How To Ensure That Your Papers Are Widely Cited

Frits Vaandrager

Institute for Computing and Information Sciences
Radboud University Nijmegen

ICIS Colloquium, 4 December 2006

Outline

Outline

Outline

Outline

Outline

Outline

The Quest for Excellence

> *"All research at the Radboud University Nijmegen is leading. In external research assessments it is rated at least very good (4) on the criteria quality, productivity, vitality and relevance."*

Strategic Plan Radboud University Nijmegen 2005-2009

Defining Productivity

"The committee assesses academic productivity by relating the output (the number of publications in total and in each category) to the input of human resources."

Assessment of Research Quality, Protocol 1998, VSNU

Productivity Definition of Committee Steels

$$Productivity = \frac{Output}{Input}$$

where

- $Output$ = number of academic publications, with weight 2 for PhD theses and books
- $Input$ = full time equivalents in research, not counting PhD students

1 research monograph of 760pp = 2 journal papers of 2pp
1 postdoc = 2.5 faculty

Analysis

Analysis

- It is very tempting to use science indicators that are "automatically" computed, even for people who should know

Analysis

- It is very tempting to use science indicators that are "automatically" computed, even for people who should know
- Indicators often unreliable

Analysis

- It is very tempting to use science indicators that are "automatically" computed, even for people who should know
- Indicators often unreliable
- Shirking of responsibility

Citation Analysis for Computer Science?

There is *some* correlation between (perceived) quality of researchers and the number of citation to their work. E.g.,

- All Dutch groups that were rated excellent on one or more criteria in previous assessment have representative in Top50 of most cited Dutch Computer Scientists according to Citeseer.
- All researchers in Top16 working in Dutch institute belonged to group that was rated excellent on one or more criteria.

Less correlation between Citeseer list and outcome last assessment.

Practical Problems (apart from Conceptual Ones)

- Citeseer and Google Scholar full of junk and mistakes; criteria for inclusion unclear
- ISI incomplete

E.g., work Lynch, Segala & Vaandrager on *Hybrid I/O Automata* cited 96× in Citeseer, 215× in Google Scholar and 20× in ISI

For many CS researchers coverage ISI better than Citeseer

Which Citations Should Count?

Which Citations Should Count?

- If we only count citations from ISI, LNCS, ACM and part of IEEE, for some groups coverage will be lower than for others.
  Coverage AI? Computational linguistics? CS is intrinsically multidisciplinary

Which Citations Should Count?

- If we only count citations from ISI, LNCS, ACM and part of IEEE, for some groups coverage will be lower than for others.
  Coverage AI? Computational linguistics? CS is intrinsically multidisciplinary
- Citations to technical reports?
  No!?

Which Citations Should Count?

- If we only count citations from ISI, LNCS, ACM and part of IEEE, for some groups coverage will be lower than for others.
  Coverage AI? Computational linguistics? CS is intrinsically multidisciplinary
- Citations to technical reports?
  No!?
- Citations to papers in e.g. EATCS Bulletin?
  No!? How about 321 citations to Coalgebra Tutorial by Jacobs & Rutten?

Which Citations Should Count?

- If we only count citations from ISI, LNCS, ACM and part of IEEE, for some groups coverage will be lower than for others.
  Coverage AI? Computational linguistics? CS is intrinsically multidisciplinary
- Citations to technical reports?
  No!?
- Citations to papers in e.g. EATCS Bulletin?
  No!? How about 321 citations to Coalgebra Tutorial by Jacobs & Rutten?
- Citations to papers in obscure workshop?
  No!? Still proceedings of such workshops are often published as LNCS

What Should We Count? (cnt)

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1795 citations to Barendregt's book in Google Scholar?

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1795 citations to Barendregt's book in Google Scholar?

- Citations from course notes?

  No!? But I am proud if my papers are discussed in courses!

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1795 citations to Barendregt's book in Google Scholar?
- Citations from course notes?

  No!? But I am proud if my papers are discussed in courses!
- Self citations?

  No!? But what if other group builds on idea that someone contributed in joint paper? Isn't this research impact?

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1795 citations to Barendregt's book in Google Scholar?
- Citations from course notes?

  No!? But I am proud if my papers are discussed in courses!
- Self citations?

  No!? But what if other group builds on idea that someone contributed in joint paper? Isn't this research impact?
- Should all citations have equal weight?

What Should We Count? (cnt)

- Citations to books?

  No!? How about 1795 citations to Barendregt's book in Google Scholar?
- Citations from course notes?

  No!? But I am proud if my papers are discussed in courses!
- Self citations?

  No!? But what if other group builds on idea that someone contributed in joint paper? Isn't this research impact?
- Should all citations have equal weight?
- How to handle different citation behavior in different subfields?

How To Define "Performance Indicators"?

How To Define "Performance Indicators"?

- Total number of citations?

  Hard to find; may be inflated by small number of big 'hits', which may not be representative of the individual if he/she is coauthor with many others

How To Define "Performance Indicators"?

- Total number of citations?

  Hard to find; may be inflated by small number of big 'hits', which may not be representative of the individual if he/she is coauthor with many others
- Citation per paper?

  Hard to find; penalizes high productivity; in CS often paper first appears in proceedings, then as TR, then in journal

How To Define "Performance Indicators"?

- Total number of citations?

  Hard to find; may be inflated by small number of big 'hits', which may not be representative of the individual if he/she is coauthor with many others
- Citation per paper?

  Hard to find; penalizes high productivity; in CS often paper first appears in proceedings, then as TR, then in journal
- Hirsch index?

  A scientist has *h-index h* if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.
  Difficult to compare individuals with different scientific age

Typical Arguments Used in Discussion

- *"True quality insensitive for precise definition"*
  A genius will surface irrespective of precise definition.
  But for most of us, choice of definition determines score 5, 4 or 3, i.e. life or death according to RU strategic plan.
- *"CS discipline should grow up"*
  Grow up to become what? Physicists?
  What if Dutch CS grows up and rest of the world doesn't?

CWTS Study

Goal:
Methodological study aimed at development of bibliometric indicators of research performance in various subfields of Computer Science

Basic approach:
Extension of ISI database with publications from LNCS, ACM and (selection of) IEEE. Analyze citations from extended database to publications of Dutch Computer Scientists from period 1996-2001 that are included in database.

Crown indicator:
CPP/FCSm, where CPP is the average number of citations per paper and FCSm is the Average Field Citation Score.

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.
3. Confirms importance of conference proceedings

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.
3. Confirms importance of conference proceedings
4. Citation impact of Dutch academic CS groups significantly above world average, especially last 2 years

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.
3. Confirms importance of conference proceedings
4. Citation impact of Dutch academic CS groups significantly above world average, especially last 2 years
5. Impact ICIS researchers only slightly above world average.

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.
3. Confirms importance of conference proceedings
4. Citation impact of Dutch academic CS groups significantly above world average, especially last 2 years
5. Impact ICIS researchers only slightly above world average.
6. Small fraction of papers responsible for most citations; average CPP is about 5; 30% of papers not cited.

Conclusions CWTS Study

1. Expanding CWTS-WoS database cumbersome; problems to extract references from pdf file in 17%.
2. Average coverage of expanded database 35%; for 25% of groups coverage below 28%.
3. Confirms importance of conference proceedings
4. Citation impact of Dutch academic CS groups significantly above world average, especially last 2 years
5. Impact ICIS researchers only slightly above world average.
6. Small fraction of papers responsible for most citations; average CPP is about 5; 30% of papers not cited.
7. Correlation between QANU quality ratings and citation indicators weak.

Should We Use Citation Analysis At All?

Yes! In fact, we already do. ISI, Citeseer and Google Scholar are used widely for evaluation purposes and help to improve quality of evaluations

Browsing for one hour through citations to publications of researcher (or group) provides a lot of insight!

Vision

Google Scholar like system that allows one to

- Enter set of target (cited) publications of individual or group
- Specify well-defined set of source (citing) publications, e.g., LNCS, ACM, IEEE, Elsevier,..
- Browse on-line through citing papers, both in and outside well-defined set
- Quantitative information *only* as catalyst for thought
- Open!

Qualitative versus Quantitative

Assuming that evaluation committee does a proper job (!), qualitative citation analysis avoids some problems quantitative approach

- Goal displacement
- Manipulation of own citation scores

How to Become Widely Cited (in a Civilized Manner)

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea
5. Mention conjectures, open problems, etc

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea
5. Mention conjectures, open problems, etc
6. Publish in ISI, LNCS, ACM or IEEE journal/proceedings

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea
5. Mention conjectures, open problems, etc
6. Publish in ISI, LNCS, ACM or IEEE journal/proceedings
7. Make your papers available on line

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea
5. Mention conjectures, open problems, etc
6. Publish in ISI, LNCS, ACM or IEEE journal/proceedings
7. Make your papers available on line
8. Explain relationship of your work to that of others

How to Become Widely Cited (in a Civilized Manner)

1. Produce good results and well-written papers!
2. Quality more important than quantity
3. Think twice before joining as coauthor
4. Be either the first or the last in a subarea
5. Mention conjectures, open problems, etc
6. Publish in ISI, LNCS, ACM or IEEE journal/proceedings
7. Make your papers available on line
8. Explain relationship of your work to that of others
9. Communicate with other researchers about your work

Conclusions

Conclusions

1. Coverage CWTS study too low

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups
3. Citations to publications not in database should be taken into account

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups
3. Citations to publications not in database should be taken into account
4. Citation information will be used in future research evaluations

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups
3. Citations to publications not in database should be taken into account
4. Citation information will be used in future research evaluations
5. Be prepared!

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups
3. Citations to publications not in database should be taken into account
4. Citation information will be used in future research evaluations
5. Be prepared!
6. As consequences of research assessments are getting so big, their quality should improve accordingly; evaluation committees should properly justify their judgment

Conclusions

1. Coverage CWTS study too low
2. Acceptable coverage ($> 80\%$) possible for most groups
3. Citations to publications not in database should be taken into account
4. Citation information will be used in future research evaluations
5. Be prepared!
6. As consequences of research assessments are getting so big, their quality should improve accordingly; evaluation committees should properly justify their judgment
7. Quantitative information only as catalyst for thought; emphasis on qualitative citation analysis