPath?

- simple, declarative language for web data extraction
- XPath extension:
 - actions
 - iteration
 - extraction

What Is XPath?

- query language
- XML document as a tree of nodes
- XPath expressions as location paths

What Is XPath?



File-Path Examples

- 1 C:\Program Files\Microsoft Office
- 2 C:\Users\Jane Doe

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1/results/record/issue
```

Result Set

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1 /results/record/issue
```

Result Set

```
1 (
2   <issue>11</issue>,
3   <issue>10</issue>,
4   <issue>9</issue>
5 )
```

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1 /results/record/url/text()
```

Result Set

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1 /results/record/url/text()
```

Result Set

```
1 (
2   "http://.../toc/tadr20/30/11",
3   "http://.../toc/tadr20/30/10",
4   "http://.../toc/tadr20/30/9"
5 )
```

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1 /results/record[@class="current"]
```

Result Set

What Is XPath?

Queried XML File

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <record class="current">
4     <volume>30</volume>
5     <issue>11</issue>
6     <year>2016</year>
7     <url>http://.../tadr20/30/11</url>
8   </record>
9   <record>
10    <volume>30</volume>
11    <issue>10</issue>
12    <year>2016</year>
13    <url>http://.../tadr20/30/10</url>
14  </record>
15  <record>
16    <volume>30</volume>
17    <issue>9</issue>
18    <year>2016</year>
19    <url>http://.../tadr20/30/9</url>
20  </record>
21 </results>
```

XPath Expression

```
1 /results/record[@class="current"]
```

Result Set

```
1 (
2   <record class="current">
3     <volume>30</volume>
4     <issue>11</issue>
5     <year>2016</year>
6     <url>[...]</url>
7   </record>
8 )
```

What Does XPath Add?

Action:

- fill in forms
- click links, buttons, etc.

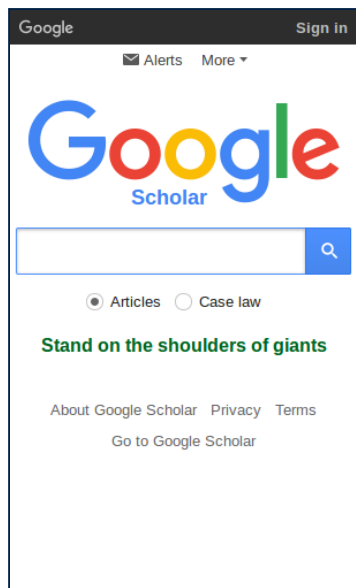
Extraction:

- add markers to extract selected nodes

Iteration:

- loops, e.g. for paginated content

Example: Navigating Google Scholar



The screenshot shows the Google Scholar homepage. At the top left is the 'Google' logo and 'Sign in' link. Below that are 'Alerts' and 'More' options. The main 'Google Scholar' logo is prominently displayed. A search bar is located below the logo, with a magnifying glass icon on the right. Underneath the search bar are radio buttons for 'Articles' (selected) and 'Case law'. The main heading 'Stand on the shoulders of giants' is shown in green. At the bottom, there are links for 'About Google Scholar', 'Privacy', and 'Terms', along with a 'Go to Google Scholar' button.

XPath Expression

```
1 doc('https://scholar.google.com')
```

Example: Navigating Google Scholar



The screenshot shows the Google Scholar homepage. At the top left is the 'Google' logo and 'Sign in' link. Below that are 'Alerts' and 'More' options. The main 'Google Scholar' logo is prominently displayed. A search bar contains the text 'XPath' with a magnifying glass icon to its right. Below the search bar are radio buttons for 'Articles' (selected) and 'Case law'. The main heading is 'Stand on the shoulders of giants'. At the bottom, there are links for 'About Google Scholar', 'Privacy', and 'Terms', along with a 'Go to Google Scholar' button.

XPath Expression

```
1 doc('https://scholar.google.com')  
2 //input[@id='gs_hdr_tsi']/{"XPath"}
```

Example: Navigating Google Scholar

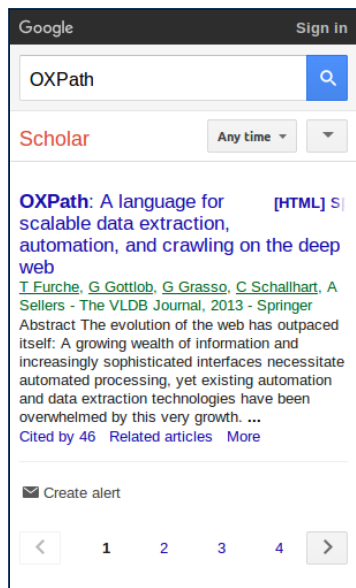


The screenshot shows the Google Scholar homepage. At the top left is the Google logo and 'Scholar' text. To the right are 'Alerts' and 'More' links. Below the logo is a search bar containing the text 'XPath'. To the right of the search bar is a magnifying glass icon. Below the search bar are radio buttons for 'Articles' (selected) and 'Case law'. Below that is the search result 'Stand on the shouldered of giants'. At the bottom are links for 'About Google Scholar', 'Privacy', and 'Terms', and a 'Go to Google Scholar' button.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{XPath}
3 ../following-sibling::button/{click/}
```

Example: Navigating Google Scholar

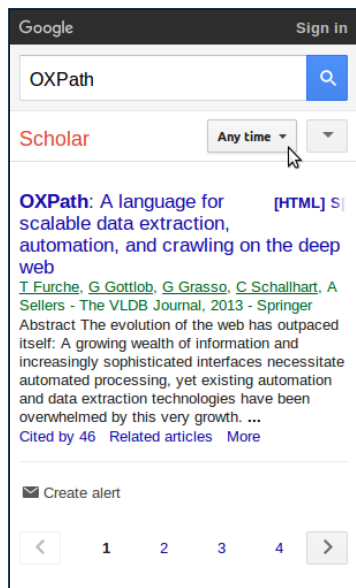


The screenshot shows the Google Scholar search interface. At the top, there is a search bar with the text "XPath" and a magnifying glass icon. To the right of the search bar is a "Sign in" link. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a dropdown menu set to "Any time". Below this, the search results for "XPath" are shown. The first result is a link to a page titled "XPath: A language for scalable data extraction, automation, and crawling on the deep web". The link is in blue and has "[HTML] S" next to it. Below the link is the author information: "T Furche, G Gottlob, G Grasso, C Schallhart, A Sellers - The VLDB Journal, 2013 - Springer". Below the author information is the abstract: "Abstract The evolution of the web has outpaced itself: A growing wealth of information and increasingly sophisticated interfaces necessitate automated processing, yet existing automation and data extraction technologies have been overwhelmed by this very growth. ...". Below the abstract are the statistics "Cited by 46" and links for "Related articles" and "More". At the bottom of the search results, there is a "Create alert" button with an envelope icon. Below the search results is a pagination bar with a left arrow, the numbers 1, 2, 3, 4, and a right arrow.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{"XPath"}
3 ../following-sibling::button/{click/}
```

Example: Navigating Google Scholar



The screenshot shows the Google Scholar search interface. At the top, there is a search bar containing the text "XPath" and a magnifying glass icon. To the right of the search bar is a "Sign in" link. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a dropdown menu currently set to "Any time". Below this, there is a search result for "XPath: A language for scalable data extraction, automation, and crawling on the deep web". The result includes the authors "T Furche, G Gottlob, G Grasso, C Schallhart, A Sellers" and the publication "The VLDB Journal, 2013 - Springer". Below the title and authors is an abstract snippet: "Abstract The evolution of the web has outpaced itself: A growing wealth of information and increasingly sophisticated interfaces necessitate automated processing, yet existing automation and data extraction technologies have been overwhelmed by this very growth. ...". At the bottom of the result are links for "Cited by 46", "Related articles", and "More". Below the result is a "Create alert" button. At the very bottom of the page, there is a pagination bar with a left arrow, the number "1" (highlighted), "2", "3", "4", and a right arrow.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{XPath}
3 ../following-sibling::button/{click}
4 //*[@id='gs_res_ab_yy-b']/{click}
```

Example: Navigating Google Scholar

The screenshot shows the Google Scholar search interface. At the top, there is a search bar containing the text "XPath" and a "Sign in" link. Below the search bar, the word "Scholar" is displayed in red. A dropdown menu is open, showing filter options: "Any time", "Since 2016", "Since 2015", "Since 2012", "Sort by relevance", and "Sort by date". The "Since 2016" option is highlighted with a mouse cursor. Below the menu, the search results for "XPath" are visible, including a snippet of text: "XPath: A scalable and automation web". At the bottom of the page, there is a "Create alert" button and a pagination bar showing page numbers 1, 2, 3, and 4.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{"XPath"}
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{"XPath"}
5 //following::*[@role='menuitemradio'][contains(.,
  '2016')]/{"XPath"}
```

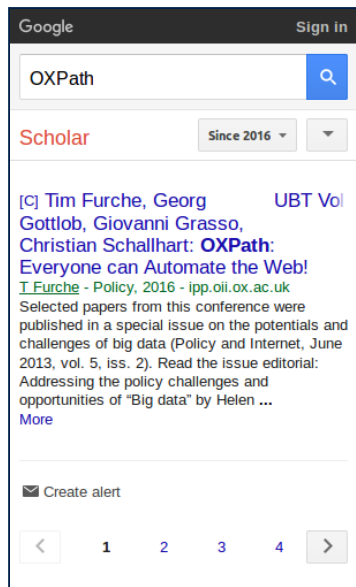
Example: Navigating Google Scholar

The screenshot shows the Google Scholar interface. At the top, there is a search bar with the text "XPath" and a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". The main content area shows a search result for "XPath: Everyone can Automate the Web!" by Tim Furche, Georg Gottlob, Giovanni Grasso, and Christian Schallhart. The result includes a link to a policy paper from 2016 and a summary of the paper's content. At the bottom of the result, there is a "Create alert" button and a pagination bar with page numbers 1, 2, 3, and 4.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{ "XPath" }
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{click/}
5 //following::*[@role='menuitemradio'][contains(.,
  '2016')]/{click/}
```

Example: Navigating Google Scholar



The screenshot shows the Google Scholar interface. At the top, there is a search bar with the text "OXPath" and a search button. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". The main content area shows a search result for a paper by Tim Furche, Georg Gottlob, Giovanni Grasso, and Christian Schallhart. The title is "OXPath: Everyone can Automate the Web!". Below the title is the author information "T Furche - Policy, 2016 - ipp.oii.ox.ac.uk". The abstract text follows: "Selected papers from this conference were published in a special issue on the potentials and challenges of big data (Policy and Internet, June 2013, vol. 5, iss. 2). Read the issue editorial: Addressing the policy challenges and opportunities of 'Big data' by Helen ...". At the bottom of the result, there is a "More" link and a "Create alert" button. A pagination bar at the very bottom shows page numbers 1, 2, 3, 4, with page 1 being the active page.

OXPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{OXPath}
3 ../following-sibling::button/{click}
4 /*[@id='gs_res_ab_yy-b']/{click}
5 //following::*[@role='menuitemradio'][contains(.,
6 //div[@class='gs_ri']//h3/a:<title=string.>)
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4 </results>
```


Example: Navigating Google Scholar

The screenshot shows the Google Scholar search interface. At the top, there is a search bar with the text "XPath" and a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". The search results for "XPath" are shown, with the top result by Tim Furche, Georg Gottlob, Giovanni Grasso, and Christian Schallhart. The title of the result is "XPath: Everyone can Automate the Web!". Below the title is a link to the paper: "T Furche - Policy, 2016 - ipp.oii.ox.ac.uk". A short abstract follows: "Selected papers from this conference were published in a special issue on the potentials and challenges of big data (Policy and Internet, June 2013, vol. 5, iss. 2). Read the issue editorial: Addressing the policy challenges and opportunities of 'Big data' by Helen ...". At the bottom of the result, there is a "More" link. Below the result, there is a "Create alert" button with an envelope icon. At the very bottom, there is a pagination bar with a left arrow, the number "1" (highlighted in blue), "2", "3", "4", and a right arrow.

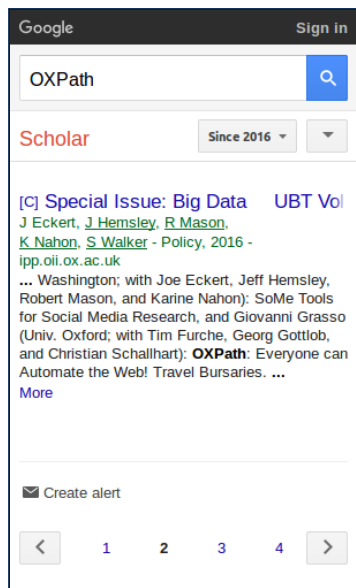
XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{ "XPath" }
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{click/}
5 //following::*[@role='menuitemradio'][contains(.,
6   '2016')]/{click/}
7 /(/*[@id='gs_nm']/button[2][not(@disabled)]/{click/})*
  //div[@class='gs_ri']//h3/a:<title=string(.)>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4 </results>
```

Example: Navigating Google Scholar



The screenshot shows the Google Scholar search interface. At the top, there is a search bar with the text "OXPath" and a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". Below this, a search result is shown for a "Special Issue: Big Data" edited by J Eckert, J Hemsley, R Mason, K Nahon, and S Walker. The snippet includes the text "OXPath: Everyone can Automate the Web! Travel Bursaries. ...". At the bottom of the search results, there is a "Create alert" button and a pagination bar with page numbers 1, 2, 3, and 4.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{OXPath}
3 ../following-sibling::button/{click}
4 /*[@id='gs_res_ab_yy-b']/{click}
5 //following::*[@role='menuitemradio'][contains(.,
6   '2016')]/{click}
7 /(//*[[@id='gs_nm']/button[2][not(@disabled)]/{click}]*/
  //div[@class='gs_ri']//h3/a:<title=string(.>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4 </results>
```

Example: Navigating Google Scholar

The screenshot shows the Google Scholar interface. At the top, there is a search bar with the text "OXPath" and a search button. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". Below this, a search result is shown for "[c] Special Issue: Big Data UBT Vol". The result includes the authors "J Eckert, J Hemsley, R Mason, K Nahon, S Walker" and the title "Policy, 2016 - ipp.oi.ox.ac.uk". A snippet of the abstract follows: "... Washington; with Joe Eckert, Jeff Hemsley, Robert Mason, and Karine Nahon): SoMe Tools for Social Media Research, and Giovanni Grasso (Univ. Oxford; with Tim Furche, Georg Gottlob, and Christian Schallhart): OXPath: Everyone can Automate the Web! Travel Bursaries. ...". A "More" link is visible below the snippet. At the bottom of the page, there is a "Create alert" button and a pagination bar with numbers 1, 2, 3, 4 and navigation arrows.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{OXPath}
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{click/}
5 //following::*[@role='menuitemradio'][contains(.,
6   '2016')]/{click/}
7 /(//*[[@id='gs_nm']/button[2][not(@disabled)]/{click/}]*/
8   //div[@class='gs_ri']//h3/a:<title=string(.>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4   <title>Special Issue: Big Data [...]</title>
5 </results>
```

Example: Navigating Google Scholar

The screenshot shows the Google Scholar search interface. At the top, the Google logo and 'Sign in' button are visible. The search bar contains the text 'OXPath' and a magnifying glass icon. Below the search bar, the word 'Scholar' is displayed in red. To the right of 'Scholar' is a filter dropdown menu set to 'Since 2016'. The search results list a paper by Eckert, Hemsley, Mason, Nahon, and Walker titled 'Special Issue: Big Data' and 'UBT Vol'. The abstract snippet mentions 'Washington; with Joe Eckert, Jeff Hemsley, Robert Mason, and Karine Nahon: SoMe Tools for Social Media Research, and Giovanni Grasso (Univ. Oxford; with Tim Furche, Georg Gottlob, and Christian Schallhart): OXPath: Everyone can Automate the Web! Travel Bursaries. ...'. A 'More' link is present below the snippet. At the bottom of the page, there is a 'Create alert' button and a pagination bar with page numbers 1, 2, 3, 4 and navigation arrows.

XPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{"OXPath"}
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{click/}
5 //following::*[@role='menuitemradio'][contains(.,
6   '2016')]/{click/}
7 /(//*[[@id='gs_nm']/button[2][not(@disabled)]/{click/}]*/
8   //div[@class='gs_ri']//h3/a:<title=string(.>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4   <title>Special Issue: Big Data [...]</title>
5   <!-- [...] -->
6 </results>
```

Example: Navigating Google Scholar

The screenshot shows the Google Scholar interface. At the top, there is a search bar with the text "OXPath" and a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red. To the right of "Scholar" is a filter dropdown menu set to "Since 2016". The main content area displays search results for the query. The first result is titled "Τεχνικές ανακάλυψης ενδιαφέρουσας πληροφορίας σε βάσεις δεδομένων ΚΑΜ Μπιλάλης, ΆΑΜ Γούναρης, ΣΑΜ Πεπελάσης - 2016 - repository.library.teimes.gr". Below the title, it says "Page 1. ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΤΕΧΝΙΚΕΣ ΑΝΑΚΑΛΥΨΗΣ ΕΝΔΙΑΦΕΡΟΥΣΑΣ ΠΛΗΡΟΦΟΡΙΑΣ ΣΕ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ...". There is a "More" link below the text. At the bottom of the search results, there is a "Create alert" button with an envelope icon. Below that is a pagination bar with a left arrow, the number "3", the number "4", the number "5", the number "6", and a right arrow.

OXPath Expression

```
1 doc('https://scholar.google.com')
2 //input[@id='gs_hdr_tsi']/{OXPath}
3 ../following-sibling::button/{click/}
4 /*[@id='gs_res_ab_yy-b']/{click/}
5 //following::*[@role='menuitemradio'][contains(.,
6   '2016')]/{click/}
7 /(/*[@id='gs_nm']/button[2][not(@disabled)]/{click/})*
8 //div[@class='gs_ri']//h3/a:<title=string(.)>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <title>Tim Furche, Georg Gottlob, [...]</title>
4   <title>Special Issue: Big Data [...]</title>
5   <!--[...]-->
6 </results>
```

Why XPath?

XPath

- static web
- plain HTML
- complete content

XPath

- dynamic web
- AJAX
- content on demand

How Will XPath Be Used?

Provide working environment:

- integrate existing tools for XPath
- collect XPath expressions prototypical of bibliographic domain
- devise and test XPath expressions
- explore use of XPath for stream monitoring

Tool support

Language plugin for Atom text editor

- Syntax highlighting for keywords
- Helps spotting errors and improves readability
- Intended to lower barriers for beginners

```
1 doc!https://www.cambridge.org/core/journals/robotica/issue/S001F08C4F1ICE84983101A0D390A8B0'
2 /**[@class="cookie-close"]{/click /}
3 //hr[!]/following:.*[@class="pagination"]|/!/[not(@class="unavailable")]/a[contains(., 'Next')]/(click)/}*
4 //div[data-prod-id]/li[@class="title"]<crecord
5 |./li[@class="title"]/a[@class="part-link"
6 |<titlenormalize-space.>
7 |<ee-fallback=concat('https://www.cambridge.org', @href)
8 |
9 |]
10 [? ./li[@class="author"]<authors-string-join./a[@class="more-by-this-author"], ', ')>
11 [? ./li[@class="source"]/a <doi=normalize-space.>]
12 [? ./li[@class="published"
13 |? ./li[@class="date"]<date=normalize-space.>]
14 [? ./following-sibling:.*[@class="pages"] <pages=replace., ".*(?!\w\d+(-\w\d+)?)", "$1">]
15 |
16 [? ./preceeding:~h4[!]<section=normalize-space.>]
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
```


Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 OXPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration

Monitoring

- Monitor harvesting process, notify in case of failures
→ similar to e.g. REPOX
- Identify incomplete or incorrect data
- Retrospective monitoring of own data pool
→ schedule harvesting

Ranking of Data Streams for Scheduling

Experiment on dblp: How can we rank all our conferences such that those most urgent for the next ingestion are on top?

Datasets:

- historical dblp data
- Microsoft Academic Graph
- Conference Ratings

Features:

- expected next entry and overdue measure
- average conference rating
- conference internationality
- average citation
- average author prominence

Ranking of Data Streams for Scheduling

Example data

Table 1: Example values for conference features used in the rankings, computed for December 2016.

| conf | overdue | rating | int. | disc. | citations | prom. |
|-------|---------|--------|------|-------|-----------|---------|
| jcdl | 15 | 3.5 | 5 | 3 | 5.9611 | 46.9888 |
| tpdl | 12 | 2.5 | 15 | 3 | 4.8133 | 52.9205 |
| icadl | 1 | 3 | 10 | 0 | 1.8881 | 52.2408 |
| dl | 48 | 0 | 1 | 38 | 18.4059 | 66.8803 |

Ranking of Data Streams for Scheduling

Evaluation:

- sliding window over months of 2016
- "gold standard" defined via interval-based pseudo-relevance
- for each month, ndcg is calculated for the produced ranking
- comparison of influence of the different features

Ranking of Data Streams for Scheduling

Comparison of ndcg values on different cut-offs. Statistical differences to the baseline tested with two-sided t-test ($*** = p < 0.001$, $** = p < 0.01$, $* = p < 0.05$).

| system | ndcg-10 | ndcg-20 | ndcg-100 | ndcg-200 |
|-------------------|-----------|-----------|-----------|-----------|
| delay (baseline) | 0,5270 | 0,5300 | 0,4954 | 0,4312 |
| +discontinued | 0,7299*** | 0,7011*** | 0,6341*** | 0,5892*** |
| +conf. rating | 0,7613*** | 0,7267*** | 0,6380*** | 0,5878*** |
| +internationality | 0,6520* | 0,6396* | 0,5979*** | 0,5695*** |
| +citations | 0,5730* | 0,5667 | 0,5468** | 0,5413*** |
| +prominence | 0,6718* | 0,6556** | 0,5956*** | 0,5683*** |

ACM News Windows

- metadata delivery might be unreliable (e.g. incomplete)
- observe several news windows,
 - last 2 weeks

Recently loaded issues and proceedings:

(available in the DL within the past 2 weeks)

Proceedings of the 10th International Conference on Security of Information and Networks

[SIN '17](#)

Proceedings of the 12th International Workshop on Variability Modelling of Software-Intensive Systems

[VAMOS 2018](#)

Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia

[MoMM2017](#)

Proceedings of the 1st Reversing and Offensive-oriented Trends Symposium

ACM News Windows

| Acronym | Proceeding Title | SIG Specimen | Year |
|--------------|--|--------------|------|
| ACM1994.12 | Proceedings of the 19th International Conference on Supercomputing | | 1994 |
| A-1997.09.02 | Proceedings of the 19th ACM SIGSOFT International Workshop on Automated Software Tools | WSOFT97 | 1997 |
| AMMIS.12 | Proceedings of the 18th Conference on Autonomous Agents and Multi-Agent Systems | | 2017 |
| AMAC.12 | Proceedings of the 18th ACM Workshop on Mobile Assisted Systems | MOBAC | 2017 |
| AC.19 | Proceedings of the 20th Australian Computer Education Conference | | 2019 |
| ACSE.12 | Proceedings of the Fourth International Conference on Agent-Centered Interaction | | 2012 |
| ACM-SE.11 | Proceedings of the SouthEast Conference | | 2011 |
| ACM-TLAC.12 | Proceedings of the ACM Tenth Latin American Conference - Chile | | 2012 |
| ACM-ICSE.12 | Proceedings of the 34th ACM International Conference on Software Engineering, Companion Volume and Lecture Notes | SEPO | 2012 |
| ACMCC.2012 | Proceedings of the 30th Annual Computer Security Applications Conference | | 2012 |
| ACMCS.12 | Proceedings of the 2012 Australian Document Conference | | 2012 |
| ACMSE.12 | Proceedings of the 2012 ACM SIGPLAN International Workshop on Programming Based on Actors, Agents, and Distributed Cms | | 2012 |
| ACM-12 | Proceedings of the 30th Australian Supercomputing Conference | | 2012 |
| AMC.12 | Proceedings of the 2012 International Conference on Artificial Intelligence, Automation and Control Technologies | | 2012 |
| AMC.12 | Proceedings of the 2012 Annual Computer Security Applications Conference | | 2012 |
| AM.12 | Proceedings of the Advances in Robotics | | 2012 |
| AMM.12 | Proceedings of the 18th ACM Workshop on AI Work Informatics and Metadata | | 2012 |

- metadata delivery might be unreliable (e.g. incomplete)
- observe several news windows, unreliable as well
 - last 2 weeks
 - last 12 months

ACM News Windows

Searched for "workflows" (refined count) [advanced search]

Found The ACM Full-Text Collection: 483,312 records Expand your search to The ACM Queue to Computing Literature: 2,747,894 records

Refinements (remove all) click each refinement below to refine

Published since: 2017

ACM Publications: Proceedings

21,618 results found Export Results: HTML | Endnote | CSV

ACM News Window Result 1 - 20 of 21,619 Result page 1 2 3 4 5 6 7 8 9 10 11

Sort by: publication date

Refine by People: Names, Institutions, Authors, Subjects

Refine by Publications: Publication Titles, ACM Publications, All Publications, Content Formats, Publication

Refine by Conferences: Conferences, Events, Proceeding Series

Refine by Publication Year: 2017 2016 Published Since 2017

- Unsupervised Workflow Extraction from First Person Videos of Mechanical Assembly**
Tung-kuo Pao, Yu-Hao Feiyang 2018 *HotMobile '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*
Publication: ACM
Eid: 3486636
Recently Augmented Reality (AR) applications have proved to help improve the efficiency in accomplishing assembly tasks. However, due to the lack of approaches to automate workflow extraction, the existing AR-based assembly assistance applications require manual authoring, which hampers scalability. Moreover, most of these applications only support information visualization and delivery...
Keywords: video analytics, workflow extraction
- Internet-of-Things (IoT) Device Pairing through Heterogeneous Sensing Signals**
Hong-Pan, Guo-Ping, Jun-Han, Adnan Barakat, Patrick Tigger, Han-Yang, Liu, Shi-Zhong Feiyang 2018 *HotMobile '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*
Publication: ACM
Eid: 3486637
Early establishing pairing between Internet of Things (IoT) devices is important for fast deployment in many smart home scenarios. Traditional pairing methods, including pairing QR code and NFC, often require specific user interaction, such as alignment, or additional appendages. The growing number of low-cost IoT devices without an interface may not meet these...
Keywords: heterogeneous sensing, internet-of-things, pairing
- CARs Collaborative Augmented Reality for Socialization**
Wen-xiao Zhang, Bo-Miao, Pan-Hui, Yiqun, Depanar-Srinani, Eric Zaverly, Peng-Qian Feiyang 2018 *HotMobile '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications*
Publication: ACM

- metadata delivery might be unreliable (e.g. incomplete)
- observe several news windows, unreliable as well
 - last 2 weeks
 - last 12 months
 - last 3 months

Searched for "tree search" [edit search] [advanced search] [advanced search]

Searched The ACM Full-Text Collection: 483,312 records Expand your search to The ACM Queue to Computing Literature: 2,747,894 records

Notifications [remove all] click each notification below to receive
Published since: 2017
ACM Publications: Proceeding

21418 results found Export Results: [html] [endnote] [xml] [pdf]

ACM news board Result 1 - 20 of 21,418 Result page 1 2 3 4 5 6 7 8 9 10 11 Sort by: [publication date]

Refine by People
Names
Institutions
Authors
Addresses

Refine by Publications
Publication Types
ACM Publications
All Publications
Content Elements
Publication

Refine by Conferences
Sponsors
Events
Proceeding Series

Refine by Publication Year

2017 2016
Published Since 2017

1 **Unsupervised Workflow Extraction from First Person Videos of Mechanical Assembly**
Tung-Hsiang Yu, Hsin-Feng
February 2018 Hablestick '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications
Publisher: ACM
Eid: 10.1145/3180000
Recently, Augmented Reality (AR) applications have proved to help improve the efficiency in accomplishing assembly tasks. However, due to the lack of approaches to automate workflow extraction, the existing AR-based assembly assistance applications require manual authoring, which hampers scalability. Moreover, most of these applications only support information visualization and delivery...
Keywords: video analytics, workflow extraction

2 **UniverSonic: IoT Device Posing through Heterogeneous Sensing Signals**
Jing Fan, Guohua Hu, Jun Han, Adnan Bashir, Patrick Tigger, Hao Young Lee, Peng Zhang
February 2018 Hablestick '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications
Publisher: ACM
Eid: 10.1145/3180000
Early establishing posing between Internet of Things (IoT) devices is important for fast deployment in many smart home scenarios. Traditional posing methods, including poseNet, QR code, and RFID, either require specific user interfaces, outdoor's arrangement, or additional appliances. The growing number of low-cost IoT devices without an interface may not meet these...
Keywords: heterogeneous sensing, internet of things, posing

3 **CARS: Collaborative Augmented Reality for Socialization**
Weizao Zhang, Bin Han, Pan Hai, Yiqun Ding, Xianbin Shi, Eric Zaveri, Peng Qian
February 2018 Hablestick '18: Proceedings of the 18th International Workshop on Mobile Computing Systems and Applications
Publisher: ACM

XPath Expression

```

1 doc('http://dl.acm.org/advsearch.cfm')
2 /**[@id='fld0']/optgroup[1]/option[5]/{click /}
3 /**[@id='how0']/option[4]/{click /}
4 /**[@id='dte0']/option[@value='${start.year}']/{click /}
5 /**[@id='submit']/input/{click /}
6 /**[@id='ACM_Publications_list']//a[contains(,
7 'Proceeding')]/{click /}
8 /**[@id='sortmenu']//option[@value='publicationDate']
9 /{click /}
10 /(.not(../**[@class='publicationDate' and
11 contains(, '$poison.pill')]))]
12 /**[@class="pagelogic"][1]/span[./strong]
13 /following-sibling::span[./a][1]
14 /{click[wait=5] /}*
15 /**[@id='results']/div[contains(@class,
16 'details')]:<record>
17 [? ../div[@class='source']/span[2]
18 :<header=normalize-space(.)>
19 :<title=substring-after(normalize-space(.), ':
20 ')>
21 :<conference=substring-before(normalize-space(.),
22 ': ')>
23 ]
24 [? ../div[@class='title']//a
25 :<section=normalize-space(.)>
26 :<details=qualify-url(@href)>

```

Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 XPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration

EDM 2014: Simple Extraction

EDM 2014

Proceedings

Citation Information
Stumpert, J., Radon, Z., Mavrikis, M., McLaren, B.M. (eds.) Proceedings of the 7th International Conference on Educational Data Mining

Online Proceedings
Click [here](#) to download a PDF file of the full proceedings.

Full Papers
Adaptive Practice of Facts in Domains with Varied Prior Knowledge
Jan Popowicz, Radik Peineke and VV Stanislav
Pages 6-13 [pdf]
Alternating Recursive Method for Q-matrix Learning
Yuan Sun, Shihwei He, Shunyu Inoue and Yi Sun
Pages 14-20 [pdf]

Organized by the International Educational Data Mining Society (IEDMS)

Sponsors:
Gold
Carnegie Learning
MARI
PEARSON

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>  
2 <results>  
1 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]-->
3 <div id="content">
4 <!--[...]-->
5 <strong>Online Proceedings</strong>
6 <!--[...]-->
7 <strong>Full Papers</strong>
8 <!--[...]-->
9 <p>Adaptive Practice of [...]
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!--[...]-->
17 </p>
18 <!--[...]-->
19 </div>
20 <!--[...]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
1 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]-->
3 <div id="content">
4 <!--[...]-->
5 <strong>Online Proceedings</strong>
6 <!--[...]-->
7 <strong>Full Papers</strong>
8 <!--[...]-->
9 <p>Adaptive Practice of [...]  
10 <br/>  
11 <em> Jan Papousek, [...]</em>  
12 <br/>  
13 Pages 6-13 [  
14 <a href="uploads/[...].pdf">pdf</a>  
15 ]  
16 <!--[...]-->  
17 </p>  
18 <!--[...]-->  
19 </div>  
20 <!--[...]-->  
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <record></record>
4 <record></record>
5 <!--[...]-->
6 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]-->
3 <div id="content">
4 <!--[...]-->
5 <strong>Online Proceedings</strong>
6 <!--[...]-->
7 <strong>Full Papers</strong>
8 <!--[...]-->
9 <p>Adaptive Practice of [...]  
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [  
14 <a href="uploads/[...].pdf">pdf</a>  
15 ]  
16 <!--[...]-->
17 </p>
18 <!--[...]-->
19 </div>
20 <!--[...]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--[...]-->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 </record>
7 <!--[...]-->
8 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]->
3 <div id="content">
4 <!--[...]->
5 <strong>Online Proceedings</strong>
6 <!--[...]->
7 <strong>Full Papers</strong>
8 <!--[...]->
9 <p>Adaptive Practice of [...]
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!--[...]->
17 </p>
18 <!--[...]->
19 </div>
20 <!--[...]->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3   [./em:<authors=string(.)>]
4   [./text()[1]:<title=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--[...]->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 </record>
8 <!--[...]->
9 </results>
```


EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]->
3 <div id="content">
4 <!--[...]->
5 <strong>Online Proceedings</strong>
6 <!--[...]->
7 <strong>Full Papers</strong>
8 <!--[...]->
9 <p>Adaptive Practice of [...]
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!--[...]->
17 </p>
18 <!--[...]->
19 </div>
20 <!--[...]->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3   [./em:<authors=string(.)>]
4   [./text()[1]:<title=string(.)>]
5   [./br[2]/following-sibling::text()[1]
6     :<pages=substring-after(., "Pages ")>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--[...]->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 <pages>6-13 [</pages>
8 </record>
9 <!--[...]->
10 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!-- [...] -->
3 <div id="content">
4 <!-- [...] -->
5 <strong>Online Proceedings</strong>
6 <!-- [...] -->
7 <strong>Full Papers</strong>
8 <!-- [...] -->
9 <p>Adaptive Practice of [...]  
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [  
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!-- [...] -->
17 </p>
18 <!-- [...] -->
19 </div>
20 <!-- [...] -->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3   [./em:<authors=string(.)>]
4   [./text()[1]:<title=string(.)>]
5   [./br[2]/following-sibling::text()[1]
6     :<pages=substring-after(., "Pages ")>]
7   [./a:<url=string(@href)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!-- [...] -->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 <pages>6-13 [</pages>]
8 <url>uploads/[...].pdf</url>
9 </record>
10 <!-- [...] -->
11 </results>
```

EDM 2014: Simple Extraction

HTML Source

```
1 <html xmlns="["...]" xml:lang="en">
2 <!--["..."]-->
3 <div id="content">
4 <!--["..."]-->
5 <strong>Online Proceedings</strong>
6 <!--["..."]-->
7 <strong>Full Papers</strong>
8 <!--["..."]-->
9 <p>Adaptive Practice of [...]  
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [  
14 <a href="uploads/["..."].pdf">pdf</a>
15 ]
16 <!--["..."]-->
17 </p>
18 <!--["..."]-->
19 </div>
20 <!--["..."]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=string(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]
6 :<pages=substring-after(., "Pages ")>]
7 [./a:<url=string(@href)>]
8 [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--["..."]-->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 <pages>6-13 [</pages>]
8 <url>uploads/["..."].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--["..."]-->
12 </results>
```

EDM 2014: Advanced Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]-->
3 <div id="content">
4 <!--[...]-->
5 <strong>Online Proceedings</strong>
6 <!--[...]-->
7 <strong>Full Papers</strong>
8 <!--[...]-->
9 <p>Adaptive Practice of [...]  
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [  
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!--[...]-->
17 </p>
18 <!--[...]-->
19 </div>
20 <!--[...]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=string(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]
6 :<pages=substring-after(., "Pages ")>]
7 [./a:<url=string(@href)>]
8 [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--[...]-->
4 <record>
5 <authors> Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 <pages>6-13 [</pages>]
8 <url>uploads/[...].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--[...]-->
12 </results>
```

EDM 2014: Advanced Extraction

HTML Source

```
1 <html xmlns="["...]" xml:lang="en">
2 <!--["..."]-->
3 <div id="content">
4 <!--["..."]-->
5 <strong>Online Proceedings</strong>
6 <!--["..."]-->
7 <strong>Full Papers</strong>
8 <!--["..."]-->
9 <p>Adaptive Practice of [...]  
10 <br/>
11 <em> Jan Papousek, [...]</em>
12 <br/>
13 Pages 6-13 [  
14 <a href="uploads/["..."].pdf">pdf</a>  
15 ]  
16 <!--["..."]-->
17 </p>
18 <!--["..."]-->
19 </div>
20 <!--["..."]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=normalize-space(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]  
 :<pages=substring-after(., "Pages ")>]
6 [./a:<url=string(@href)>]
7 [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--["..."]-->
4 <record>
5 <authors>Jan Papousek, [...]</authors>
6 <title>Adaptive Practice of [...]</title>
7 <pages>6-13 [</pages>]
8 <url>uploads/["..."].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--["..."]-->
12 </results>
```

EDM 2014: Advanced Extraction

HTML Source

```
1 <html xmlns="[...]" xml:lang="en">
2 <!--[...]-->
3 <div id="content">
4 <!--[...]-->
5 <strong>Online Proceedings</strong>
6 <!--[...]-->
7 <strong>Full Papers</strong>
8 <!--[...]-->
9 <p>Adaptive Practice of [...>
10 <br/>
11 <em> Jan Papousek, [...</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/[...].pdf">pdf</a>
15 ]
16 <!--[...]-->
17 </p>
18 <!--[...]-->
19 </div>
20 <!--[...]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=normalize-space(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]
6 :<pages=replace(normalize-space(.),
7 ".*?(\\d+(-\\d+)?.)*", "$1">]
8 [./a:<url=string(@href)>]
9 [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--[...]-->
4 <record>
5 <authors>Jan Papousek, [...</authors>
6 <title>Adaptive Practice of [...</title>
7 <pages>6-13</pages>
8 <url>uploads/[...].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--[...]-->
12 </results>
```

EDM 2014: Advanced Extraction

HTML Source

```
1 <html xmlns="["...]" xml:lang="en">
2 <!--["..."]-->
3 <div id="content">
4 <!--["..."]-->
5 <strong>Online Proceedings</strong>
6 <!--["..."]-->
7 <strong>Full Papers</strong>
8 <!--["..."]-->
9 <p>Adaptive Practice of ["..."]
10 <br/>
11 <em> Jan Papousek, ["..."]</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/["..."].pdf">pdf</a>
15 ]
16 <!--["..."]-->
17 </p>
18 <!--["..."]-->
19 </div>
20 <!--["..."]-->
21 </html>
```

XPath Expression

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=normalize-space(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]
6 :<pages=replace(normalize-space(.),
7 ".*?(\\d+(-\\d+)?.)*", "$1">]
8 [./a:<url=qualify-url(@href)>]
9 [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--["..."]-->
4 <record>
5 <authors>Jan Papousek, ["..."]</authors>
6 <title>Adaptive Practice of ["..."]</title>
7 <pages>6-13</pages>
8 <url>http://["..."]uploads/["..."].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--["..."]-->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source (EDM 2014)

```
1 <html xmlns="["...]" xml:lang="en">
2 <!--["..."]-->
3 <div id="content">
4 <!--["..."]-->
5 <strong>Online Proceedings</strong>
6 <!--["..."]-->
7 <strong>Full Papers</strong>
8 <!--["..."]-->
9 <p>Adaptive Practice of [...>
10 <br/>
11 <em> Jan Papousek, [...</em>
12 <br/>
13 Pages 6-13 [
14 <a href="uploads/["..."].pdf">pdf</a>
15 ]
16 <!--["..."]-->
17 </p>
18 <!--["..."]-->
19 </div>
20 <!--["..."]-->
21 </html>
```

XPath Expression (EDM 2014)

```
1 doc('http://edm2014.org/?page=proceedings')
2 /**[@id='content']/p[./em]:<record>
3 [./em:<authors=string(.)>]
4 [./text()[1]:<title=string(.)>]
5 [./br[2]/following-sibling::text()[1]
6 :<pages=substring-after(., "Pages ")>]
7 [./a:<url=string(@href)>]
8 [./preceding::strong[1]:<header=string(.)>]
```

XML Output (EDM 2014)

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 <!--["..."]-->
4 <record>
5 <authors> Jan Papousek, [...</authors>
6 <title>Adaptive Practice of [...</title>
7 <pages>6-13 [</pages>
8 <url>uploads/["..."].pdf</url>
9 <header>Full Papers</header>
10 </record>
11 <!--["..."]-->
12 </results>
```


EDM 2016: Adapting from EDM 2014

| | | |
|--|--|---|
| EDM16 The 9th Intl. Conf. on Educational Data Mining | |  |
| June 29 – July 2, 2016 Raleigh North Carolina, USA | | |
| EDM2016 | Proceedings | Organized by the International Educational Data Mining Society (IEDMS) |
| Speakers | This page holds the proceedings for the 9th International Conference on Educational Data Mining. The conference will be held on June 29 – July 2, 2016, in Raleigh, North Carolina, USA. | Sponsors |
| Keynotes | Individual papers |  |
| Industry Panel | Invited Talks |  |
| Proceedings | Data-Driven Education: Some opportunities and Challenges Rakesh Agrawal |  |
| Awards | WISE Ways to Strengthen Inquiry Science Learning Marisa Linn (presentation) | Blackboard |
| Attendees | Enabling people to harness and control EDM for lifelong, life-wide learning July Kay |  |

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2  /**[@id='content']/p[./em]:<record>
3  [./em:<authors=string(.)>]
4  [./text()[1]:<title=string(.)>]
5  [./br[2]/following-sibling::text()[1]
6   :<pages=substring-after(., "Pages ")>]
7  [./a:<url=string(@href)>]
8  [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3 </results>
```

EDM 2016: Adapting from EDM 2014

| | | |
|--|--|---|
| EDM16 The 9 th Intl. Conf. on Educational Data Mining | |  |
| June 29 – July 2, 2016 Raleigh North Carolina, USA | | |
| EDM2016 | Proceedings | Organized by the International Educational Data Mining Society (IEDMS) |
| Speakers | This page holds the proceedings for the 9th International Conference on Educational Data Mining. The conference will be held on June 29 - July 2, 2016, in Raleigh, North Carolina, USA. | Sponsors |
| Keynotes | Individual papers |  |
| Industry Panel | Invited Talks |  |
| Proceedings | Data-Driven Education: Some opportunities and Challenges Rakesh Agrawal |  |
| Awards | WISE Ways to Strengthen Inquiry Science Learning Marisa Linn (presenter) | Blackboard |
| Attendees | Enabling people to harness and control EDM for lifelong, life-wide learning July Klay |  |

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 /**[@id='content']/p[./em]:<record>
3   [./em:<authors=string(.)>]
4   [./text()[1]:<title=string(.)>]
5   [./br[2]/following-sibling::text()[1]
6     :<pages=substring-after(., "Pages ")>]
7   [./a:<url=string(@href)>]
8   [./preceding::strong[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>???
```

EDM 2016: Adapting from EDM 2014

EDM16

The 9th Intl. Conf. on Educational Data Mining

June 29 – July 2, 2016
Raleigh
North Carolina, USA



| | | |
|--------------------|--|---|
| EDM2016 | Proceedings | Organized by the International Educational Data Mining Society (IEDMS) |
| Speakers | This page holds the proceedings for the 9th International Conference on Educational Data Mining. The conference will be held on June 29 - July 2, 2016, in Raleigh, North Carolina, USA. | Sponsors |
| Keynotes | Individual papers |  |
| Industry Panel | Invited Talks |  |
| Proceedings | Data-Driven Education: Some opportunities and Challenges Rakesh Agrawal |  |
| Awards | WISE Ways to Strengthen Inquiry Science Learning Marisa Linn (presenter) | Blackboard |
| Attendees | Enabling people to harness and control EDM for lifelong, life-wide learning July Kay |  |

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2   ???:<record>
3     [???:<authors=???>]
4     [???:<title=???>]
5     [???:<pages=???>]
6     [???:<url=???>]
7     [???:<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!-- [...] -->
4   <record>
5     <authors>???
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!--[...]-->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <!--[...]-->
9   <span class="[...]" title">9th [...]</span>
10  <span class="[...]" firstpage">2</span>
11  <span class="[...]" lastpage">2</span>
12  <span class="[...]" pdf_url">http[...]</span>
13  <br/>
14  <span class="[...]" author">Ra[...]</span>
15  <!--[...]-->
16 </p>
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2   ???<:record>
3     [???<:authors=???>]
4     [???<:title=???>]
5     [???<:pages=???>]
6     [???<:url=???>]
7     [???<:header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>???</authors>
6     <title>???</title>
7     <pages>???</pages>
8     <url>???</url>
9     <header>???</header>
10  </record>
11  <!--[...]-->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!--[...]-->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <span class="[...]title">9th [...]</span>
9   <span class="[...]firstpage">2</span>
10  <span class="[...]lastpage">2</span>
11  <span class="[...]pdf_url">http[...]</span>
12  <br/>
13  <span class="[...]author">Ra[...]</span>
14  <!--[...]-->
15 </p>
16 <!--[...]-->
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[contains(@class, 'cit')]:<record>
3   [???:<authors=???>]
4   [???:<title=???>]
5   [???:<pages=???>]
6   [???:<url=???>]
7   [???:<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>???</authors>
6     <title>???</title>
7     <pages>???</pages>
8     <url>???</url>
9     <header>???</header>
10  </record>
11  <!--[...]-->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!--[...]-->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <!--[...]-->
9   <span class="[...]" title="9th [...]">9th [...]</span>
10  <span class="[...]" firstpage="2">2</span>
11  <span class="[...]" lastpage="2">2</span>
12  <span class="[...]" pdf_url="http[...]">http[...]</span>
13  <span class="[...]" author="Ra[...]">Ra[...]</span>
14  <!--[...]-->
15 </p>
16 <!--[...]-->
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[@class='cit']]:<record>
3   [./*[@class='author']]:<authors=string(.)>]
4   [???:<title=???>]
5   [???:<pages=???>]
6   [???:<url=???>]
7   [???:<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>Rakesh Agrawal</authors>
6     <title>???</title>
7     <pages>???</pages>
8     <url>???</url>
9     <header>???</header>
10  </record>
11  <!--[...]-->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!-- [...] -->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <span class="[...]" title="9th [...]">9th [...]</span>
9   <span class="[...]" firstpage="2">2</span>
10  <span class="[...]" lastpage="2">2</span>
11  <span class="[...]" pdf_url="http[...]">http[...]</span>
12  <br/>
13  <span class="[...]" author="Ra[...]">Ra[...]</span>
14  <!-- [...] -->
15 </p>
16 <!-- [...] -->
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[@class='cit']]:<record>
3   [./*[@class='author']:<authors=string(.)>]
4   [./*[@class='title']:<title=string(.)>]
5   [???<pages=???>]
6   [???<url=???>]
7   [???<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!-- [...] -->
4   <record>
5     <authors>Rakesh Agrawal</authors>
6     <title>Data-Driven [...]</title>
7     <pages>???</pages>
8     <url>???</url>
9     <header>???</header>
10  </record>
11  <!-- [...] -->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!--[...]-->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <!--[...]-->
9   <span class="[...]" title">9th [...]</span>
10  <span class="[...]" firstpage">2</span>
11  <span class="[...]" lastpage">2</span>
12  <span class="[...]" pdf_url">http[...]</span>
13  <br/>
14  <span class="[...]" author">Ra[...]</span>
15  <!--[...]-->
16 </p>
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[@class='cit']]:<record>
3   [.*[@class='author']:<authors=string(.)>]
4   [.*[@class='title']:<title=string(.)>]
5   [.:<pages=concat(./.*[@class='firstpage'],
6     '-', ./.*[@class='lastpage'])>]
7   [???:<url=???>]
8   [???:<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>Rakesh Agrawal</authors>
6     <title>Data-Driven [...]</title>
7     <pages>2-2</pages>
8     <url???
```


EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!--[...]-->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <!--[...]-->
9   <span class="[...]" title">9th [...]</span>
10  <span class="[...]" firstpage">2</span>
11  <span class="[...]" lastpage">2</span>
12  <span class="[...]" pdf_url">http[...]</span>
13  <br/>
14  <span class="[...]" author">Ra[...]</span>
15  <!--[...]-->
16 </p>
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[@class='cit']]:<record>
3   [./*[@class='author']:<authors=string(.)>]
4   [./*[@class='title']:<title=string(.)>]
5   [.:<pages=concat(./*[@class='firstpage'],
6     '-', ./*[@class='lastpage'])>]
7   [./*[@class='url']:<url=string(.)>]
8   [???:<header=???>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!--[...]-->
4   <record>
5     <authors>Rakesh Agrawal</authors>
6     <title>Data-Driven [...]</title>
7     <pages>2-2</pages>
8     <url>http://[...].pdf</ee>
9     <header>???</header>
10  </record>
11  <!--[...]-->
12 </results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
1 <html>
2 <!-- [...] -->
3 <h1>Individual papers</h1>
4 <h3>Invited Talks</h3>
5 <p>
6   <a id="[...]" class="citation_title"
7     href="[...]">Data-Driven [...]</a>
8   <span class="[...]" title="9th [...]">9th [...]</span>
9   <span class="[...]" firstpage="2">2</span>
10  <span class="[...]" lastpage="2">2</span>
11  <span class="[...]" pdf_url="http:[...]">http:[...]</span>
12  <br/>
13  <span class="[...]" author="Ra[...]">Ra[...]</span>
14  <!-- [...] -->
15 </p>
16 <!-- [...] -->
17 </html>
```

XPath Expression

```
1 doc('http://edm2016.org/proceedings.html')
2 //p[./*[@class='cit']]:<record>
3   [./*[@class='author']:<authors=string(.)>]
4   [./*[@class='title']:<title=string(.)>]
5   [.:<pages=concat(./*[@class='firstpage'],
6     '-', ./*[@class='lastpage'])>]
7   [./*[@class='url']:<url=string(.)>]
8   [./preceding::h3[1]:<header=string(.)>]
```

XML Output

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <results>
3   <!-- [...] -->
4   <record>
5     <authors>Rakesh Agrawal</authors>
6     <title>Data-Driven [...]</title>
7     <pages>2-2</pages>
8     <url>http://[...].pdf</ee>
9     <header>Invited Talks</header>
10  </record>
11  <!-- [...] -->
12 </results>
```

Table of Contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 OXPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration

Demonstration: Researcher ID

The screenshot shows the ResearcherID profile for Peter Adami. The profile includes the following information:

- ResearcherID:** I-2057-2013
- Other Name:** Adami, P.
- URL:** <http://www.researcherid.com/rid/2057-2013>
- Subject:** Evolutionary Biology
- Keywords:** corollary ecology, wildlife ecology, phenology
- ORCID:** <http://orcid.org/0009-0003-3364-1334>
- My Institutions (from default):** Palacky University
- Sub-org(s):** Researcher (Academic)
- Just Affiliation:** Museum of Natural History

Below the profile information, there is a section for "My Publications: View" which shows a list of publications. The first publication is:

1. Title: Changes in spring arrival dates and temperature sensitivity of migratory birds over two centuries
Author(s): Robinson, C., Mello, M., Menni, A., et al.
Source: International Journal of Biometeorology Volume: 61 Issue: 7 Pages: 1279-1290 Published: 2017
Times Cited: 0
DOI: 10.1007/s00484-017-1305-0

XPath Expression

```
1 doc("http://www.researcherid.com/rid/I-2057-2013")
2  /**[@id='resultContainer']/table[1]
3    /**[@id='resultsPerPage']/option[3]/{click /}
4  /{//following::*[@id='resultContainer']/table[1]//img
5    [alt="Next Page"]
6    [onclick="showNextPage()"]/{click /}*
7  //following::*[@id='resultContainer']:<publications>
8  [./table[2]/tbody/tr/td[2]:<publication>
9  [.:<title=substring-before(
10     substring-after(normalize-space(.), "Title: ",
11     " Author(s):")>]
12  [.:<auSoPaPu=concat("Author(s): ",
13     substring-before(
14     substring-after(normalize-space(.), "Author(s):
15     ", " DOI:")>]
16  [.:<doi=substring-after(normalize-space(.), "DOI:
17     ")>]
18 ]
```

Demonstration: Booking

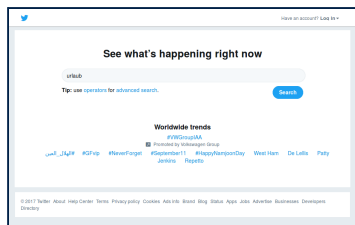
The screenshot shows the Booking.com interface for a hotel search. The main content area displays reviews for the Metropolitan Hotel in Berlin. Two reviews are visible:

- Review 1:** By user 'Silke' (5.7 rating). Text: "Für einen Wochenendaufenthalt völlig in Ordnung, super Ausgangslage, um die Stadt zu erkunden." (For a weekend stay, completely in order, great starting point to explore the city.)
- Review 2:** By user 'Margret' (7.0 rating). Text: "Einfach ein schöner Kurzurlaub." (Simply a nice short vacation.)

XPath Expression

```
1 doc("https://www.booking.com/hotel/de/metropolitan...")
2 //a[@id='show_reviews_tab']/{click /}
3 /(//*[id='review_next_page_link']/{clkwithchange
4 //div[contains(@class, 'review_list_block')]
5 //li[contains(@class, 'review_item')]
6 [not(contains(@class, 'featured_review_item'))]
7 [not(@class= 'review_item_photo
8 review_item_photo-p')]:<review>
9 [? .//*[contains(@class, 'review_item_date')]
10 :<date=normalize-space(.)>]
11 [? .//*[contains(@class, 'review_item_review_score')]
12 :<score=normalize-space(.)>]
13 [? .//*[@class='review_item_header_content_container']
14 :<title=normalize-space(.)>]
15 [? .//*[contains(@class, 'review_item_review_content')]
16 :<text=string-join(.p/text(), " ")>]
17 [? .//*[@class='reviewer_country']
18 :<country=normalize-space(.)>]
```

Demonstration: Twitter



XPath Expression

```
1 doc("https://twitter.com/search-home")
2 //input[@id='search-home-input']/{'urlaub'}/{'presenter/}
3 /{/div[contains(@class,'stream-footer')]
4   /{mouseover /})*{0, 4}
5 /. :<count=count(//li[@data-item-type='tweet'])>
6 //li[@data-item-type='tweet']:<tweet>
7   [? ../strong[@class='fullname show-popup-with-id ' ]
8     :<user_name=string(.)>
9     [? ../a[starts-with(@class,
10       'account-group')]/span[@class='username u-dir']
11       :<user_id=string(.)>
12       [? ../a[@class="tweet-timestamp js-permalink js-nav
13         js-tooltip"]/@title:<date=string(.)>]
14       [? ../p[starts-with(@class,'TweetTextSize')]
15         :<content=normalize-space(.)>]
16       [? ../button[contains(@aria-describedby,
17         'reply-count')]/span/span:<antworten=string(.)>]
18       [? ../button[contains(@aria-describedby,
19         'retweet-count')]/span/span:<retweets=string(.)>]
20       [? ../button[contains(@aria-describedby,
21         'favorite-count')]/span/span:<likes=string(.)>]
```

Discussion

Thank you for your attention!
Feel free to ask any questions now!

Contact us:

`mandy.neumann@th-koeln.de`

`michelsc@uni-trier.de`

Source:

Visit <http://www.oxpath.org>

Table of contents

- 1 Project Profile: Smart Harvesting II
- 2 Maintaining Scientific Literature Databases
- 3 XPath
- 4 Monitoring
- 5 Examples
- 6 Demonstration