Local and Regional Impact of Open Access Journal Articles: Citation analysis on research papers by Korean authors

Jeong-Wook Seo. Seoul National University College of Medicine
- Professor, Department of Pathology
- WHO Collaborating Centre for Health Information and Library Services
- CEO, Creative Commons Korea

Hosik Chung. Naver corporation

Jungmin Yun. Ajou University

Honam Choi. Korea Institute of Science and Technology Information (KISTI)

※ Support in part by KISTI(Korean Institute of Science and Technology Information) and NAVER
Content

1. What about Open Access
2. My approach on Open Access
3. Web of Science
4. NAVER Academic
5. Interpretation & Implications
1. What about Open Access?
How many journals are open access?

OA journals comprise A % of all SCI/SSCI/AHCI journal titles?

OA journals comprise B % of all Scopus journal titles?

OA journals comprise C % of all journal titles available in my library?

D % of all articles by Korean researchers are published in OA journals?

E % of all references from published articles by Korean authors are from OA journals?

- Journals at available open access registries: DOAJ (9,709), PMC (1,059), Scopus (2,583)
- Journals lists available from open access publishers: Biomedcentral (459)
- Journals published by international organization: World bank, WHO, FAO,
- Potential open access journals are not counted as open access.
- We listed journal titles cited more than once by Korean authors first and then checked whether they are OA journal as defined above.
OA journals comprise % of all SCI-, SSCI-, AHCI or % of all Scopus journal titles?

- **A** WoS: 9.8%
  - 1,216 (9.8%) journal titles are open access among 12,422 indexed by ISI.
  - 12.2% of SCI journals (982/ 8,040)
  - 11.1% of SSCI journals (165/ 2,986)
  - 4.9% of AHCI (69/ 1,396)

- **B** Scopus: 12.4%
  - 2,583 (12.4%) journal titles are open access among 20,874 indexed by Scopus.
  - 17.7% of Life science journals (758/ 4,290)
  - 15.9% of Health Science journals (1,056/ 6,656)
  - 11.4% of Physical science journals (762/6,683)
  - 7.5% of Social science journals (554/ 7,394)
  - 5.8% of Art and Humanities journals (143/2,445)
OA journals comprise [C] % of all journal titles available in my library?

- This figure is totally different in different libraries
- Even different in the same library if they count in different ways

[C] Just an example: 21% at SNU Medical library in 2013

- Open Access, 5144, 21%
- Subscribed Science journals, 11753, 49%
- Subscribed medical journals, 2656, 11%
- Local journals, 200, 1%
- Trial journals, 1967, 8%
- Subscribed journals shared, 2327, 10%
% of all articles by Korean researchers are published in OA journals?

6% 23,813 (6.2%) of articles among 387,113 articles are in OA journals during 21 years (1991–2011).

*The number of journals is the lowest estimate.
**The proportion of OAJ should be higher than 6%.

Toll, 6043, 94%

OAJ, 23,813, 6.2%

*The number of article may be 10% higher.
**The proportion of OAJ is a close approximate.
“USE of open access articles” is our question!

Journal titles

1. How many journal titles are cited in our research?
2. What proportion is from open access?

Cited articles

1. How many articles are cited from OA journals?
2. My Approach on Open Access
Why do we open?

Sports for World Cup?
- To make another best journal in the world?
- To develop an exclusive & premium information?

Sports for All?
- To provide practical solutions on Sciences for many?
- To meet demands of people and community?
Hypothesis

Traditional way: Finding useful knowledge from Available resources

Smart way: Search first and check availability If not available find alternative

※ Quantity and quality of library Collection was important.
※ Quality of Searching is the most important now.

Expected results of our analysis

Traditional way: High cites from Traditional(?) journals

Smart way: High cites from Open Access journals
Why do I think is important citation analysis?

**Hypothesis**

<table>
<thead>
<tr>
<th>Searching is to overview current knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access and reading are to understand particular knowledge.</td>
</tr>
<tr>
<td>Citing is to list up the knowledge.</td>
</tr>
<tr>
<td>What I know, and what I want others to know.</td>
</tr>
</tbody>
</table>

**Expected results of our analysis**

<table>
<thead>
<tr>
<th>Diversity of cited journals when search based selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited cited journals when embellishment prevails</td>
</tr>
<tr>
<td>Embellish the research by citing famous articles</td>
</tr>
</tbody>
</table>

Embellish the research by citing famous articles
The present scholarly citation system inadequately exposes the knowledge networks that exist within the scholarly literature, linking papers, authors, funders, research projects and datasets.

- Citation data are hidden behind subscription firewalls of commercial companies.
- Academics are not free to use their own citation data as they please.

In this Open Access age, it is a scandal that reference lists from journal articles, the core elements of the academic data cycle, are not freely available for use by the scholars who created them.

Citation data now need to be recognized as a part of the Commons – those works that are freely and legally available for sharing.
Recommended nomenclature for references and citations

- This is the nomenclature used in our SPAR (Semantic Publishing and Referencing) Ontologies

http://purl.org/spar/

Generic structured metadata required to record a citation

entities

- Citing paper
- Cited paper

relationship

- cito:cites

type

- e.g. Journal article
- bibliographic metadata
- Title
- Publication date
- Bibliographic citation
- Unique identifier
- provenance
- Source of citation info, e.g. CrossRef

Courtesy of David Shotton
Citation database was a crucial subject in our research.

**Local and regional citation data collected from Web of Science and Scopus.**

- Web of Science and Scopus supply global citation statistics.
- Parsing technologies are necessary for journal titles and related metadata.

NAVER academic supplies local citation statistics.

Matching the abbreviated journal titles with full journal names and ISSN numbers. From the ISSN numbers journals could be recognized OA journals, ISI journals, Scopus journals and so on.

**Journal database** to match name of journal with ISSN number is a key. Coverage of matched data among current database is an issue. Interpretation of the meaning of the data is critical part of our research.

→ The number of OA journals is the lowest estimate.
→ The actual proportion(%) of OAJ is higher than the figure here.
Why “Local and Regional” research?

Hypothesis

Why are “Local and Regional” research important?

- (Research ecology) Research environment and demand are different.
- (Research trends) Time related factors are contributing to research outcomes.
- (Diagnosis and discover) Knowing local and regional issues will show knowledge priority and requirements.

Expected results of our analysis

Citation pattern in “Local and Regional” research is different from

- (Practical citation) Authors cite more local articles from local journals
- (Subject area) Different patterns by subject areas
- (Embellishment) Unnecessary references are omitted in articles from local journals.
- (Interests) Interests of nation, societies, institution and research staffs are reflected.
3. Web of Science
How we collected “WoS Data.”

- As a user of SNU library subscribing Web of Science, JCR and NCR
- On-line collection of articles thru searching on address of authors.
- **Number of citing articles**: 448,280 (445,579) in 1991–2011 (Korea)
  - citing articles (SNU only): 52,295 (1998–2011), 6,100 (SNU 2011)
- **Number of cited articles**: 9,864,892 (incl. 410,000 non-journal cited articles) (Korea)

*Years of citing articles are coupled with journal titles of cited articles*
So far only 395 local journals and 9,437 global journals are matched with ISSN among 26,771 titles.

36.7% of journal titles are matched with issn but they covered 88.6% of references.

The numbers of journal titles may not be accurate but percentages of cited counts are close to real value.
WoS Data Summary for cited articles (SNU 2011) excluding unmatched journal names

Proportion by Journal Groups

Journal Title Counts

Cited Article Counts

• After exclusion of numbers of articles with un-matched journal names.
• WoS indexed articles are vast majority of articles cited (90%) and journals contributed (77%).

- **Yearly production of Citing articles**
  - Loc Toll SCI
  - Loc Toll Non-SCI
  - Loc OAJ SCI
  - Loc OAJ Non-SCI
  - Glob Toll SCI
  - Glob Toll NonSCI
  - Glob OAJ SCI
  - Glob OAJ Non-SCI

- **Yearly sum of Cited articles**
  - Glob Toll SCI
  - Glob Toll nonSCI
  - Glob OAJ SCI

*Source: WoS*
2 year citation of each journal vs ISI impact factor (All subject, SNU 2010–2011)

This part is “Magnified” in the following graphs.
2 year citation of each journal vs ISI impact factor (All subject, SNU 2010–2011, Magnified)

- Toll Access Journal
- Open Access Journal

Journals more cited by Koreans

Journals less cited by Koreans

Sum of 2010–2011 citation vs Impact Factor Low to High

25 / Local and regional impact of open access journal articles
2 year citation vs impact factor (Medicine & Health) (SNU 2010–2011, Magnified)

- J Cell Biol: Free access after 6 months
- Diabetes: Free access after 12 months
- Brit J Cancer: Free access after 12 months
2 year citation vs impact factor (BMC & Springer Open, other OA journals on Medicine & Health) (SNU 2010–2011, Magnified)

* Courtesy of BioMedCentral
OA Portion by Subjects: Journals vs Articles (SNU 2011)

- **Medicine Health**
  - Journals: 72%
  - Articles: 89%

- **Engineering**
  - Journals: 90%
  - Articles: 95%

- **Natural Science**
  - Journals: 80%
  - Articles: 91%

- **Agro-Marine**
  - Journals: 82%
  - Articles: 94%
OA Portion by Subjects: Journals vs Articles (SNU 2011)

- **Humanities**
  - Journals: 83%
  - Articles: 98%

- **Social Science**
  - Journals: 61%
  - Articles: 67%

Legend:
- Loc Toll Non-SCI
- Loc Toll SCI
- Loc OAJ Non-SCI
- Loc OAJ SCI
- Glob Toll Non-SCI
- Glob Toll SCI
- Glob OAJ Non-SCI
- Glob OAJ SCI
1. OA journals are less cited than Toll access journals
   Medicine & Health 6.0%  Engineering 3.3%
   Natural Sciences 3.4%  Agro-Marine 4.6%
   Humanities 0.3%  Social Sciences 4.1%

2. More than 90% of cited articles are from WoS indexed journals except for Social science (58%).

3. Articles from local journals are rarely cited.
4. NAVER

http://academic.naver.com
Brief Introduction of NAVER Academic Service

The Gateway to Research Information in Korea

1. Aggregation
   - Aggregation of academic data from around the world

2. Connect to Libraries
   - Connect to college libraries through Linking Resolver
   - Provided by the libraries for academic search API

3. Citation Index
   - Citation index DB to provide in-depth academic information for researchers (Non-commercial)

- Synonym DB for 'Authority Control'
  - About 110,000 journals, 2.5M Synonym
  - About 20,000 academic societies, 0.2M Synonym
  - About 10,000 Korean government offices, 0.1M Synonym
Monthly Page views on NAVER Academic and Google Scholar by Korean users

Monthly counts of page views (x1,000)

Source: Nielson Korean Click
Proportion of page views by age groups on NAVER academic vs Google Scholar

Source: Nielson Korean Click, Apr 2014
Korean Citation Index Journals: 1,897

Citation database coverage:
- 90,000 citing, 2.1 million cited articles from Humanities
- 140,000 citing, 4.8 million cited articles from Social Sciences
- 70,000 citing, 1.65 million cited articles from Natural Sciences
- 90,000 citing, 1.4 million cited articles from Engineering
- 60,000 citing, 0.65 million cited articles from Agro-Marine
- 60,000 citing, 1.4 million cited articles from Medicine & Health

510,000 citing, 12 million cited articles all together.
Distribution of journal titles of cited articles in NAVER Academic is different from Web of Science

Cited journal titles in NAVER Academic DB

- Glob Toll SCI, 3878, 33%
- Glob Toll NonSCI, 3164, 27%
- Loc Toll NonSCI, 3017, 25%
- Loc Toll SCI, 76, 1%
- Others, 874, 7%
- Loc OAJ NonSCI, 606, 5%
- Loc OAJ SCI, 23, 0%

Cited journal titles in WoS DB

- Glob Toll SCI, 7563, 28%
- Loc OAJ SCI, 232, 1%
- Loc Toll SCI, 85, 0%
- Loc OAJ NonSCI, 3, 0%
- Others, 16939, 63%
- Glob OAJ SCI, 653, 3%
- Glob OAJ NonSCI, 17, ...
- Glob Toll NonSCI, 1204, 5%

(Caution) Numbers of journal titles in each group will increase as more names are matched.
Distribution of cited articles in NAVER Academic is different from Web of Science

Cited articles in NAVER database

- Global Toll NonSCI 8%
- Local Toll NonSCI 44%
- Local OAJ NonSCI 13%
- Local OAJ SCI 0%

Cited articles in WoS database

- Global Toll NonSCI 8%
- Local Toll NonSCI 1%
- Local OAJ NonSCI 13%
- Local OAJ SCI 0%
OA Portion of Articles by Subjects in NAVER Academic

**Medicine Health**

**Engineering**
OA Portion of Articles by Subjects in NAVER Academic

Natural Science

Argo-Marine
Annual trends of cited counts (proportion / counts) of Medicine and Health journals

- Citing local journals as many as Glob SCI
- Open access journals are growing.
- Local nonSCI OA journals are often cited.
- Recent citations more from local journals.
Annual trends of cited counts (proportion / counts) of Natural Sciences journals

- 60% cited from Glob SCI in natural sciences
- Open access journals are growing.
- Recent local SCI / nonSCI journals are cited.
NAVER Academic citation data shows:

1. Local journal titles of cited articles in NAVER Academic is 57% compared to 1–2% in Web of Science.

2. Recent articles are more used from local journals but global journals have longer cited year span.

3. Local OA journals are more used particularly for medicine and health.

4. Humanities are the most local, natural science is the most global.
## Citing articles

<table>
<thead>
<tr>
<th>Type</th>
<th>NAVER (%)</th>
<th>WoS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Local</td>
<td>OAJ</td>
</tr>
<tr>
<td>Medicine &amp; Health</td>
<td>62</td>
<td>22</td>
</tr>
<tr>
<td>Engineering</td>
<td>80</td>
<td>41</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>39</td>
<td>16</td>
</tr>
<tr>
<td>Agro-Marine</td>
<td>88</td>
<td>38</td>
</tr>
<tr>
<td>Humanities</td>
<td>89</td>
<td>4</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>44</td>
<td>6</td>
</tr>
</tbody>
</table>
5. Interpretation & Implications
Open Access is not as much used as we expected.

Different Referencing by Korean Authors

<table>
<thead>
<tr>
<th></th>
<th>When they submit to “Global Journals”</th>
<th>When they submit to “Local Journals”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing OA Journals %</td>
<td>0.3~6%</td>
<td>4~41%</td>
</tr>
<tr>
<td>Citing Local Journals %</td>
<td>0.5~1.8%</td>
<td>32~89%</td>
</tr>
</tbody>
</table>
Diversity of titles of journals of cited articles

How many journal titles are cited?

Web of Science: 8523 SCI titles used. 728 (8.5%) open access SCI titles.

NAVER Academic: 4048 SCI titles used, 104 (2.6%) open access SCI titles.

Local journals are widely used in NAVER Academic database.

Total number of used journal titles: estimated 9,832–26,771 (technology issue).

Diversity of cited journal is more demanded

• Traditional way of citation dominates over internet way of search & selection.
• Embellishment of articles by citing famous articles.
• Reputation Effect: Inbreeding citation? Eliticism?
Local and Regional Characteristics of Open Access journals and articles

Local & Regional citation databases indicate – – –

- Local / regional Research Culture & Custom
- Local / regional Knowledge Demand & Trends
- Local / regional Publishing Market Opportunities
- Local / regional Evidences & Critiques on Policies on Research
  ➞ Cross publisher / cross county metric system on citation
  ➞ Regional Citation Database Network

Local & Regional strategy for Open Access

- More people to read OA,
- More scientists to write OA,
- More authors to cite OA.

Encourage Reading, Writing & Citing
End of Slides