Literature Search

IN 3305
Created by Tomas Klos. Edited by Alexandru Iosup.
February 1, 2010

Parallel and Distributed Systems Groep
http://www.pds.ewi.tudelft.nl/
Introduction

• From the IN3305 study goals:
  “kennismaken met wetenschappelijke literatuur”
  “problemen oplossen door te zoeken in literatuur”

• What is
  “scientific literature”?

• To read or not to read?
• Literature is output and input
• Measuring and assessing Quality
• Useful sites
• Recommendations and tips
How to Talk About Books You Haven’t Read

• “There is more than one way not to read”
  • Not opening the book

• You cannot read everything
  • How many books can you read?
  • How many books can a librarian read?

• Librarians can talk about every book in the library (every book out of millions)

→ There exists a system to (not) read
Literature = output

• “Publish or perish”: quality / quantity
  • (“80% of all published papers are not cited”)
• Peer-review (for conferences, journals):
  • (double) blind review:
    - Accept, with/without (major) revisions
    - Reject
  • Acceptance rate, e.g. 25%
  • (Nature: 10% is reviewed)
• Measuring scientific output: “scientometrics”
Scientometrics

- Scientometrics, “measuring and analyzing science”
- Bibliometrics, “study or measurement of texts and information”
- In particular citation analysis:
  - Which papers cite a paper / does a paper cite?
  - Authority of authors, journals, papers
  - Same principle: Google PageRank
    - Web: network of sites, linking to each other
    - Science: network of papers, citing each other
World Wide Web
Citation Networks
Citation Databases

- Commercial:
  - ScienceCitation Index (Inf.Sci. Inst.)
  - Scopus (Elsevier)
- Free:
  - Google Scholar: better coverage than ISI
  - CiteSeer (computer science)
  - RePec (economics)
Indices

- **Journals**: Journal Impact Factor
- **Personal**: h-index (Hirsch, 2005):
  
  “I propose the index $h$, defined as the number of papers with citation number $\geq h$, as a useful index to characterize the scientific output of a researcher.”

A scientist has index $h$ if $h$ of his/her $N$ papers have at least $h$ citations each, and the other $(N - h)$ papers have no more than $h$ citations each.

- **Extensions**: g-index, h-b-index
Journal Impact Factor (JIF)

• Many journals have *no* impact factor
• JIF is the average number of citations in a given year, to papers in a journal in the 2 previous years.
• For journal $x$, 2008

$$JIF(x, 2008) = \frac{\text{number of citations in 2008 to papers in journal } x \text{ from the period 2006 – 2007}}{\text{Total number of papers in journal } x \text{ in the period 2006 – 2007}}$$

• What does an average value mean?
Journal Impact factors, 2004

2004 Science Journals Impact Factors (Bron: ISI)

- Highest JIF $\sim 30$
- Very high JIF $\geq 15$

$\geq 1$ citation/publication (last 2 years)
CS impact factors, 2005

2005 Impact Factor CS Journals (Bron: ISI)

JIF

Journal Rank

CS

Highest JIF ~8
Very high JIF ≥2

All

Highest JIF ~30
Very high JIF ≥15
Google Scholar

- “cited by”
- Relevant authors
- TU Delft SFX linking
- Import into bibtex
Guided local search and its application to the traveling salesman problem - all 14 versions

Local Search in Combinatorial Optimization - all 4 versions

New genetic local search operators for the traveling salesman problem - all 11 versions

Efficient local search with search space smoothing: a case study of the traveling salesman problem - all 6 versions

Local optimization and the traveling salesman problem - all 3 versions

A genetic local search algorithm for solving symmetric and asymmetric traveling salesman problems - all 8 versions

Efficient local search with search space smoothing: a case study of the traveling salesman problem (TSP) - all 6 versions

Local optimization and the traveling salesman problem - all 3 versions

http://scholar.google.com/scholar?num=50&hl=en&lr=&client=firefox-a&cites=13460080777608056294
From home: use vpn!
DBLP

- “lists more than one million articles” (April 2008)
- Indexes:
  - Authors
    - Now also “Faceted search”, “CompleteSearch”
  - Conferences
  - Journals
  - Series
  - Subjects
Vincent Conitzer

List of publications from the DBLP Bibliography Server - FAQ

Coauthor Index - Ask others: ACM DL/Guide - CiteSeer - CSB - Google - MSN - Yahoo

2007

61 EE Vincent Conitzer: Invited talk by winner of IFAMAS Victor Lesser Distinguished Dissertation Award. **AAMAS 2007**.


56 EE Vincent Conitzer: Limited verification of identities to induce false-name-proofness. TARK 2007: 102-111


2006


52 Vincent Conitzer: Computing Slater Rankings Using Similarities among Candidates. AAAI 2006


AAMAS 2007: Honolulu, HI, USA

Invited talks

- John Strassner:
  Using agents and autonomic computing to build next generation seamless mobility services. 1
  *Electronic Edition* (ACM DL) [BIBTEX]
- Jeffrey O. Kephart:
  Multagent systems for autonomic computing. 2
  *Electronic Edition* (ACM DL) [BIBTEX]
- Sarit Kraus:
  Automated negotiation in open environments; (by ACM/SIGART autonomous agents research award winner). 3
  *Electronic Edition* (ACM DL) [BIBTEX]
- Gal A. Kaminka:
  Robots are agents, too! 4
  *Electronic Edition* (ACM DL) [BIBTEX]

Embodied agents and architectures: full papers

- Pedro Siqueira, Marco Vals, Ana Paiva:
  What can I do with this?: finding possible interactions between characters and objects. 5
  *Electronic Edition* (ACM DL) [BIBTEX]
- John Thangarajah, James Harland, David Morley, Neil Yorke-Smith:
  Aborting tasks in BDI agents. 6
  *Electronic Edition* (ACM DL) [BIBTEX]
- Sebastian Sardina, Lin Padgham:
International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)

Agent Theories, Architectures, and Languages (ATAL)

International Foundation for Autonomous Agents and Multiagent Systems (IFAMAS)
ATAL Home Page
ATAL Wiki
ACM DL: AAMAS / Agents

AAMAS 2008: Estoril, Portugal
AAMAS 2008 Home Page

AAMAS 2007: Honolulu, HI, USA
Contents menu - AAMAS 2007 Home Page

AAMAS 2006: Hakodate, Japan
Contents menu - AAMAS 2006 Home Page

Done
<table>
<thead>
<tr>
<th>Year</th>
<th>Submitted</th>
<th>Accepted</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>531</td>
<td>122</td>
<td>23%</td>
</tr>
<tr>
<td>2006</td>
<td>553</td>
<td>127</td>
<td>23%</td>
</tr>
<tr>
<td>2001</td>
<td>248</td>
<td>66</td>
<td>27%</td>
</tr>
<tr>
<td>1998</td>
<td>180</td>
<td>57</td>
<td>32%</td>
</tr>
<tr>
<td>1997</td>
<td>171</td>
<td>59</td>
<td>35%</td>
</tr>
</tbody>
</table>
Harzing’s Publish or Perish

- Uses google scholar data
- Calculates many indices
  - Number of citations (also per year / article / author / ...)
  - Hirsch’s h-index
  - Zhang’s e-index (excess in h-index set)
  - Egghe’s g-index
  - ...
<table>
<thead>
<tr>
<th>Rank</th>
<th>Per Year</th>
<th>Authors</th>
<th>Title</th>
<th>Year</th>
<th>Publication</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64.33</td>
<td>V. Conter, T. Sandh...</td>
<td>Now complexity results about Nash ...</td>
<td>2004</td>
<td>Games and Economic Behavior</td>
<td>Elsevier</td>
</tr>
<tr>
<td>2</td>
<td>17.22</td>
<td>V. Conter, T. Sandh...</td>
<td>Complexity of mechanism design</td>
<td>2003</td>
<td>Anik reprint</td>
<td>anik.org</td>
</tr>
<tr>
<td>99</td>
<td>12.30</td>
<td>V. Conter, T. Sandh...</td>
<td>Universal voting protocol breaks to ...</td>
<td>2002</td>
<td>Anik reprint</td>
<td>anik.org</td>
</tr>
<tr>
<td>67</td>
<td>9.67</td>
<td>V. Conter, T. Sandh...</td>
<td>Complexity of manipulating elections</td>
<td>2001</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>76</td>
<td>9.23</td>
<td>V. Conter, T. Sandh...</td>
<td>Nonexistence of voting rules that ...</td>
<td>2001</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>65</td>
<td>10.28</td>
<td>V. Conter, T. Sandh...</td>
<td>Who are elections with few ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>64</td>
<td>9.08</td>
<td>V. Conter, T. Sandh ...</td>
<td>How many candidates are needed ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>18</td>
<td>4.29</td>
<td>V. Conter, T. Sandh...</td>
<td>Computing Shapley values, manipulations</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>49</td>
<td>8.17</td>
<td>V. Conter, T. Sandh...</td>
<td>Vote evaluation, complexity and strategy</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>67</td>
<td>5.75</td>
<td>V. Conter, T. Sandh...</td>
<td>Complexity of determining nonempty ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>65</td>
<td>4.53</td>
<td>V. Conter, T. Sandh...</td>
<td>Computational modeling of the rev ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>41</td>
<td>3.98</td>
<td>V. Conter, T. Sandh...</td>
<td>Computing Social rankings using ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>55</td>
<td>5.03</td>
<td>V. Conter, T. Sandh...</td>
<td>Communication complexity of ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>54</td>
<td>6.00</td>
<td>V. Conter, T. Sandh...</td>
<td>Complexity of constructing solutions ...</td>
<td>2000</td>
<td>Artificial Intelligence</td>
<td>Elsevier</td>
</tr>
<tr>
<td>24</td>
<td>5.67</td>
<td>V. Conter, T. Sandh...</td>
<td>Coalition games in open anonymous ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>72</td>
<td>4.17</td>
<td>V. Conter, T. Sandh...</td>
<td>Combinatorial auctions with ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>41</td>
<td>4.03</td>
<td>V. Conter, T. Sandh...</td>
<td>Improved bounds for computing ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>60</td>
<td>5.09</td>
<td>V. Conter, T. Sandh...</td>
<td>Common voting rules as maximum ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>50</td>
<td>6.00</td>
<td>V. Conter, T. Sandh...</td>
<td>Combinatorial auctions with lower ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>29</td>
<td>4.14</td>
<td>V. Conter, T. Sandh...</td>
<td>Self-interested automated mechanisms ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>87</td>
<td>4.60</td>
<td>V. Conter, T. Sandh...</td>
<td>Complexity of (labeled) dominance ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>27</td>
<td>3.38</td>
<td>V. Conter, T. Sandh...</td>
<td>Applications of automated mechanisms ...</td>
<td>2000</td>
<td>UAI-01 workshop on Bayesian ...</td>
<td>UAI-01</td>
</tr>
<tr>
<td>56</td>
<td>6.50</td>
<td>V. Conter, T. Sandh...</td>
<td>Worst-case optimal redistribution ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
<tr>
<td>17</td>
<td>5.00</td>
<td>V. Conter, T. Sandh...</td>
<td>Computing the optimal strategy ...</td>
<td>2000</td>
<td>Proceedings of the National ...</td>
<td>aaas.org</td>
</tr>
</tbody>
</table>
Off-topic: How to Game the Citation System? (part of) Collaboration graph
All authors with Erdős number 1
Collaboration Graph Degree Distribution

Erdős
Collaboration Graph: Connected Components Distribution

The number of connected components

Giant Component

size
Interested?

- Erdős Number Project
  http://www.oakland.edu/enp/
- Kevin Bacon Oracle
  http://oracleofbacon.org/
- Mark Newman: “who is the best connected scientist?”
Literature = input

• Citations
  • Place your work in context
  • Give credit to previous work
  • Support your arguments
  • Show your marginal contribution
  • Prevent plagiarism
• Read what you cite! (prevent superfluous citing)
  This does NOT mean:
  • “You should read everything”
  • “You cannot also read what you don’t cite”
Sources: peer-reviewed

- **Textbook/monograph:** for teaching and background
  - Complete treatment of a topic
  - Cite a textbook? Mention chapter or page number

- **Journal article**
  - More space, detail, thorough than conference paper
  - Sometimes old news at publication date (lag)

- **Paper in edited volume:**
  - Multiple papers, review of state-of-the-art
  - Cite individual papers

- **Paper in conference proceedings**
  - Recent results
  - Conference quality; publisher of proceedings?
Sources: not peer-reviewed

- Working papers, Preprints
  - Up-to-date, spread ideas
  - “Open access”
  - Computing Research Repository (CoRR)
    http://arxiv.org/corr/home
- Websites
- ‘Personal communication’
Quality?


- SCIgen - An Automatic CS Paper Generator
  http://pdos.csail.mit.edu/scigen/
  accepted (non-reviewed) for: 2005 World Multi-Conference on Systemics, Cybernetics and Informatics
  (another one: an Elsevier journal!)
Finding Sources

- Browse:
  - DBLP: http://dblp.uni-trier.de/
  - CiteSeer: http://citeseerx.ist.psu.edu/
  - Google Scholar: http://scholar.google.com/
- Author homepages
- Follow links and citations (forward and backward)
TU Delft Library

- Search
  - http://www.library.tudelft.nl/ws/search/
  - e.g. “information by subject” -> computer science
- TUlib
  - “how to find and use scientific information”
  - http://www.library.tudelft.nl/tulib/
How to Talk About Books You Haven’t Read

There exists a system to (not) read

1. Know where to find the sources
   - Trustworthy: DBLP, ACM DL, Google Scholar
   - Less trustworthy: CoRR, ...

2. Know how to find the good sources
   - Number of citations: ACM DL, Google Scholar
   - H-index: Publish or Perish (the program)
   - Try to avoid or weight citation cliques

3. Select from the good sources
Questions?