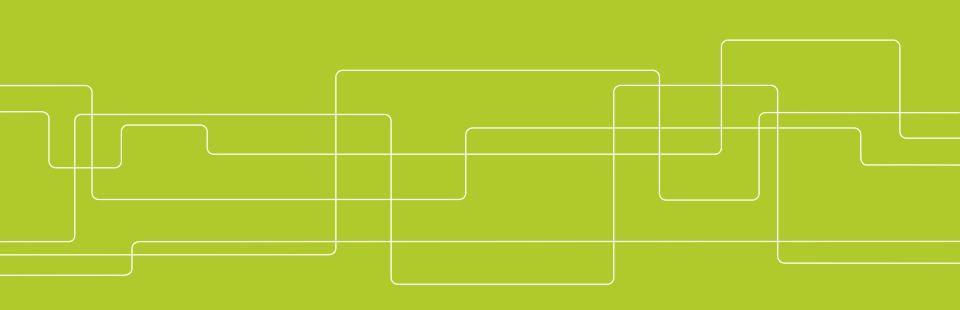


# A comparison of coverage and citation matching in Google Scholar, Web of Science and Scopus

Peter Sjögårde KTH Royal Institute of Technology, ECE School, Publication Infrastructure





- Why don't you use Google Scholar for bibliometrics?
- My publications get more citations in Google Scholar



### Why don't you use Google Scholar for bibliometrics?

- No access to structured data
- Questionable quality of data (duplicates)
- Possible to manipulate
- Contents not defined
- Contains non-peer reviewed papers, popular science, educational material



Aguillo, Isidro. "Is Google Scholar Useful for Bibliometrics? A Webometric Analysis." *Scientometrics* 91, no. 2 (2012): 343–351. doi:10.1007/s11192-011-0582-8.

Lopez-Cozar, Emilio Delgado, Nicolas Robinson-Garcia, and Daniel Torres-Salinas. "Manipulating Google Scholar Citations and Google Scholar Metrics: Simple, Easy and Tempting." arXiv:1212.0638 [cs], December 4, 2012. http://arxiv.org/abs/1212.0638.



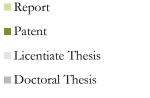
## My publications get more citations in Google Scholar

- Is this true?
- If true:
  - How big is the difference?
  - Where do the citations come from?
  - Why are they not counted in Web of Science and Scopus?

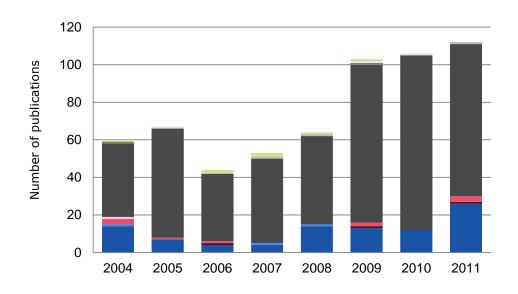


### Case study: Embedded electronics and computer systems at KTH

Publications in local repository 2004-2011



- Conference Proceedings (editor)
- Conference Paper
- Collection/Anthology (editor)
- Chapter in book
- Book Review
- ■Book
- Article, review/survey
- Article in journal (other)
- Article in journal (peer reviewed)





### **Data**

- Publications from 2009
- Peer reviewed Articles and Conference papers
- 91 publications



### **Publications and Citations**

	<b>Publications</b>	Coverage C	itations C	it./Publ.
Web of Science	58	64%	73	1.26
Scopus	77	85%	206	2.68
Google Scholar	91	100%	714	7.85



### Where do the citations come from?

Random sample of 100 of the GS citations

Manually search in Google Scholar for each of the 100 documents citing the publications published by the UoA. Noted document type and source.

Manually search in Web of Science and Scopus for the same documents, to see if they were indexed in these databases and if so, if they had resulted in a citation.



VETH VETHSKAP OCH KONST A New Methodology for Constructing a Publication-Level Classification System of Science

#### Referencing object

most popular classification is level classification systems tons: They offer only a list have difficulties with multi these limitations, we lettreonstructing classification

Tor Science and rechnology studies as well as Kevin Boyack, Lutz Bornmann, Dick Klavans, and Javier Ruiz Castillo for helpful discussion and feedback. We also thank the three anonymous reviewers for comments on this

paper.

ntroduction
In bibliometric and s

esterns of science are an inestern of science assigns ja research areas. Such a symplify literature search, it is of scientific disciplinmearch evaluations.

This many introduces

Bostoni March 1, 2012; netted O 2012 ASSAT • Pubbled oil Library (villayoillasSfrary cons.) References

Ahlgren, P., & Colliander, C. (2009). Document-document similarity approaches and science mapping: Experimental comparison of five approaches. Journal of Informetrics, 3(1), 49–63.

Archambault, É., Beauchesne, O.H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In E.C.M. Noyons, P. Ngulube, & J. Leta (Eds.), Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (pp. 66–77).

classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. Journal of the American Society for Information Science and Technology, 60(9), 1823–1835.

Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. Physical Review E, 74(1), 016110.

Rosvall, M., & Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. In Proceedings of the National Academy of Sciences, 105(4), 1118–1123.

osvail, M., & Bergstrom, C.1. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. PLoS ONE, 6(4), e18209.

Rotta, R., & Noack, A. (2011). Multilevel local search algorithms for modularity clustering. Journal of Experimental Algorithmics, 16(2), article 2.3.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. Journal of the American Society for Information Science and Technology, 60(3), 571–580.

Small H (2006) Tracking and predicting arough proces in colonea

#### Reference

### Citation

andal VD Guil

Google

● Articles (✓ include patents)
 ○ Case law

Stand on the shoulders of giants

### Target object

Maps of random walks on complex networks reveal community structure

Martin Resvali<sup>e†</sup> and Carl T. Bergst

"Department of Biology, University of Steinbergton, Soutile, WIG 90195-1800, and "Sunta the Institute, 1999 hyde Rank Blood, Sonta th Edited by Brian Skyrms, University of California, Indina, CA, and approved December 10, 2007 December for review July 21, 2007).

To comprehend the multiporthic apparentation of large-scale biological and toold spictors, one invasion as intermediate theoretic largest and toold spictors, one invasion as intermediate theoretic region of the control of the control of the control of the control rection detectors. We use the probability flows of endeem walks or a antenior, as a possion for information them to the real spices and to tion of the probability flows. The search is a may that been impellitudtions of the probability flows. The search is a may that been impellitudtion, the flowests the enthal by reality a many of steediletical control of the control of the control of the control partness. He discusses a molectoric emporative with thirt their very demonstrating in our and degree of retargetion risks the enphysics, climitary, an order to the control of the control of the physics, climitary, an order to the control of the co

thatanny | compressor | information theory | map of a

Designation of some terms on effectivents, analyzing and applications of administration, and section common constraints. Some different common constraints, these distractions can provide (1, 2), but not an extension, these distractions can provide (1, 2), the case as constraints, these distractions conducted the constraints of the con

Northook Blags and Coding Theory
In this article, we one maps to describe the detauries is links and nodes in directed, weighted networks that replevel interactions arrang the solumin of a system. To

in system because, we need to innovate mation on the network. We therefore idea compose the network by fielding an effic description of how information flows on a of onder assume which information flows o be aggregated and described as a single we Neffman Coding. A straightforward method of giving name to be a books is to use a Haffman code (18). Haffman codes sove year

The article of the County (Section).

The where comparedness should be addressed it mail research estimates about the paint of the County (Section 2) and the comparedness should be addressed it mail research estimates as the county of the article county of the county

-1128 | PRAS | January 28, 2006 | upl 105 | no. 4

ww.proc.org.(g),460,78,5075 peak 8786



## Are the references in Google Scholar hitting the correct target?

Comment	Count
Correct reference in referencing object	95
Not possible to verify	4
Incorrect citation	1
Total	100



### **Document type of referencing objects**

Document type	Count
Conference papers	65
Articles	12
Doctorate thesis	12
Chapters in books	3
? (Chinese)	3
No source	1
Report	1
Student thesis	2
Duplicates	1
Total	100



## Publishers of the referencing conference papers and articles

Source per document type	N
Conference Paper	65
IEEE	55
ACM	2
Now Publishers Inc.	1
World Academic Publishing	1
Academic Press Inc.	1
Foundation of Computer Science	1
Article in Journal	12
Springer	3
Elsevier	3
ACM	2
Now Publishers Inc.	1
World Academic Publishing	1
Academic Press Inc.	1
Foundation of Computer Science	1
Total	77



## Coverage of referencing objects and target objects in databases

	Google Scholar We	b of Science	Scopus
Referencing object in database	77	19	65
Both referencing and target object in database	77	16 <b>↑</b>	62 <b>↑</b>
Reference matched to target object resulting in a citation	77	9	31



# References not matched even though both referencing and target object are indexed in the database

	Web of Science	Scopus
Correct reference	6	1
Incorrect reference	1	30
Totalt	7	31



### Example of incorrect reference in Scopus

#### Reference in orginial publication

#### REFERENCES

- C. Claus, B. Zhang, W. Stechele, L. Braun, M. Hübner, J. Becker, A multi-platform controller allowing for maximum Dynamic Partial Reconfiguration throughput, Proc. of the International Conference on Field Programmable Logic and Applications, pp. 535-539, 2008.
- [2] Liu Ming, W. Kuehn, Lu Zhonghai, A. Jantsch, Run-time Partial Reconfiguration speed investigation and architectural design space exploration, Proc. of the International Conference on Field Programmable Logic and Applications, pp. 498-502, 2009.
- [3] C. Claus, R. Ahmed, F. Altenried, W. Stechele, Towards rapid dynamic partial reconfigurationin video-based driver assistance systems, 6th Internation Symposium of Reconfigurable Computing: Architectures, Tools and Applications, pp. 55-67, 2010
- [4] A. Usman, M. B. Malik, K. Munawar, FPGA/soft-processor based realtime object tracking system, Proc. of the 5th Southern Conference on Programmable Logic, pp. 33-37, 2009.
- [5] K. Yamaoka, T. Morimoto, H. Adachi, K. Awane, T. Koide, H.J. Mattausch, Multi-object tracking VLSI architecture using image-scan based region growing and featured matching, Proc. of the International Symposium on Circuits and Systems, ISCAS 2006.
- [6] Alpha Data, http://www.alpha-data.com
- [7] Xilinx, Virtex-4 FPGA user guide, ug070 v2.6, 2008.
- [8] Xilinx, Virtex-4 FPGA configuration user guide, ug071 v1.11, 2009.
- [9] Xilinx, Early access partial reconfiguration user guide, ug208 v1.2, 2008.

#### Reference in Scopus

Claus, C., Zhang, B., Stechele, W., Braun, L., Hübner, M., Becker, J.

A multi-platform controller allowing for maximum dynamic partial reconfiguration throughput

(2008) Proceedings - 2008 International Conference on Field Programmable Logic and Applications, FPL, art. no. 4630002, pp. 535-538. Cited 27 times ISBN: 978-142441961-6 doi: 10.1109/FPL.2008.4630002 Ø▶▶ KTHB Ming, L., Kuehn, W., Zhonghai, L., Jantsch, A. (2009) Proc. of the International Conference on Field Programmable Logic and Applications, pp. 498-502. Ø▶▶ KTHB Claus, C., Ahmed, R., Altenried, F., Stechele, W. (2010) 6th Internation Symposium of Reconfigurable Computing: Architectures, Tools and Applications, pp. 55-67. Cited 4 times. Ø▶▶ KTHB Usman, A., Malik, M.B., Munawar, K. (2009) Proc. of the 5th Southern Conference on Programmable Logic, pp. 33-37, Cited 2 times. Ø▶▶ KTHB Yamaoka, K., Morimoto, T., Adachi, H., Awane, K., Koide, T., Mattausch, H.J. (2006) Proc. of the International Symposium on Circuits and Systems, ISCAS Ø▶▶ KTHB Ø▶▶ KTHB (2008) Virtex-4 FPGA User Guide. Cited 47 times. Xilinx, ug070 v2.6 Ø▶▶ KTHB (2009) Virtex-4 FPGA Configuration User Guide. Cited 28 times. Xilinx, ug071 v1.11 Ø▶▶ KTHB (2008) Early Access Partial Reconfiguration User Guide. Cited 91 times Xilinx, ug208 v1.2 Ø▶▶ KTHB



# References not matched even though both referencing and target object are indexed in the database

	Web of Science	Scopus
Correct reference	6	1
Incorrect reference	1	30
Totalt	7	31



### Example of references not resulting in a citation

#### Target object in Web of Science

Title: Development and experimental verification of analytical models for printable interdigital capacitor sensors on paperboard

Author(s): Feng, Yi; Hallstedt, Julius; Chen, Qiang; et al.

Book Group Author(s): IEEE

Conference: 8th IEEE Conference on Sensors Location: Christchurch, NEW ZEALAND Date: OCT 25-28, 2009

Sponsor(s): IEEE Sensors Council

Source: 2009 IEEE SENSORS, VOLS 1-3 Pages: 1034-1039 DOI: 10.1109/ICSENS.2009.5398531 Published: 2009

#### Reference in Web of Science

Title: Development and experimental verification of analytical models for printable inter-digital capacitor sensors on paperboard

Author(s): Yi, F.; Hallstedt, J.; Qiang, C.; et al.

Conference: Proc. IEEE Sensors

Source: PIEEE SENSORS Pages: 1034-1039 Published: 2009



### Summary

- Google Scholar covers more publications published by the UoA and more of the referencing objects.
- The sources of the citations in GS come from verifiable sources. A majority come from large publishers.
- Web of Science have a poor coverage of publications published by the UoA and of the referencing objects.
- The citation-matching algorithm in Web of Science systematically miss citations to conference papers.
- Scopus has a better coverage of publications published by the UoA and of the referencing objects.
- A large proportion of the references in referencing conference papers in Scopus are erroneously indexed, resulting in loss of citations.



### **Conclusions**

- Non of the sources proved reliable for evaluation of the analyzed Unit of Assessment.
- It is problematic to use bibliometrics for evaluation of research groups that publish a high proportion of their papers in conference papers.
- When using Web of Science and Scopus, the impact of conference papers seems to be systematically underestimated.
- Be cautious when drawing conclusions about the impact of conference papers from bibliometric data.