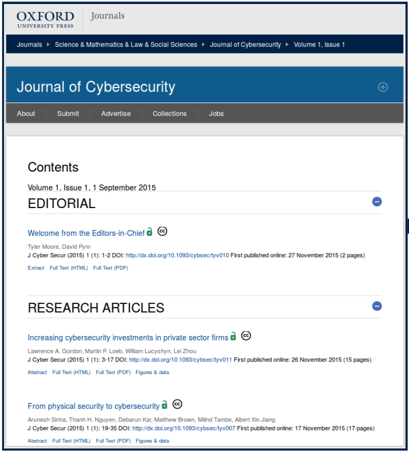
Smart Harvesting with OXPath

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

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

June 15, 2018

Harvesting Bibliographic Data



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

Journal of Cybersecurity

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

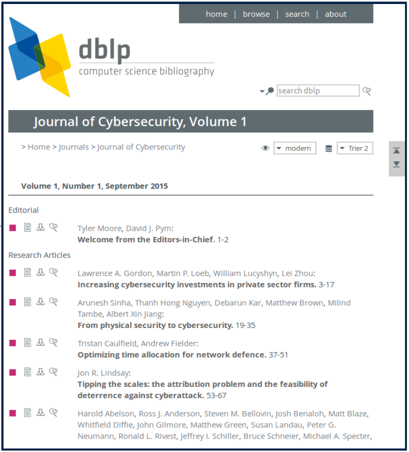
From physical security to cybersecurity

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

?



home | browse | search | about
dblp
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

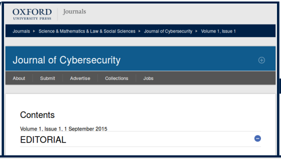
Tristan Caulfield, Andrew Fielder:
Optimizing time allocation for network defence. 37-51

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

?



OXFORD UNIVERSITY PRESS | Journals

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

Journal of Cybersecurity

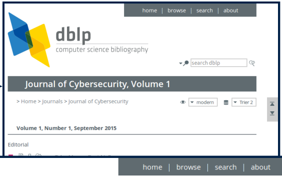
About | Submit | Advertise | Collections | Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

?



home | browse | search | about

dblp computer science bibliography

search dblp

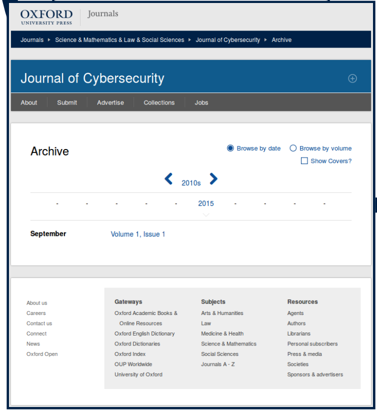
Journal of Cybersecurity, Volume 1

> Home > Journals > Journal of Cybersecurity

modern | Trier 2

Volume 1, Number 1, September 2015

Editorial



OXFORD UNIVERSITY PRESS | Journals

Journals ▸ Science & Mathematics & Law & Social Sciences ▸ Journal of Cybersecurity ▸ Archive

Journal of Cybersecurity

About | Submit | Advertise | Collections | Jobs

Archive

● Browse by date ○ Browse by volume

□ Show Covers?

‹ 2010s ›

- - - - - 2015 - - - -

September Volume 1, Issue 1

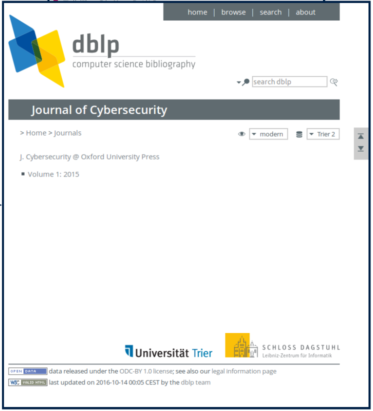
About us
Careers
Contact us
Connect
News
Oxford Open

Gateways
Oxford Academic Books & Online Resources
Oxford English Dictionary
Oxford Dictionaries
Oxford Index
OUP Worldwide
University of Oxford

Subjects
Arts & Humanities
Law
Medicine & Health
Science & Mathematics
Social Sciences
Journals A - Z

Resources
Agents
Authors
Librarians
Personal subscribers
Press & media
Societies
Sponsors & advertisers

?



home | browse | search | about

dblp computer science bibliography

search dblp

Journal of Cybersecurity

> Home > Journals

modern | Trier 2

J. Cybersecurity @ Oxford University Press

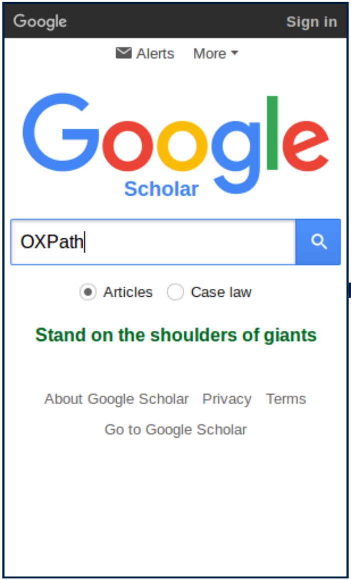
- Volume 1: 2015

Universität Trier | SCHLOSS DAGSTUHL Leibniz-Zentrum für Informatik

OPEN DATA data released under the ODC-BY 1.0 license; see also our legal information page

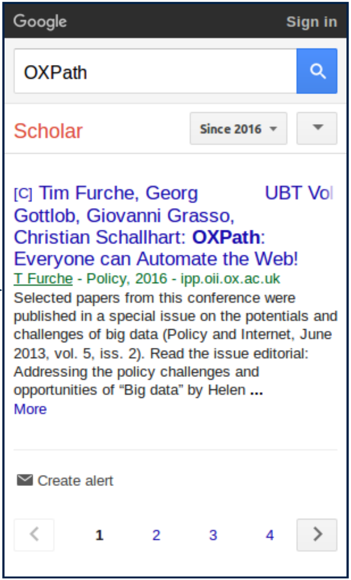
W3C VALID HTML last updated on 2016-10-14 00:05 CEST by the dblp team

Accessing Bibliographic Data



Stand on the shoulders of giants

?



[C] Tim Furche, Georg Gottlob, Giovanni Grasso, Christian Schallhart: **OXPath**: Everyone can Automate the Web! UBT Vol

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

The Role of OXPath in Smart Harvesting II

Motivation:

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

Solution:

- provide working environment and tools to use OXPath

OXPath:

- simple, declarative language for web data extraction

Table of Contents

1. The Role of OXPath in Smart Harvesting II
2. Maintaining Scientific Literature Databases
3. OXPath
4. Examples
5. Demonstration

Table of Contents

1 The Role of OXPath in Smart Harvesting II

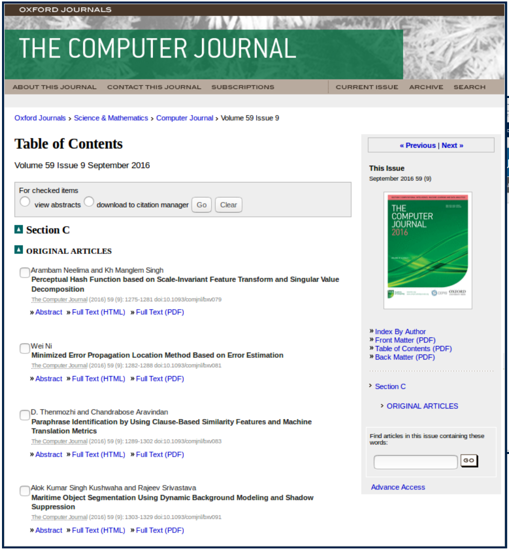
2 Maintaining Scientific Literature Databases

3 OXPath

4 Examples

5 Demonstration

Sources of Raw Web 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 [Go] [Clear]

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): 1275-1281 doi:10.1093/comjnl/bxv079
» 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/bxv081
» Abstract » Full Text (HTML) » Full Text (PDF)

□ D. Thenmozhi 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/bxv083
» 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/bxv091
» Abstract » Full Text (HTML) » Full Text (PDF)

« Previous | Next »

This Issue
September 2016 59 (9)

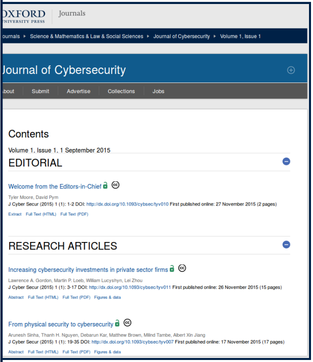
THE COMPUTER JOURNAL 2016

» 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:
[GO]

Advance Access



XFORD
NIVERSITY PRESS
Journals

ournals ▸ Science & Mathematics & Law & Social Sciences ▸ Journal of Cybersecurity ▸ Volume 1, Issue 1

Journal of Cybersecurity

bout Submit Advertise Collections Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

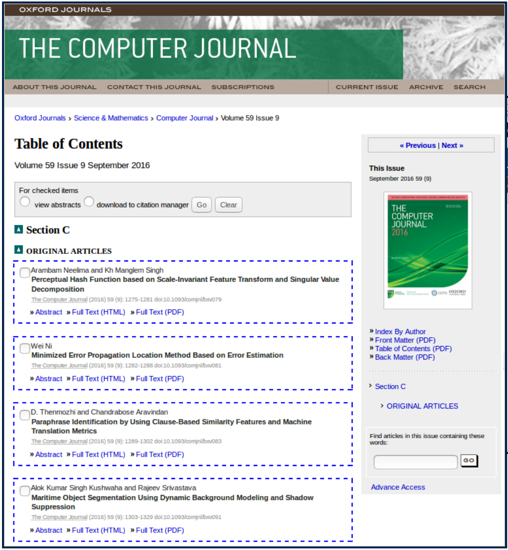
Welcome from the Editors-in-Chief
Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: http://dx.doi.org/10.1093/cybsec/tyv010 First published online: 27 November 2015 (2 pages)
Extract Full Text (HTML) Full Text (PDF)

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 Web 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 Go Clear

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): 1275-1281 doi:10.1093/comjnl/bxv079
» 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/bxv081
» Abstract » Full Text (HTML) » Full Text (PDF)

□ D. Thenmozhi 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/bxv083
» 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/bxv091
» Abstract » Full Text (HTML) » Full Text (PDF)

« Previous | Next »

This Issue
September 2016 59 (9)

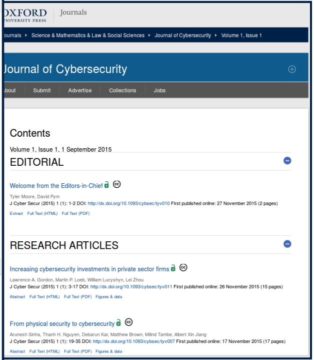
THE COMPUTER JOURNAL 2016

» 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:
GO

Advance Access



XFORD UNIVERSITY PRESS   Journals

ournals ▸ Science & Mathematics & Law & Social Sciences ▸ Journal of Cybersecurity ▸ Volume 1, Issue 1

Journal of Cybersecurity

bout   Submit   Advertise   Collections   Jobs

Contents

Volume 1, Issue 1, 1 September 2015

EDITORIAL

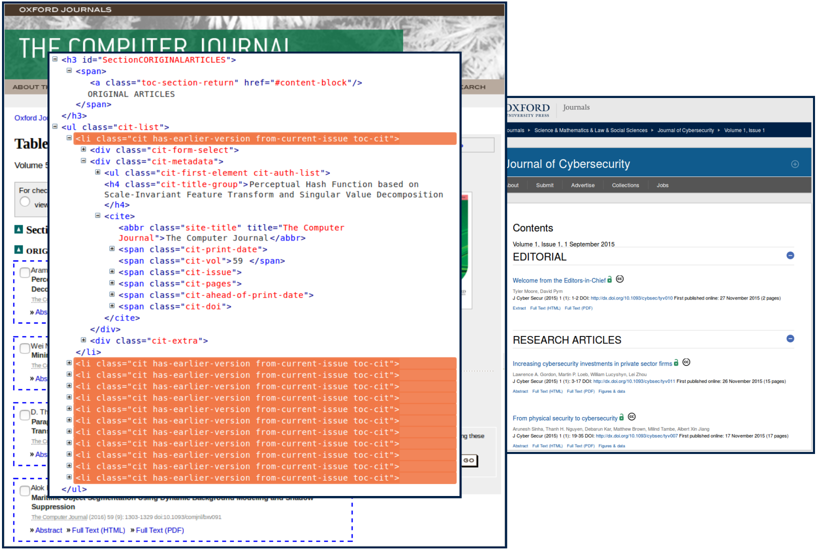
Welcome from the Editors-in-Chief
Tyler Moore, David Pym
J Cyber Secur (2015) 1 (1): 1-2 DOI: http://dx.doi.org/10.1093/cybsec/tyv010 First published online: 27 November 2015 (2 pages)
Extract   Full Text (HTML)   Full Text (PDF)

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 Web Data



OXFORD JOURNALS

THE COMPUTER JOURNAL

ABOUT T… …ARCH

Oxford Jo…

Table…

Volume 5…

For chec…
○ view…

Secti…

ORIG…

□ Aram…
Perc…
Deco…
The C…
» Abs…

□ Wei N…
Mini…
The C…
» Abs…

□ D. Th…
Para…
Trans…
The C…
» Abs…

□ Alok …
Maritime Object Segmentation Using Dynamic Background Modeling and Shadow Suppression
The Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/bxv091
» Abstract » Full Text (HTML) » Full Text (PDF)

…ng these

GO

```
⊟ <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">
        </cite>
      </div>
    ⊞ <div class="cit-extra">
    </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>
```

…XFORD UNIVERSITY PRESS
Journals
…ournals ▸ Science & Mathematics & Law & Social Sciences ▸ Journal of Cybersecurity ▸ Volume 1, Issue 1
Journal of Cybersecurity
…bout
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

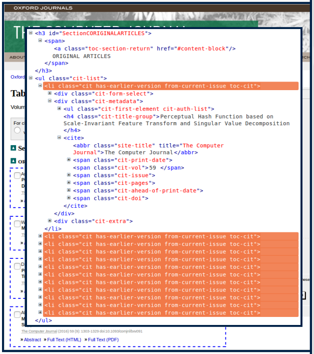
Increasing cybersecurity investments in private sector firms

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

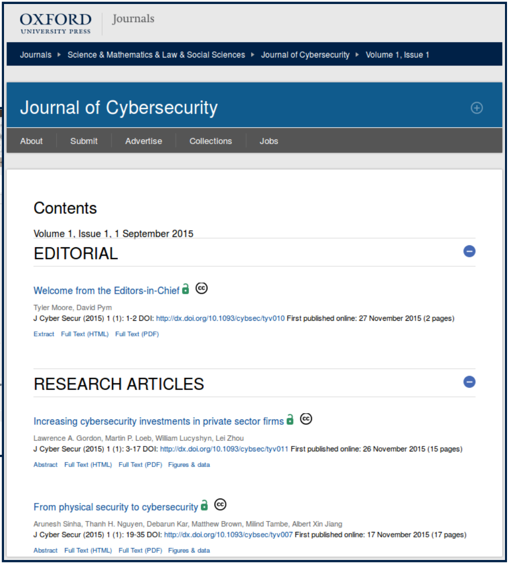
Sources of Raw Web Data



```
⊟ <h3 id="SectionCORIGINALARTICLES">
  ⊟ <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">
        </cite>
      </div>
    ⊞ <div class="cit-extra">
    </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 Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/bxv091

» Abstract » Full Text (HTML) » Full Text (PDF)



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

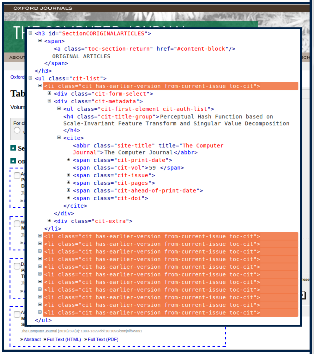
From physical security to cybersecurity

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 Web Data

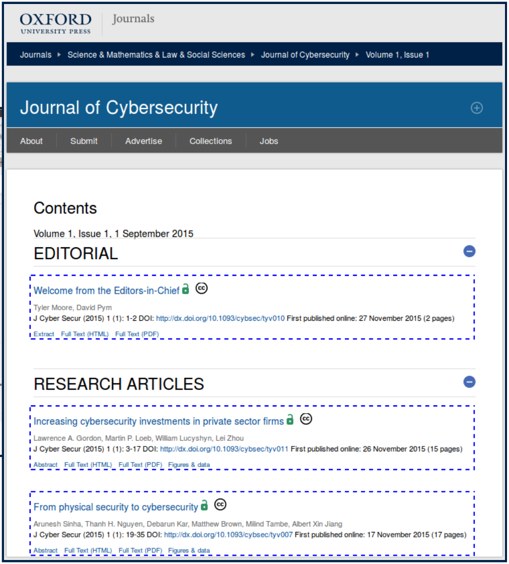


OXFORD JOURNALS

```
<h3 id="SectionCORIGINALARTICLES">
  <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">
      </cite>
    </div>
    <div class="cit-extra">
  </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 Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/bxv091

» Abstract » Full Text (HTML) » Full Text (PDF)



OXFORD UNIVERSITY PRESS | Journals

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

Journal of Cybersecurity

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

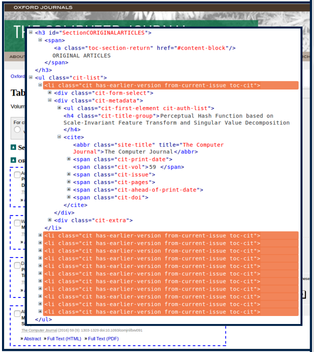
From physical security to cybersecurity

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

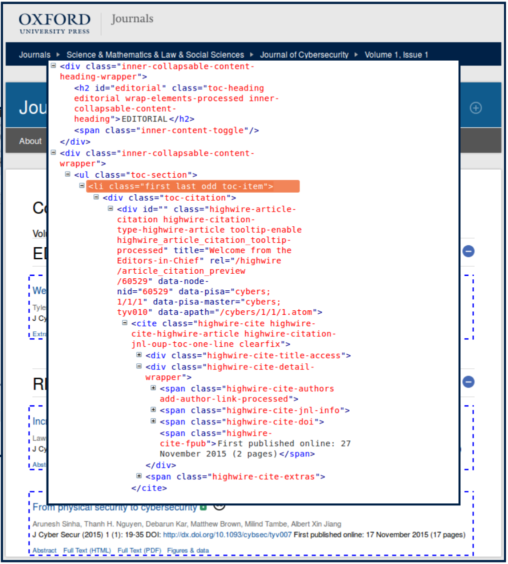
Sources of Raw Web Data



```
<h3 id="SectionCORIGINALARTICLES">
  <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">
      </cite>
    </div>
    <div class="cit-extra">
  </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 Computer Journal (2016) 59 (9): 1303-1329 doi:10.1093/comjnl/bxv091

» Abstract » Full Text (HTML) » Full Text (PDF)



OXFORD
UNIVERSITY PRESS
Journals
Journals ▸ Science & Mathematics & Law & Social Sciences ▸ Journal of Cybersecurity ▸ Volume 1, Issue 1
About

```
<div class="inner-collapsable-content-
heading-wrapper">
    <h2 id="editorial" class="toc-heading
    editorial wrap-elements-processed inner-
    collapsable-content-
    heading">EDITORIAL</h2>
    <span class="inner-content-toggle"/>
</div>
<div class="inner-collapsable-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-fpub">First published online: 27
              November 2015 (2 pages)</span>
            </div>
            <span class="highwire-cite-extras">
          </cite>
```

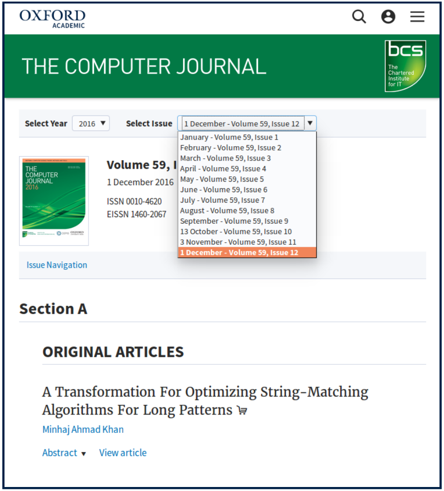
From physical security to cybersecurity

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 Web Data



OXFORD ACADEMIC

THE COMPUTER JOURNAL

bcs 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, I

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

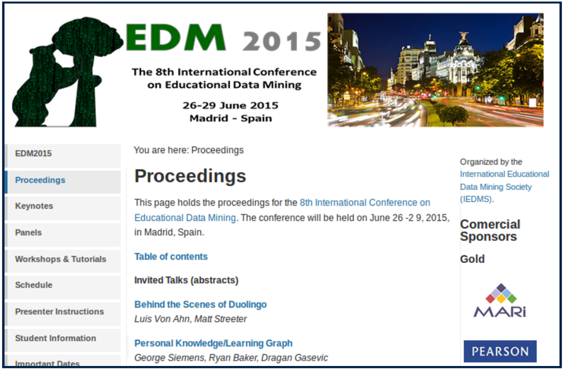
Minhaj Ahmad Khan

Abstract ▼ View article

Oxford Academic:

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

Sources of Raw Web Data



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.

Table of contents

Invited Talks (abstracts)

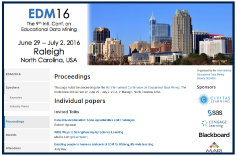
Behind the Scenes of Duolingo
Luis Von Ahn, Matt Streeter

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

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

Comercial Sponsors

Gold



The 9th Intl. Conf. on Educational Data Mining

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

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

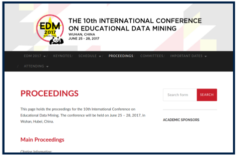
Data-Driven Education: Some opportunities and Challenges
Rakesh Agrawal

WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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

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

Sponsors


EDM 2017
EDM 2017 ▾ / KEYNOTES / SCHEDULE ▾ / PROCEEDINGS / COMMITTEES / IMPORTANT DATES ▾
/ ATTENDING ▾
Search form
SEARCH

THE 10th INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING
WUHAN, CHINA
JUNE 25 - 28, 2017

PROCEEDINGS

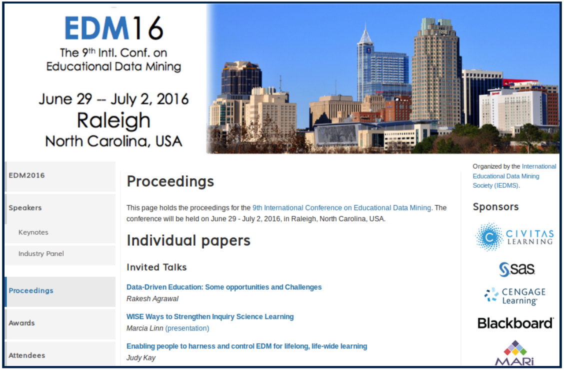
This page holds the proceedings for the 10th International Conference on Educational Data Mining. The conference will be held on June 25 – 28, 2017, in Wuhan, Hubei, China.

ACADEMIC SPONSORS

Main Proceedings

Citation Information:

Sources of Raw Web Data



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.

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

Sponsors

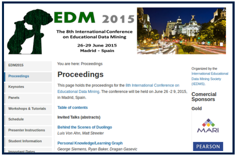
Individual papers

Invited Talks

Data-Driven Education: Some opportunities and Challenges
Rakesh Agrawal

WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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



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

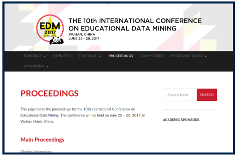
Gold

Table of contents

Invited Talks (abstracts)

Behind the Scenes of Duolingo
Luis Von Ahn, Matt Streeter

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



PROCEEDINGS

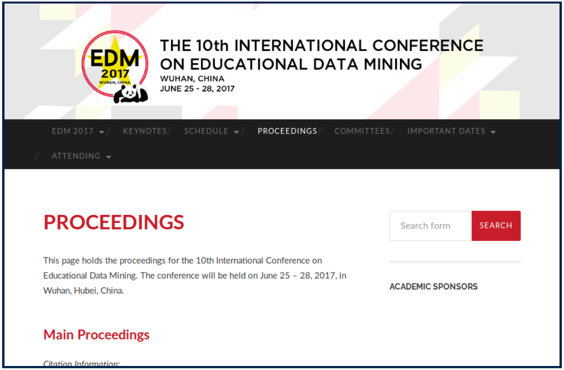
This page holds the proceedings for the 10th International Conference on Educational Data Mining. The conference will be held on June 25 – 28, 2017, in Wuhan, Hubei, China.

ACADEMIC SPONSORS

Main Proceedings

Citation Information:

Sources of Raw Web Data



THE 10th INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING
WUHAN, CHINA
JUNE 25 - 28, 2017

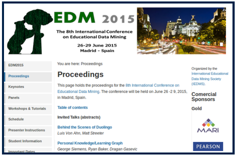
PROCEEDINGS

This page holds the proceedings for the 10th International Conference on Educational Data Mining. The conference will be held on June 25 – 28, 2017, in Wuhan, Hubei, China.

ACADEMIC SPONSORS

Main Proceedings

Citation Information:



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.

Table of contents

Invited Talks (abstracts)

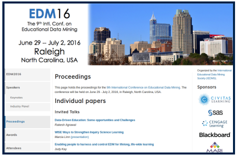
Behind the Scenes of Duolingo
Luis Von Ahn, Matt Streeter

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

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

Comercial Sponsors

Gold



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

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

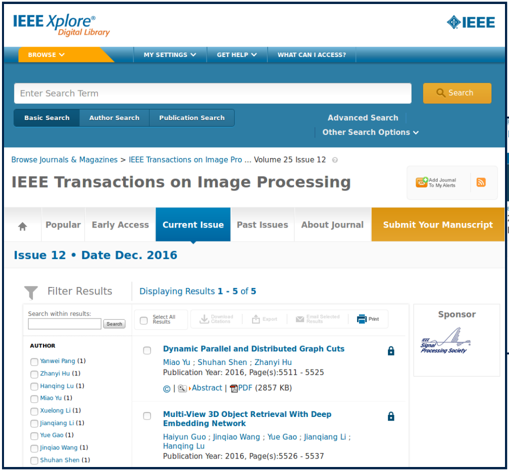
WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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

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

Sponsors

Sources of Raw Web Data



Browse Journals & Magazines > IEEE Transactions on Image Pro ... Volume 25 Issue 12

IEEE Transactions on Image Processing

Issue 12 • Date Dec. 2016

Displaying Results **1 - 5** of **5**

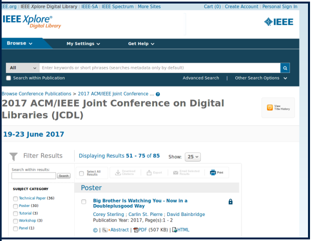
Search within results: Search

AUTHOR

- □ Yanwei Pang (1)
- □ Zhanyi Hu (1)
- □ Hanqing Lu (1)
- □ Miao Yu (1)
- □ Xuelong Li (1)
- □ Jianqiang Li (1)
- □ Yue Gao (1)
- □ Jinqiao Wang (1)
- □ Shuhan Shen (1)

□ **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


EE.org | IEEE Xplore Digital Library | IEEE-SA | IEEE Spectrum | More Sites
Cart (0) | Create Account | Personal Sign In
IEEE Xplore® Digital Library
IEEE
Browse
My Settings
Get Help
All
Enter keywords or short phrases (searches metadata only by default)
Search within Publication
Advanced Search
Other Search Options
View Title History
Filter Results
Show: 25
Select All Results
Download Citations
Export
Email Selected Results
Print

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

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

19-23 June 2017

Displaying Results **51 - 75** of **85**

Search within results: Search

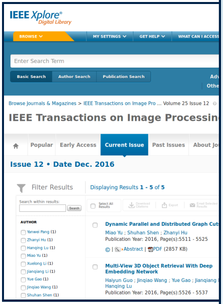
SUBJECT CATEGORY

- □ 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

Sources of Raw Web Data



Browse Journals & Magazines > IEEE Transactions on Image Pro ... Volume 25 Issue 12

IEEE Transactions on Image Processin

Issue 12 • Date Dec. 2016

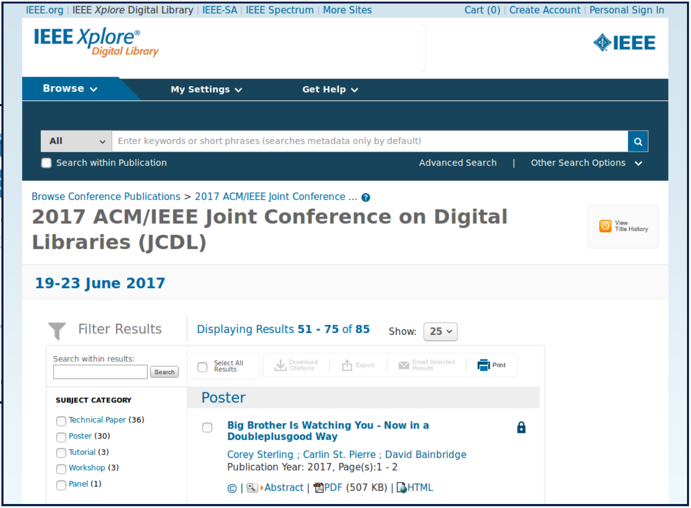
Displaying Results 1 - 5 of 5

AUTHOR

- □ Yanwei Pang (1)
- □ Zhanyi Hu (1)
- □ Hanqing Lu (1)
- □ Miao Yu (1)
- □ Xuelong Li (1)
- □ Jianqiang Li (1)
- □ Yue Gao (1)
- □ Jinqiao Wang (1)
- □ Shuhan Shen (1)

□ **Dynamic Parallel and Distributed Graph Cut**
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
Hanqing Lu
Publication Year: 2016, Page(s):5526 - 5537


IEEE.org | IEEE Xplore Digital Library | IEEE-SA | IEEE Spectrum | More Sites
Cart (0) | Create Account | Personal Sign In
IEEE Xplore® Digital Library
IEEE
Browse
My Settings
Get Help
All
Enter keywords or short phrases (searches metadata only by default)
Search within Publication
Advanced Search
Other Search Options
View Title History
Filter Results
Show: 25
Search within results:
Search
Select All Results
Download Citations
Export
Email Selected Results
Print

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

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

19-23 June 2017

Displaying Results 51 - 75 of 85

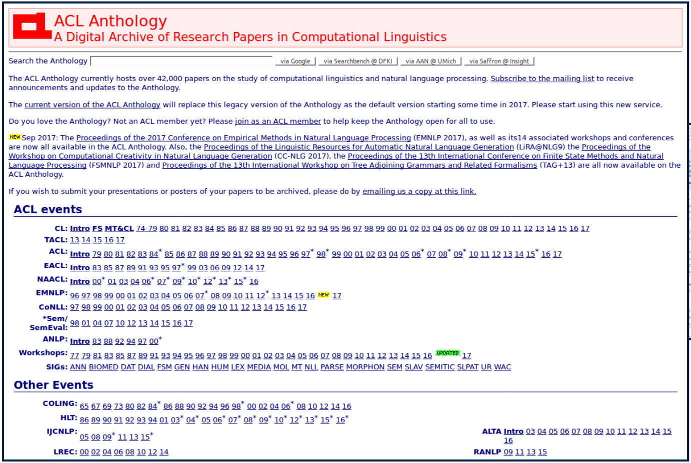
SUBJECT CATEGORY

- □ 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

Sources of Raw Web Data



ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

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 its14 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 (FSMNLP 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.

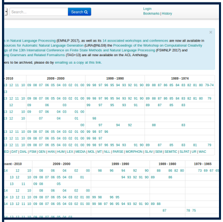
If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this link.

ACL events

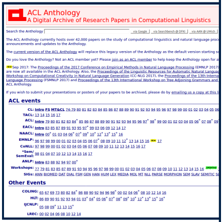
CL: **Intro** **FS** **MT&CL** 74-79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17
TACL: 13 14 15 16 17
ACL: **Intro** 79 80 81 82 83 84* 85 86 87 88 89 90 91 92 93 94 95 96 97* 98* 99 00 01 02 03 04 05 06* 07 08* 09* 10 11 12 13 14 15* 16 17
EACL: **Intro** 83 85 87 89 91 93 95 97* 99 03 06 09 12 14 17
NAACL: **Intro** 00* 01 03 04 06* 07* 09* 10* 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 UPDATED 17
SIGs: ANN BIOMED DAT DIAL FSM GEN HAN HUM LEX MEDIA MOL MT NLL PARSE MORPHON SEM SLAV SEMITIC SLPAT 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* 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



Sources of Raw Web Data



ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

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

The ACL Anthology currently hosts over 42,000 papers on the study of computational linguistics and natural language proce
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 so

Do you love the Anthology? Not an ACL member yet? Please join as an ACL member to help keep the Anthology open for a

NEW Sep 2017: The Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)
are now all available in the ACL Anthology. Also, the Proceedings of the Linguistic Resources for Automatic Natural Languag
Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017), the Proceedings of the 13th Interna
Language Processing (FSMNLP 2017) and Proceedings of the 13th International Workshop on Tree Adjoining Grammars and
ACL Anthology.

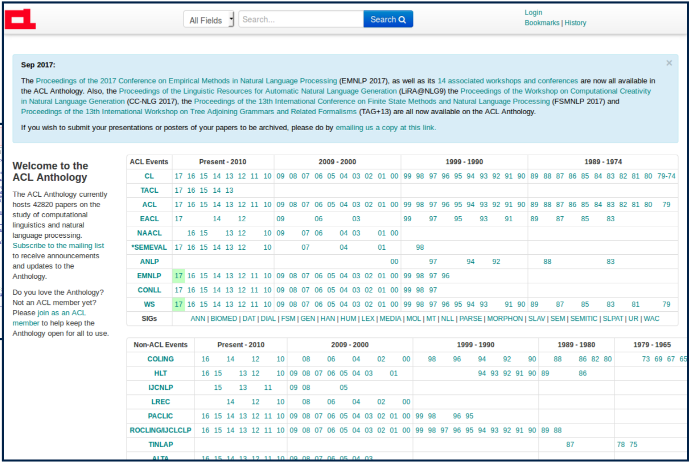
If you wish to submit your presentations or posters of your papers to be archived, please do by emailing us a copy at this li

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
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*
EACL: Intro 83 85 87 89 91 93 95 97* 99 03 06 09 12 14 17
NAACL: Intro 00* 01 03 04 06* 07* 09* 10* 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 UPDATED
SIGs: ANN BIOMED DAT DIAL FSM GEN HAN HUM LEX MEDIA MOL MT NLL PARSE MORPHON SEM SLAV SEMITIC SI

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* 12* 13* 15* 16*
IJCNLP: 05 08 09* 11 13 15*
LREC: 00 02 04 06 08 10 12 14



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 (FSMNLP 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.

Welcome to the ACL Anthology

The ACL Anthology currently hosts 42820 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to receive announcements and updates to the Anthology.

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.

ACL Events	Present - 2010	2009 - 2000	1999 - 1990	1989 - 1974
CL	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79-74
TACL	17 16 15 14 13			
ACL	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79
EACL	17 14 12	09 06 03	99 97 95 93 91	89 87 85 83
NAACL	16 15 13 12 10	09 07 06 04 03 01 00		
*SEMEVAL	17 16 15 14 13 12 10	07 04 01	98	
ANLP		00	97 94 92	88 83
EMNLP	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96	
CONLL	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97	
WS	17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 91 90	89 87 85 83 81 79
SIGs	ANN | BIOMED | DAT | DIAL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MT | NLL | PARSE | MORPHON | SLAV | SEM | SEMITIC | SLPAT | UR | WAC			

Non-ACL Events	Present - 2010	2009 - 2000	1999 - 1990	1989 - 1980	1979 - 1965
COLING	16 14 12 10	08 06 04 02 00	98 96 94 92 90	88 86 82 80	73 69 67 65
HLT	16 15 13 12 10	09 08 07 06 05 04 03 01	94 93 92 91 90	89 86	
IJCNLP	15 13 11	09 08 05			
LREC	14 12 10	08 06 04 02 00			
PACLIC	16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 96 95		
ROCLING/IJCLCLP	16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88	
TINLAP				87	78 75
ALTA	16 15 14 13 12 11 10	09 08 07 06 05 04 03			

Problem of Data Heterogeneity

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

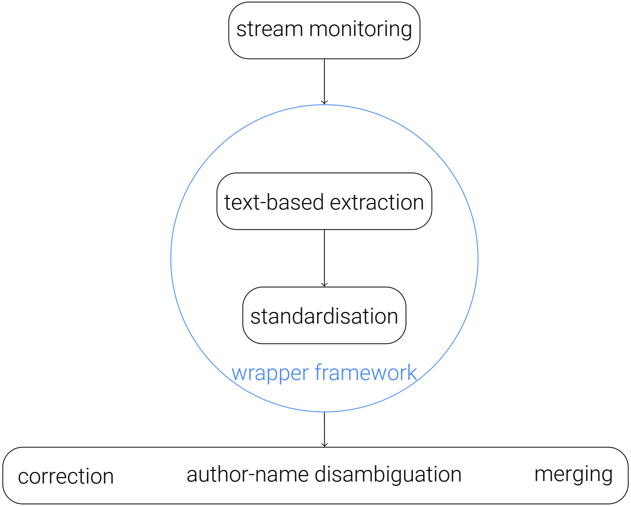
- website layouts
- change

Problem of Data Integration

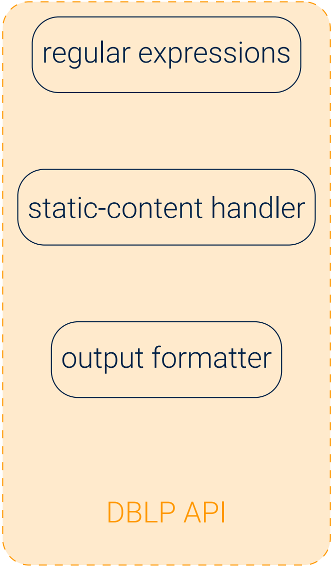
- question of feasibility: automated vs. manual harvesting

- expensive maintenance

DBLP as a Case Example



DBLP as a Case Example



regular expressions

static-content handler

output formatter

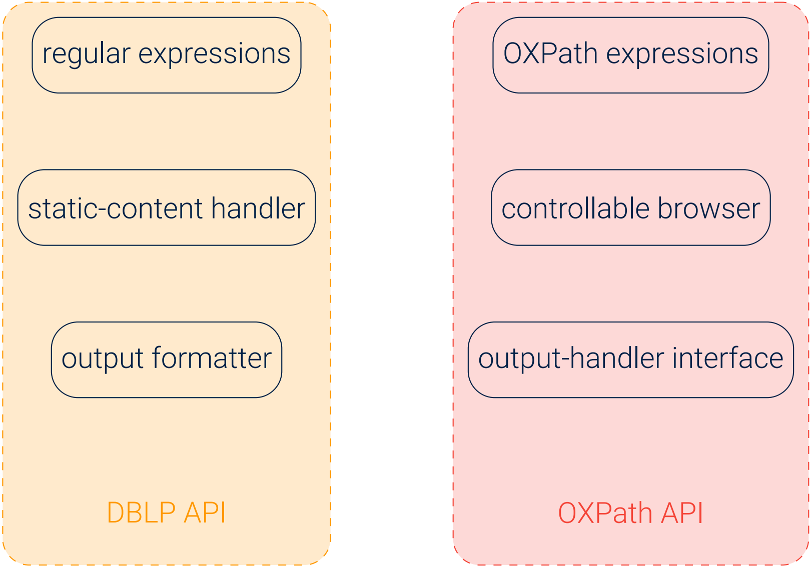
DBLP API

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

OXPath expressions

controllable browser

output-handler interface

OXPath API

Table of Contents

1. The Role of OXPath in Smart Harvesting II
2. Maintaining Scientific Literature Databases
3. OXPath
4. Examples
5. Demonstration

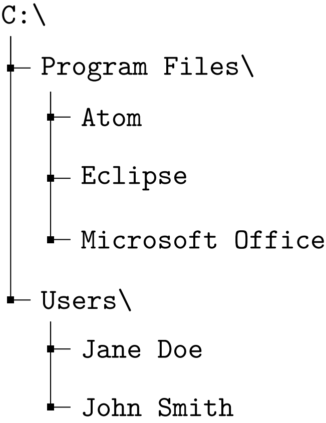
What Is OXPath?

- 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?



C:\

Program Files\

Atom

Eclipse

Microsoft Office

Users\

Jane Doe

John Smith

File-Path Examples

```
C:\Program Files\Microsoft Office
C:\Users\Jane Doe
```

What Is XPath?

Queried XML File

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

What Is XPath?

Queried XML File

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

```
(
  <issue>11</issue>,
  <issue>10</issue>,
  <issue>9</issue>
)
```

What Is XPath?

Queried XML File

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

What Is XPath?

Queried XML File

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

```
(
  "http://.../toc/tadr20/30/11",
  "http://.../toc/tadr20/30/10",
  "http://.../toc/tadr20/30/9"
)
```

What Is XPath?

Queried XML File

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

What Is XPath?

Queried XML File

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/11</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>10</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/10</url>
  </record>
  <record>
    <volume>30</volume>
    <issue>9</issue>
    <year>2016</year>
    <url>http://.../tadr20/30/9</url>
  </record>
</results>
```

XPath Expression

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

Result Set

```
(
  <record class="current">
    <volume>30</volume>
    <issue>11</issue>
    <year>2016</year>
    <url>[...]</url>
  </record>
)
```

What Does OXPath Add?

Action:

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

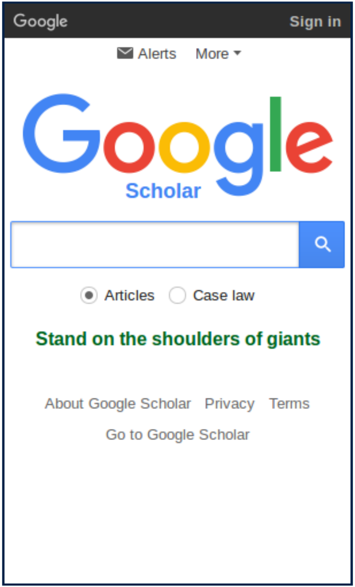
Extraction:

- add markers to extract selected nodes

Iteration:

- loops, e.g. for paginated content

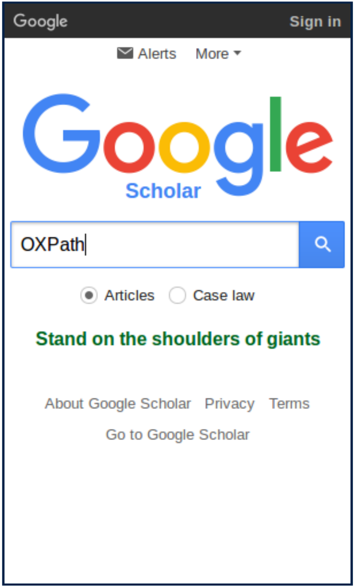
Example: Navigating Google Scholar



OXPath Expression

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

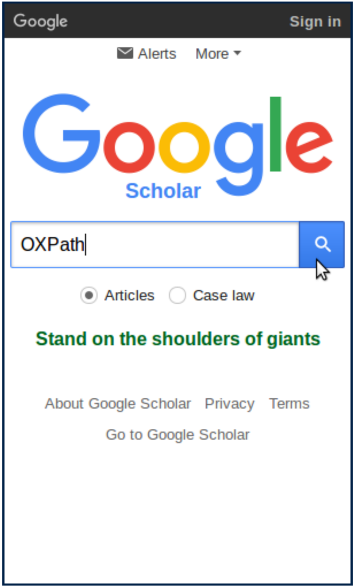
Example: Navigating Google Scholar



OXPath Expression

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

Example: Navigating Google Scholar



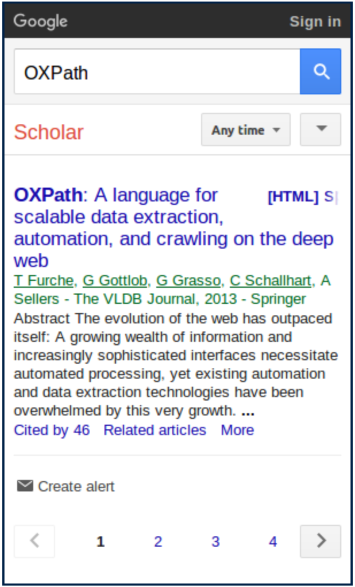
OXPath

Stand on the shoulders of giants

OXPath Expression

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

Example: Navigating Google Scholar



OXPath

OXPath: A language for scalable data extraction, automation, and crawling on the deep web [HTML] S

T Furche, G Gottlob, G Grasso, C Schallhart, A Sellers - The VLDB Journal, 2013 - Springer

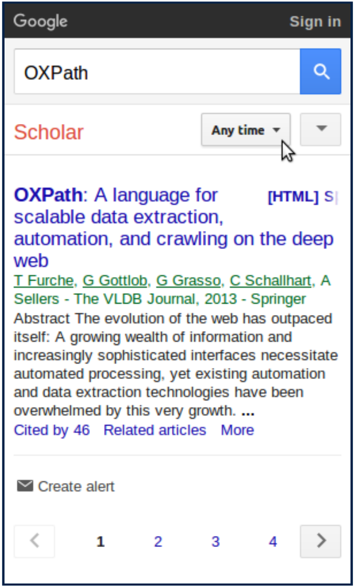
Abstract The evolution of the web has outpaced itself: A growing wealth of information and increasingly sophisticated interfaces necessitate automated processing, yet existing automation and data extraction technologies have been overwhelmed by this very growth. **...**

Cited by 46 Related articles More

OXPath Expression

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

Example: Navigating Google Scholar



Google Sign in

OXPath

Scholar Any time

OXPath: A language for scalable data extraction, automation, and crawling on the deep web [HTML] S

T Furche, G Gottlob, G Grasso, C Schallhart, A Sellers - The VLDB Journal, 2013 - Springer

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. ...

Cited by 46 Related articles More

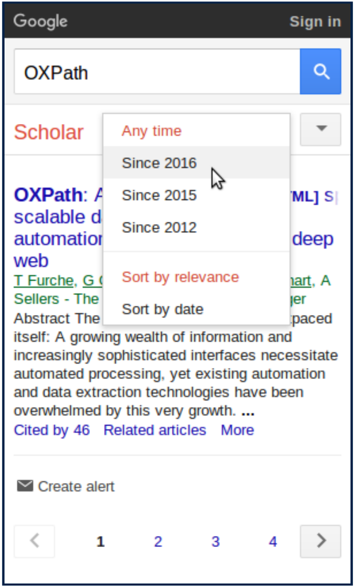
Create alert

< 1 2 3 4 >

OXPath Expression

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

Example: Navigating Google Scholar



OXPath: A ... ML] S... scalable d... automation ... deep web

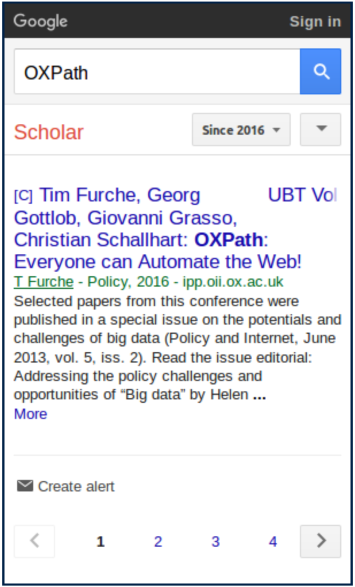
T Furche, G ... hart, A Sellers - The ... ger

Abstract The ... paced 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. ...

OXPath Expression

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

Example: Navigating Google Scholar



[C] Tim Furche, Georg Gottlob, Giovanni Grasso, Christian Schallhart: **OXPath**: Everyone can Automate the Web! UBT Vol

T Furche - Policy, 2016 - ipp.oii.ox.ac.uk

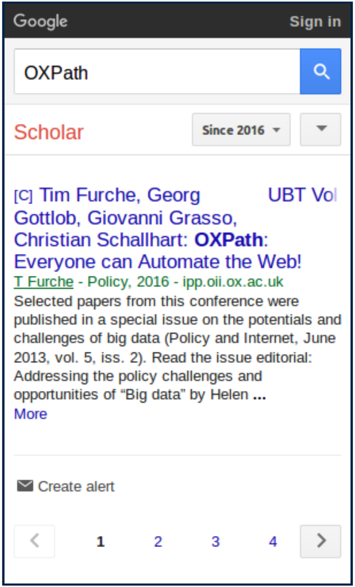
Selected papers from this conference were published in a special issue on the potentials and challenges of big data (Policy and Internet, June 2013, vol. 5, iss. 2). Read the issue editorial: Addressing the policy challenges and opportunities of "Big data" by Helen **...**

More

OXPath Expression

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

Example: Navigating Google Scholar



[C] Tim Furche, Georg Gottlob, Giovanni Grasso, Christian Schallhart: **OXPath**: Everyone can Automate the Web!

UBT Vol

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

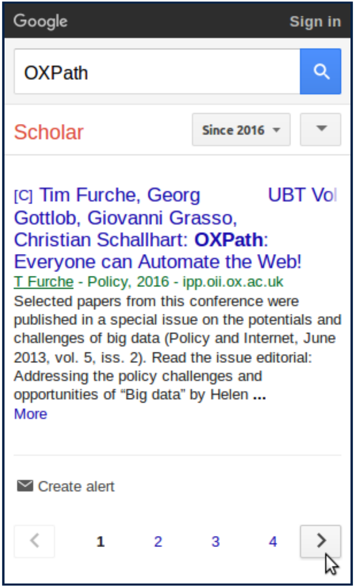
OXPath Expression

```
doc('https://scholar.google.com')
  //input[@id='gs_hdr_tsi']/{"OXPath"}
  /../following-sibling::button/{click/}
    //*[@id='gs_res_ab_yy-b']/{click}
      //following::*[@role='menuitemradio'][contains(.,
         '2016')]/{click/}
    //div[@class='gs_ri']//h3/a:<title=string(.)>
```

XML Output

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

Example: Navigating Google Scholar



[C] Tim Furche, Georg Gottlob, Giovanni Grasso, Christian Schallhart: **OXPath**: Everyone can Automate the Web!

UBT Vol

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

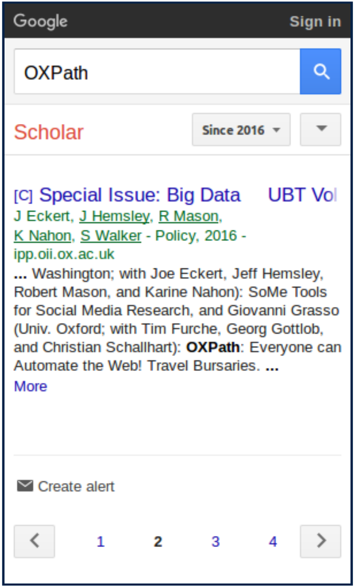
OXPath Expression

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

XML Output

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

Example: Navigating Google Scholar



Google Sign in

OXPath

Scholar Since 2016

[C] Special Issue: Big Data UBT Vol

J Eckert, J Hemsley, R Mason, K Nahon, S Walker - Policy, 2016 - ipp.oii.ox.ac.uk

... 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. ...

More

Create alert

< 1 2 3 4 >

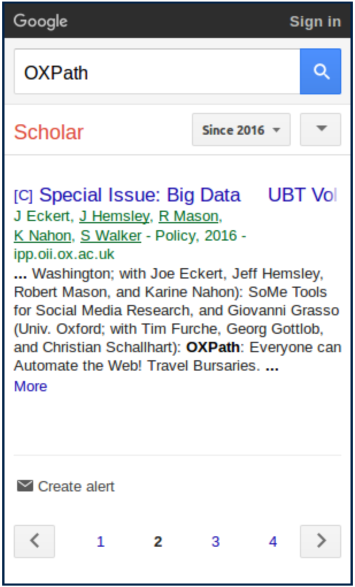
OXPath Expression

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

XML Output

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

Example: Navigating Google Scholar



[C] Special Issue: Big Data UBT Vol

J Eckert, J Hemsley, R Mason, K Nahon, S Walker - Policy, 2016 - ipp.oii.ox.ac.uk

... 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. ...

More

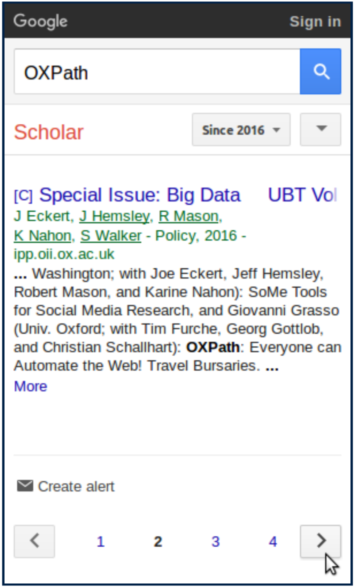
OXPath Expression

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

XML Output

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

Example: Navigating Google Scholar



[C] Special Issue: Big Data UBT Vol
J Eckert, J Hemsley, R Mason,
K Nahon, S Walker - Policy, 2016 -
ipp.oii.ox.ac.uk

... 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. ...
More

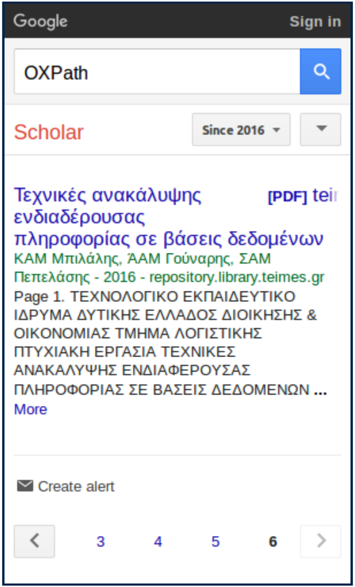
OXPath Expression

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

XML Output

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

Example: Navigating Google Scholar



Google Sign in

OXPath

Scholar Since 2016

Τεχνικές ανακάλυψης ενδιαφέρουσας πληροφορίας σε βάσεις δεδομένων [PDF] tei

ΚΑΜ Μπιλάλης, ΆΑΜ Γούναρης, ΣΑΜ Πεπελάσης - 2016 - repository.library.teimes.gr

Page 1. ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΤΕΧΝΙΚΕΣ ΑΝΑΚΑΛΥΨΗΣ ΕΝΔΙΑΦΕΡΟΥΣΑΣ ΠΛΗΡΟΦΟΡΙΑΣ ΣΕ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ...

More

Create alert

< 3 4 5 6 >

OXPath Expression

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

XML Output

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

Why OXPath?

XPath

- static web
- plain HTML
- complete content

OXPath

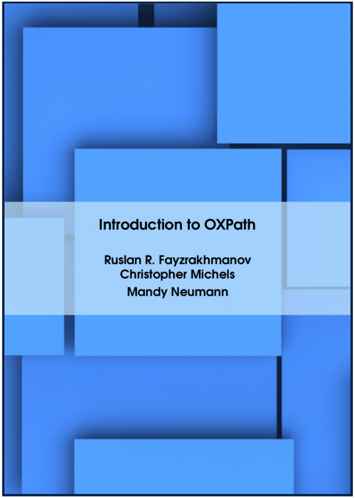
- dynamic web
- AJAX
- content on demand

OXPath In Use

In current working environment and production:

- OXPath-based wrappers in dblp
- collection of OXPath expressions prototypical of bibliographic domain
- OXPath Tutorial

OXPath Tutorial



Introduction to OXPath

Ruslan R. Fayzrakhmanov
Christopher Michels
Mandy Neumann

- supported by the Oxford research team that developed OXPath
- including:
  - a concise summary of XPath
  - setup and use instructions for OXPath
  - list of action keywords
  - list of functions for extraction and DOM-tree navigation
  - starter examples from the bibliographic domain

OXPath In Use

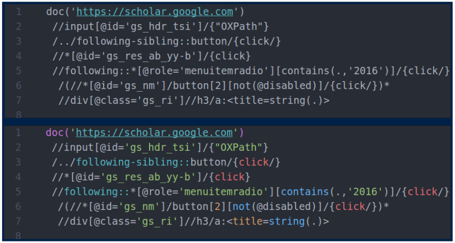
In current working environment and production:

- OXPath-based wrappers in dblp
- collection of OXPath expressions prototypical of bibliographic domain
- OXPath Tutorial
- integrated support for OXPath language in open-source editor *Atom*

Tool Support: Atom

Language plugin for Atom text editor

- syntax highlighting for keywords
- helps spotting errors and improves readability
- intended to lower barriers for beginners



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

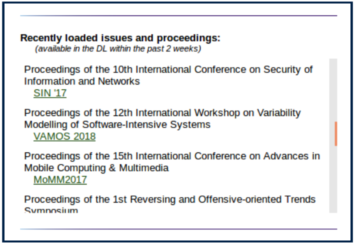
```
doc('https://scholar.google.com')
 //input[@id='gs_hdr_tsi']/{"OXPath"}
 /../following-sibling::button/{click/}
 //*[@id='gs_res_ab_yy-b']/{click}
 //following::*[@role='menuitemradio'][contains(.,'2016')]/{click/}
  /(//*[@id='gs_nm']/button[2][not(@disabled)]/{click/})*
  //div[@class='gs_ri']//h3/a: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

OXPath In Use

In current working environment and production:

- OXPath-based wrappers in dblp
- collection of OXPath expressions prototypical of bibliographic domain
- OXPath Tutorial
- integrated support for OXPath language in open-source editor *Atom*
- OXPath-based monitoring for digital libraries of large-scale publishers

Monitoring: ACM



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

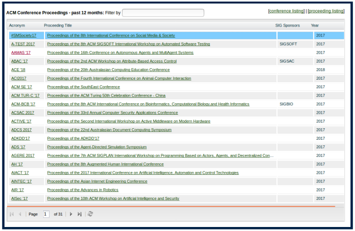
Proceedings of the 12th International Workshop on Variability Modelling of Software-Intensive Systems
VAMOS 2018

Proceedings of the 15th International Conference on Advances in Mobile Computing & Multimedia
MoMM2017

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

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

Monitoring: ACM



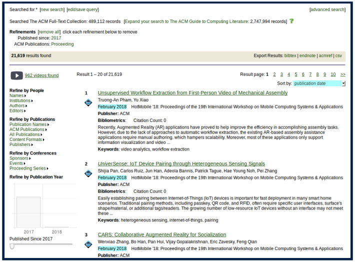
ACM Conference Proceedings - past 12 months: Filter by

[conference listing] | [proceeding listing]

Acronym	Proceeding Title	SIG Sponsors	Year
#SMSociety17	Proceedings of the 8th International Conference on Social Media & Society		2017
A-TEST 2017	Proceedings of the 8th ACM SIGSOFT International Workshop on Automated Software Testing	SIGSOFT	2017
AAMAS '17	Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems		2017
ABAC '17	Proceedings of the 2nd ACM Workshop on Attribute-Based Access Control	SIGSAC	2017
ACE '18	Proceedings of the 20th Australasian Computing Education Conference		2018
ACI2017	Proceedings of the Fourth International Conference on Animal-Computer Interaction		2017
ACM SE '17	Proceedings of the SouthEast Conference		2017
ACM TUR-C '17	Proceedings of the ACM Turing 50th Celebration Conference - China		2017
ACM-BCB '17	Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics	SIGBIO	2017
ACSAC 2017	Proceedings of the 33rd Annual Computer Security Applications Conference		2017
ACTIVE '17	Proceedings of the Second International Workshop on Active Middleware on Modern Hardware		2017
ADCS 2017	Proceedings of the 22nd Australasian Document Computing Symposium		2017
ADKDD'17	Proceedings of the ADKDD'17		2017
ADS '17	Proceedings of the Agent-Directed Simulation Symposium		2017
AGERE 2017	Proceedings of the 7th ACM SIGPLAN International Workshop on Programming Based on Actors, Agents, and Decentralized Con…		2017
AH '17	Proceedings of the 8th Augmented Human International Conference		2017
AIACT '17	Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies		2017
AINTEC '17	Proceedings of the Asian Internet Engineering Conference		2017
AIR '17	Proceedings of the Advances in Robotics		2017
AISec '17	Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security		2017

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

Monitoring: ACM



Searched for * [new search] [edit/save query] [advanced search]

Searched The ACM Full-Text Collection: 489,112 records [Expand your search to The ACM Guide to Computing Literature: 2,747,994 records]

Refinements [remove all] click each refinement below to remove
Published since: 2017
ACM Publications: Proceeding

21,619 results found — Export Results: bibtex | endnote | acmref | csv

962 videos found — Result 1 – 20 of 21,619 — Result page: **1** 2 3 4 5 6 7 8 9 10 >>

Sort by: publication date

Refine by People
Names
Institutions
Authors
Editors

Refine by Publications
Publication Names
ACM Publications
All Publications
Content Formats
Publishers

Refine by Conferences
Sponsors
Events
Proceeding Series

Refine by Publication Year

Published Since 2017

1 Unsupervised Workflow Extraction from First-Person Video of Mechanical Assembly
Truong-An Pham, Yu Xiao
February 2018 HotMobile '18: Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications
Publisher: ACM
Bibliometrics: Citation Count: 0
Recently, Augmented Reality (AR) applications have proved to help improve the efficiency in accomplishing assembly tasks. However, due to the lack of approaches to automatic 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 video ...
Keywords: video analytics, workflow extraction

2 *UniverSense*: IoT Device Pairing through Heterogeneous Sensing Signals
Shijia Pan, Carlos Ruiz, Jun Han, Adeola Bannis, Patrick Tague, Hae Young Noh, Pei Zhang
February 2018 HotMobile '18: Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications
Publisher: ACM
Bibliometrics: Citation Count: 0
Easily establishing pairing between Internet-of-Things (IoT) devices is important for fast deployment in many smart home scenarios. Traditional pairing methods, including passkey, QR code, and RFID, often require specific user interfaces, surface's shape/material, or additional tags/readers. The growing number of low-resource IoT devices without an interface may not meet these ...
Keywords: heterogeneous sensing, internet-of-things, pairing

3 CARS: Collaborative Augmented Reality for Socialization
Wenxiao Zhang, Bo Han, Pan Hui, Vijay Gopalakrishnan, Eric Zavesky, Feng Qian
February 2018 HotMobile '18: Proceedings of the 19th International Workshop on Mobile Computing Systems & Applications
Publisher: 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

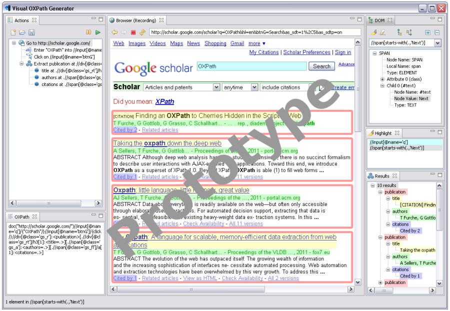
OXPath In Use

In future use:

- interactive tool for devising OXPath expressions
- integrate further tools for OXPath

Visual OXPath

- provide a visual interface
- integrate tools for XPath
- goal: semi-supervised wrapper induction



Go to http://scholar.google.com/
Enter "OXPath" into //input[@name
Click on //input[@name='btnG']
Extract publication at //div[@class=
title at .//div[@class='gs_rt']/h
authors at .//span[@class='gs_
citations at .//span[@class='gs

```
doc("http://scholar.google.com/")//input[@name='q']/{"OXPath"/}//input[@name='btnG']/{click/}//div[@class='gs_r']:<publication>[.//div[@class='gs_rt']/h3[1]:<title=.>][.//span[@class='gs_a']:<authors=.>][.//span[@class='gs_fl']/a[1]:<citations=.>]
```

1 element in //span[starts-with(.,'Next')]

http://scholar.google.com/scholar?q=OXPath&hl=en&btnG=Search&as_sdt=1%2C5&as_sdtp=on

Did you mean: *XPath*

[CITATION] Finding an **OXPath** to Cherries Hidden in the Script Web
T Furche, G Gottlob, G Grasso, C Schallhart... - ... . rep., diadem project ... path
Cited by 2 - Related articles

Taking the **oxpath** down the deep web
A Sellers, T Furche, G Gottlob... - Proceedings of the ..., 2011 - portal.acm.org
ABSTRACT Although deep web analysis has been studied extensively, there is no succinct formalism to describe user interactions with AJAX-enabled web applications. Toward this end, we introduce **OXPath** as a superset of XPath 1.0. Beyond XPath, **OXPath** is able (1) to fill web forms ...
Cited by 1 - Related articles - Check Availability - All 11 versions

Oxpath: little language, little memory, great value
AJ Sellers, T Furche, G Gottlob... - Proceedings of the ..., 2011 - portal.acm.org
ABSTRACT Data about everything is readily available on the web—but often only accessible through elaborate user interactions. For automated decision support, extracting that data is es- sential, but infeasible for existing heavy-weight data ex- traction systems. In this ...
Cited by ... - Related articles - Check Availability - All 11 versions

Oxpath: A language for scalable, memory-efficient data extraction from web applications
T Furche, G Gottlob, G Grasso, C Schallhart... - Proceedings of the VLDB ..., 2011 - fox7.eu
ABSTRACT The evolution of the web has outpaced itself. The growing wealth of information and the increasing sophistication of interfaces ne- cessitate automated processing. Web automation and extraction technologies have been overwhelmed by this very growth. To address this ...
Cited by 1 - Related articles - View as HTML - Check Availability - All 2 versions

//span[starts-with(.,'Next')]

```
SPAN
  Node Name: SPAN
  Local Name: span
  Type: ELEMENT
  Attribute 0 (class)
  Child 0 (#text)
    Node Name: #text
    Node Value: Next
    Type: TEXT
```

//input[@name='q']
//span[starts-with(.,'Next')]

```
10 results
  publication
    title
      [CITATION] Findin
    authors
      T Furche, G Gottlo
    citations
      Cited by 2
  publication
    title
      Taking the oxpath
    authors
      A Sellers, T Furche
    citations
      Cited by 1
  publication
```

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`, e.g.
to find the tutorial

Table of Contents

1 The Role of OXPath in Smart Harvesting II

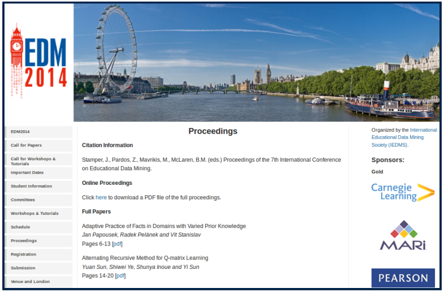
2 Maintaining Scientific Literature Databases

3 OXPath

4 Examples

5 Demonstration

EDM 2014: Simple Extraction



Proceedings

Citation Information

Stamper, J., Pardos, 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 Papousek, Radek Pelánek and Vit Stanislav
Pages 6-13 [pdf]

Alternating Recursive Method for Q-matrix Learning
Yuan Sun, Shiwei Ye, Shunya Inoue and Yi Sun
Pages 14-20 [pdf]

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

Sponsors:

Gold

OXPath Expression

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

XML Output

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

EDM 2014: Simple Extraction

HTML Source

```html
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
    <br/>
    <em> Jan Papousek, [...]</em>
    <br/>
    Pages 6-13 [
    <a href="uploads/[...].pdf">pdf</a>
    ]
    <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

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

EDM 2014: Simple Extraction

HTML Source

```html
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
   <!--[...]-->
   <strong>Online Proceedings</strong>
   <!--[...]-->
   <strong>Full Papers</strong>
   <!--[...]-->
   <p>Adaptive Practice of [...]
     <br/>
     <em> Jan Papousek, [...]</em>
     <br/>
     Pages 6-13 [
     <a href="uploads/[...].pdf">pdf</a>
     ]
     <!--[...]-->
   </p>
   <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <record></record>
  <record></record>
  <!--[...]-->
</results>
```

EDM 2014: Simple Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
  <!--[...]-->
  <div id="content">
    <!--[...]-->
    <strong>Online Proceedings</strong>
    <!--[...]-->
    <strong>Full Papers</strong>
    <!--[...]-->
    <p>Adaptive Practice of [...]
      <br/>
      <em> Jan Papousek, [...]</em>
      <br/>
      Pages 6-13 [
      <a href="uploads/[...].pdf">pdf</a>
      ]
      <!--[...]-->
    </p>
    <!--[...]-->
  </div>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors> Jan Papousek, [...]</authors>
  </record>
  <!--[...]-->
</results>
```

EDM 2014: Simple Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
  <!--[...]-->
  <div id="content">
    <!--[...]-->
    <strong>Online Proceedings</strong>
    <!--[...]-->
    <strong>Full Papers</strong>
    <!--[...]-->
    <p>Adaptive Practice of [...]
      <br/>
      <em> Jan Papousek, [...]</em>
      <br/>
      Pages 6-13 [
      <a href="uploads/[...].pdf">pdf</a>
      ]
      <!--[...]-->
    </p>
    <!--[...]-->
  </div>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors> Jan Papousek, [...]</authors>
    <title>Adaptive Practice of [...]</title>
  </record>
  <!--[...]-->
</results>
```

EDM 2014: Simple Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
    <br/>
    <em> Jan Papousek, [...]</em>
    <br/>
    Pages 6-13 [
    <a href="uploads/[...].pdf">pdf</a>
    ]
    <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
   <authors> Jan Papousek, [...]</authors>
   <title>Adaptive Practice of [...]</title>
   <pages>6-13 [</pages>
 </record>
 <!--[...]-->
</results>
```

EDM 2014: Simple Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
   <br/>
   <em> Jan Papousek, [...]</em>
   <br/>
   Pages 6-13 [
   <a href="uploads/[...].pdf">pdf</a>
   ]
   <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
   <authors> Jan Papousek, [...]</authors>
   <title>Adaptive Practice of [...]</title>
   <pages>6-13 [</pages>
   <url>uploads/[...].pdf</url>
 </record>
 <!--[...]-->
</results>
```

EDM 2014: Simple Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
   <br/>
   <em> Jan Papousek, [...]</em>
   <br/>
   Pages 6-13 [
   <a href="uploads/[...].pdf">pdf</a>
   ]
   <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
  <authors> Jan Papousek, [...]</authors>
  <title>Adaptive Practice of [...]</title>
  <pages>6-13 [</pages>
  <url>uploads/[...].pdf</url>
  <header>Full Papers</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2014: Advanced Extraction

HTML Source

```html
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
   <br/>
   <em> Jan Papousek, [...]</em>
   <br/>
   Pages 6-13 [
   <a href="uploads/[...].pdf">pdf</a>
   ]
   <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
  <authors> Jan Papousek, [...]</authors>
  <title>Adaptive Practice of [...]</title>
  <pages>6-13 [</pages>
  <url>uploads/[...].pdf</url>
  <header>Full Papers</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2014: Advanced Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
  <!--[...]-->
  <div id="content">
    <!--[...]-->
    <strong>Online Proceedings</strong>
    <!--[...]-->
    <strong>Full Papers</strong>
    <!--[...]-->
    <p>Adaptive Practice of [...]
      <br/>
      <em> Jan Papousek, [...]</em>
      <br/>
      Pages 6-13 [
      <a href="uploads/[...].pdf">pdf</a>
      ]
      <!--[...]-->
    </p>
    <!--[...]-->
  </div>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>Jan Papousek, [...]</authors>
    <title>Adaptive Practice of [...]</title>
    <pages>6-13 [</pages>
    <url>uploads/[...].pdf</url>
    <header>Full Papers</header>
  </record>
  <!--[...]-->
</results>
```

EDM 2014: Advanced Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
   <br/>
   <em> Jan Papousek, [...]</em>
   <br/>
   Pages 6-13 [
   <a href="uploads/[...].pdf">pdf</a>
   ]
   <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
  <authors>Jan Papousek, [...]</authors>
  <title>Adaptive Practice of [...]</title>
  <pages>6-13</pages>
  <url>uploads/[...].pdf</url>
  <header>Full Papers</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2014: Advanced Extraction

HTML Source

```
<html xmlns="[...]" xml:lang="en">
 <!--[...]-->
 <div id="content">
  <!--[...]-->
  <strong>Online Proceedings</strong>
  <!--[...]-->
  <strong>Full Papers</strong>
  <!--[...]-->
  <p>Adaptive Practice of [...]
   <br/>
   <em> Jan Papousek, [...]</em>
   <br/>
   Pages 6-13 [
   <a href="uploads/[...].pdf">pdf</a>
   ]
   <!--[...]-->
  </p>
  <!--[...]-->
 </div>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
  <authors>Jan Papousek, [...]</authors>
  <title>Adaptive Practice of [...]</title>
  <pages>6-13</pages>
  <url>http://[...]uploads/[...].pdf</url>
  <header>Full Papers</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source (EDM 2014)

```
<html xmlns="[...]" xml:lang="en">
  <!--[...]-->
  <div id="content">
    <!--[...]-->
    <strong>Online Proceedings</strong>
    <!--[...]-->
    <strong>Full Papers</strong>
    <!--[...]-->
    <p>Adaptive Practice of [...]
      <br/>
      <em> Jan Papousek, [...]</em>
      <br/>
      Pages 6-13 [
      <a href="uploads/[...].pdf">pdf</a>
      ]
      <!--[...]-->
    </p>
    <!--[...]-->
  </div>
  <!--[...]-->
</html>
```

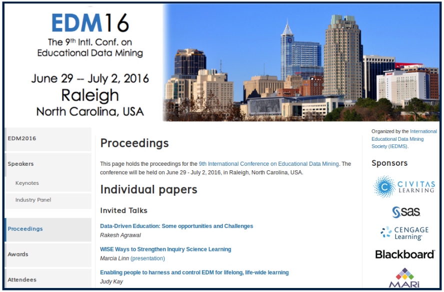
OXPath Expression (EDM 2014)

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

XML Output (EDM 2014)

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors> Jan Papousek, [...]</authors>
    <title>Adaptive Practice of [...]</title>
    <pages>6-13 [</pages>
    <url>uploads/[...].pdf</url>
    <header>Full Papers</header>
  </record>
  <!--[...]-->
</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
- 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

WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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

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

Sponsors

CIVITAS LEARNING
sas
CENGAGE Learning
Blackboard
MARi

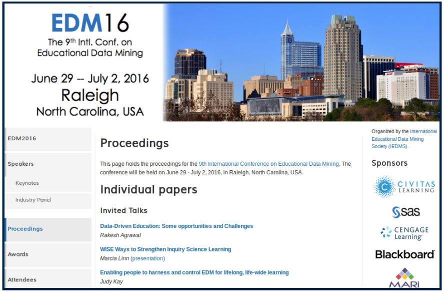
OXPath Expression

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

XML Output

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

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

WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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

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

Sponsors

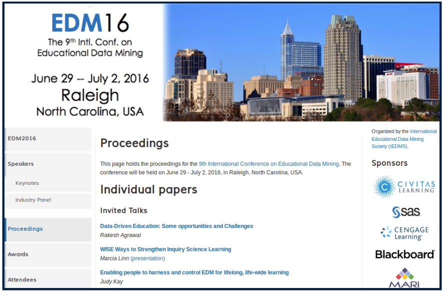
OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
   <authors>???</authors>
   <title>???</title>
   <pages>???</pages>
   <url>???</url>
   <header>???</header>
 </record>
 <!--[...]-->
</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
- 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

WISE Ways to Strengthen Inquiry Science Learning
Marcia Linn (presentation)

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

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

Sponsors

OXPath Expression

```
doc('http://edm2016.org/proceedings.html')
 ???:<record>
   [???:<authors=???>]
   [???:<title=???>]
   [???:<pages=???>]
   [???:<url=???>]
   [???:<header=???>]
```

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>???</authors>
    <title>???</title>
    <pages>???</pages>
    <url>???</url>
    <header>???</header>
  </record>
  <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```html
<html>
 <!--[...]-->
 <h1>Individual papers</h1>
 <h3>Invited Talks</h3>
 <p>
  <a id="[...]" class="citation_title"
      href="[...]">Data-Driven [...]</a>
  <!--[...]-->
  <span class="[...]title">9th [...]</span>
  <span class="[...]firstpage">2</span>
  <span class="[...]lastpage">2</span>
  <span class="[...]pdf_url">http[...]</span>
  <br/>
  <span class="[...]author">Ra[...]</span>
  <!--[...]-->
 </p>
 <!--[...]-->
</html>
```

OXPath Expression

```
doc('http://edm2016.org/proceedings.html')
 ???:<record>
   [???:<authors=???>]
   [???:<title=???>]
   [???:<pages=???>]
   [???:<url=???>]
   [???:<header=???>]
```

XML Output

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
   <authors>???</authors>
   <title>???</title>
   <pages>???</pages>
   <url>???</url>
   <header>???</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
<html>
  <!--[...]-->
  <h1>Individual papers</h1>
  <h3>Invited Talks</h3>
  <p>
    <a id="[...]" class="citation_title"
        href="[...]">Data-Driven [...]</a>
    <!--[...]-->
    <span class="[...]title">9th [...]</span>
    <span class="[...]firstpage">2</span>
    <span class="[...]lastpage">2</span>
    <span class="[...]pdf_url">http[...]</span>
    <br/>
    <span class="[...]author">Ra[...]</span>
    <!--[...]-->
  </p>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>???</authors>
    <title>???</title>
    <pages>???</pages>
    <url>???</url>
    <header>???</header>
  </record>
  <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
<html>
  <!--[...]-->
  <h1>Individual papers</h1>
  <h3>Invited Talks</h3>
  <p>
    <a id="[...]" class="citation_title"
        href="[...]">Data-Driven [...]</a>
    <!--[...]-->
    <span class="[...]title">9th [...]</span>
    <span class="[...]firstpage">2</span>
    <span class="[...]lastpage">2</span>
    <span class="[...]pdf_url">http[...]</span>
    <br/>
    <span class="[...]author">Ra[...]</span>
    <!--[...]-->
  </p>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>Rakesh Agrawal</authors>
    <title>???</title>
    <pages>???</pages>
    <url>???</url>
    <header>???</header>
  </record>
  <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```html
<html>
  <!--[...]-->
  <h1>Individual papers</h1>
  <h3>Invited Talks</h3>
  <p>
    <a id="[...]" class="citation_title"
        href="[...]">Data-Driven [...]</a>
    <!--[...]-->
    <span class="[...]title">9th [...]</span>
    <span class="[...]firstpage">2</span>
    <span class="[...]lastpage">2</span>
    <span class="[...]pdf_url">http[...]</span>
    <br/>
    <span class="[...]author">Ra[...]</span>
    <!--[...]-->
  </p>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```xml
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>Rakesh Agrawal</authors>
    <title>Data-Driven [...]</title>
    <pages>???</pages>
    <url>???</url>
    <header>???</header>
  </record>
  <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
<html>
  <!--[...]-->
  <h1>Individual papers</h1>
  <h3>Invited Talks</h3>
  <p>
    <a id="[...]" class="citation_title"
         href="[...]">Data-Driven [...]</a>
    <!--[...]-->
    <span class="[...]title">9th [...]</span>
    <span class="[...]firstpage">2</span>
    <span class="[...]lastpage">2</span>
    <span class="[...]pdf_url">http[...]</span>
    <br/>
    <span class="[...]author">Ra[...]</span>
    <!--[...]-->
  </p>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

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

EDM 2016: Adapting from EDM 2014

HTML Source

```
<html>
 <!--[...]-->
 <h1>Individual papers</h1>
 <h3>Invited Talks</h3>
 <p>
   <a id="[...]" class="citation_title"
       href="[...]">Data-Driven [...]</a>
   <!--[...]-->
   <span class="[...]title">9th [...]</span>
   <span class="[...]firstpage">2</span>
   <span class="[...]lastpage">2</span>
   <span class="[...]pdf_url">http[...]</span>
   <br/>
   <span class="[...]author">Ra[...]</span>
   <!--[...]-->
 </p>
 <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
 <!--[...]-->
 <record>
   <authors>Rakesh Agrawal</authors>
   <title>Data-Driven [...]</title>
   <pages>2-2</pages>
   <url>http://[...].pdf</ee>
   <header>???</header>
 </record>
 <!--[...]-->
</results>
```

EDM 2016: Adapting from EDM 2014

HTML Source

```
<html>
  <!--[...]-->
  <h1>Individual papers</h1>
  <h3>Invited Talks</h3>
  <p>
    <a id="[...]" class="citation_title"
        href="[...]">Data-Driven [...]</a>
    <!--[...]-->
    <span class="[...]title">9th [...]</span>
    <span class="[...]firstpage">2</span>
    <span class="[...]lastpage">2</span>
    <span class="[...]pdf_url">http[...]</span>
    <br/>
    <span class="[...]author">Ra[...]</span>
    <!--[...]-->
  </p>
  <!--[...]-->
</html>
```

OXPath Expression

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

XML Output

```
<?xml version="1.0" encoding="UTF-8"?>
<results>
  <!--[...]-->
  <record>
    <authors>Rakesh Agrawal</authors>
    <title>Data-Driven [...]</title>
    <pages>2-2</pages>
    <url>http://[...].pdf</ee>
    <header>Invited Talks</header>
  </record>
  <!--[...]-->
</results>
```

Table of Contents

1 The Role of OXPath in Smart Harvesting II

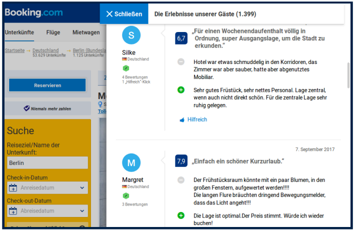
2 Maintaining Scientific Literature Databases

3 OXPath

4 Examples

5 Demonstration

Demonstration: Booking



Die Erlebnisse unserer Gäste (1.399)

Silke
Deutschland
4 Bewertungen
1 „Hilfreich"-Klick

6,7 „Für einen Wochenendaufenthalt völlig in Ordnung, super Ausgangslage, um die Stadt zu erkunden."

Hotel war etwas schmuddelig in den Korridoren, das Zimmer war aber sauber, hatte aber abgenutztes Mobiliar.

Sehr gutes Früstück, sehr nettes Personal. Lage zentral, wenn auch nicht direkt schön. Für die zentrale Lage sehr ruhig gelegen.

7. September 2017

Margret
Deutschland
3 Bewertungen

7,9 „Einfach ein schöner Kurzurlaub."

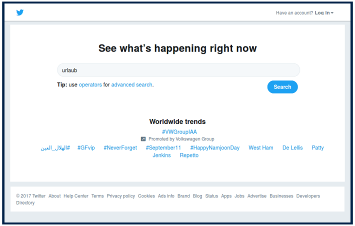
Der Frühstücksraum könnte mit ein paar Blumen, in den großen Fenstern, aufgewertet werden!!!!
Die langen Flure bräuchten dringend Bewegungsmelder, dass das Licht angeht!!!

Die Lage ist optimal.Der Preis stimmt. Würde ich wieder buchen!

OXPath Expression

```
doc("https://www.booking.com/hotel/de/metropolitan...")
  //a[@id='show_reviews_tab']/{click /}
  /(//*[@id='review_next_page_link']/{clkwithchange
       /})*{0,2}
  //div[contains(@class, 'review_list_block')]
    //li[contains(@class, 'review_item')]
         [not(contains(@class, 'featured_review_item'))]
         [not(@class= 'review_item_photo
         review_item_photo-p')]:<review>
  [? .//*[contains(@class, 'review_item_date')]
    :<date=normalize-space(.)>]
  [? .//*[contains(@class, 'review_item_review_score')]
    :<score=normalize-space(.)>]
  [? .//*[@class='review_item_header_content_container']
    :<title=normalize-space(.)>]
  [? .//*[contains(@class, 'review_item_review_content')]
    :<text=string-join(./p/text(), " ")>]
  [? .//*[@class='reviewer_country']
    :<country=normalize-space(.)>]
```

Demonstration: Twitter



Have an account? Log in ▾

See what's happening right now

urlaub

Tip: use operators for advanced search.

Search

Worldwide trends

#VWGroupIAA

Promoted by Volkswagen Group

#الهلال_العين #GFvip #NeverForget #September11 #HappyNamjoonDay West Ham De Lellis Patty Jenkins Repetto

© 2017 Twitter About Help Center Terms Privacy policy Cookies Ads info Brand Blog Status Apps Jobs Advertise Businesses Developers Directory

OXPath Expression

```
doc("https://twitter.com/search-home")
 //input[@id='search-home-input']/{'urlaub'}/{pressenter/}
 /(//div[contains(@class,'stream-footer')]
   /{mouseover /})*{0, 4}
 /.:<count=count(//li[@data-item-type='tweet'])>
 //li[@data-item-type='tweet']:<tweet>
   [? .//strong[@class='fullname show-popup-with-id ']
      :<user_name=string(.)>]
   [? .//a[starts-with(@class,
      'account-group')]/span[@class='username u-dir']
      :<user_id=string(.)>]
   [? .//a[@class="tweet-timestamp js-permalink js-nav
      js-tooltip"]/@title:<date=string(.)>]
   [? .//p[starts-with(@class,'TweetTextSize')]
      :<content=normalize-space(.)>]
   [? .//button[contains(@aria-describedby,
      'reply-count')]/span/span:<antworten=string(.)>]
   [? .//button[contains(@aria-describedby,
      'retweet-count')]/span/span:<retweets=string(.)>]
   [? .//button[contains(@aria-describedby,
      'favorite-count')]/span/span:<likes=string(.)>]
```