

Researcher Profiling for Researcher Analysis Service

Mikyoung Lee¹, Minhee Cho¹, Changhoo Jeong¹, Hanmin Jung¹

¹ Korea Institute of Science and Technology Information (KISTI),
245 Daehak-ro, Yuseong-gu, Daejeon, South Korea
{jerryis, mini, chjeong, jhm}@kisti.re.kr

Abstract. This study examines a method of generating comprehensive profiling information for a researcher analysis service. In addition to basic and performance-based information about researchers necessary to generate profiling information, we introduce researcher performance index models for researcher analysis service. The models can that measure qualitative and quantitative performance, researcher influence, and growth potential, which are necessary to analyze the skills of a researcher from multiple perspectives. We measure the qualitative performance index of researchers by using the citation index of the papers they have published. The quantitative performance index can be measured based on a researcher's published papers. The Influence index measures the social influence of researchers according to their academic work. The growth potential index measures the speed at which a researcher improves research performance. We expect to develop a researcher analysis service that can evaluate a researcher's performance and enhance his or her research abilities.

Keywords: Researcher Profiling, Research Performance analysis, Researcher Analysis service

1 Introduction

Many services exist that provide researchers with a wealth of academic information such as scholarly literature, including Google Scholar, MS Academic Search and Elsevier SciVal. However, services are lacking that help improve the research abilities of researchers by analyzing their research skills. In Informetrics, various studies have represented researcher research skills in the form of objective numbers by examining various bibliometric indicators for research policy [1]. Currently, the h-index, which uses a citation index, is widely used to measure researcher skills. However, the h-index fails to produce objective assessments of the abilities of individual researchers. This is because it reflects only the combined achievements of co-researchers and is limited in the manner in which it comparatively analyzes them with the achievements of researchers in other fields. In addition, the h-index does not reflect the assessments of competent new researchers. Numerous studies have been conducted to compensate for the existence of a index, which is insufficient for analyzing the comprehensive research skills of researchers.

¹ Corresponding author

This study describes researcher profiling information used in a researcher analysis service for improving research skills.

2 Related Work

InSciTe Advisory(<http://inscite-advisory.kisti.re.kr/search>) deals with various textual big data including papers, patents, Web, social data, and linked data. It analyzes researcher's competitiveness and recommends attainable strategy and plan based on prescriptive analytics as well as descriptive analytics. It consists of two main analytics: descriptive analytics and prescriptive analytics. The former includes both activity history analysis and research power analysis for a selected researcher [2][3].

- **Commerciality:** Ability to produce practical products and profits
- **Scholarity:** Ability to produce new knowledge and academic outputs
- **Influentiality:** Ability to spread leverage to other researchers
- **Diversity:** Ability to extend research scope and degree of variation of research field
- **Durability:** Ability to keep research consistently in some research field
- **Technology emergability:** Degree of emerging about research area performed by researcher
- **Partner Trend:** Status of research area changing about partner researchers
- **Market Trend:** Status of market size changing about researcher area

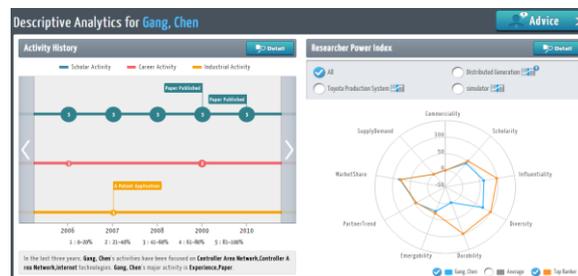


Fig. 1. Descriptive analytics of InSciTe Advisory

ArnetMiner(<http://www.arnetminer.org>) aims to provide comprehensive search and mining services for researcher social networks. In this system, we focus on: creating a semantic-based profile for each researcher by extracting information from the distributed Web; integrating academic data from multiple sources; accurately searching the heterogeneous network; analyzing and discovering interesting patterns from the built researcher social network [4].

ArnetMiner offers insights into the capabilities of researchers by using the seven indices of activity, papers, citation, h-index, g-index, sociability, and diversity. As Fig. 2 shows, users can judge a researcher's ability by examining the graph related to each index.

- **Citation:** The number of citations of all publications by an expert.
- **Papers:** The number of all publications by an expert.

- **H-index:** An expert has index h if h of his or her N papers have at least h citations each, and the other papers have at most h citations each.
- **Activity:** People's activity is simply defined based on one's papers published in the last years.
- **Diversity:** An expert's research may include several different research fields. Diversity is defined to quantitatively reflect the degree
- **Sociability:** The score of sociability is basically defined based on how many coauthors an expert has.

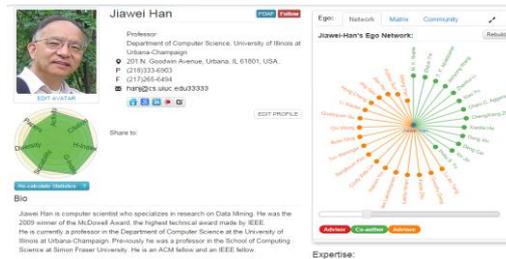


Fig. 2. Researcher Information on ArnetMiner

3 Researcher Performance Index Model

This chapter discusses the researcher performance index model used for researcher profiling. This index model collects the bibliography and citation information of researchers and compares them numerically to measure their research skills. We define four indices based on qualitative and quantitative performance, influence, and growth potential.

3.1 Qualitative Performance Index

We measure the qualitative performance index of researchers by using the citation index of the papers they have published. The qualitative performance index is a measurement of the influence of published papers and, thus, an indicator of their research quality. In general, indices such as citation index, h-index [5] and g-index [6] are used to measure researcher performance. A citation index can be used because the h-index and g-index for all authors in our collection are unavailable.

3.2 Quantitative Performance Index

The quantitative performance index can be measured based on a researcher's published papers. The more active the researcher has been producing recent studies, the more weight is given to the annual performance index.

The qualitative performance index is not proportional to the quantitative performance index because the more papers that are published, the higher is the

researcher's score on the quantitative index, thus indicating active research. However, the higher the quality of the papers, the higher is the researcher's quantitative performance index because the rank of the journals and conferences are reflected in the calculation of the quantitative performance index. If researchers being considered are the main or corresponding authors, they are more likely to score high on the quantitative index when they are actively involved in research. Using (1), the quantitative performance index can be calculated based on the rank of the journals and conferences, the weight of the author order.

$$QP = \sum_1^n Journal_{rank} \times author_{order} + \sum_1^m Proceeding_{rank} \times author_{order} \quad (1)$$

The rank of journals can be calculated by the journal's impact factor. The journals from each field are ranked according to the impact factor. Regarding conference proceedings, the rankings of the conferences in each field are used. With respect to authors, main and corresponding authors who write the majority of a paper can earn one point, whereas co-authors receive points equal to $1/n$ th, where n is the total number of authors.

3.3 Influence Index

Using a co-author network, the influence index measures the social influence of researchers according to their academic work. The influence index is an application of Google's PageRank [7] based on the co-author network and assumes that co-authors of influential research also have high influence.

$$II(A) = (1 - d)/n + d(\sum_1^n II(Tn)/C(Tn)) \quad (2)$$

In (2), $II(A)$ refers to the value of researcher A's influence, $II(TI)$ refers to the value of researcher TI 's influence, and $C(TI)$ represents the number of co-authors with researcher TI . In addition, d equals 0.85, as in the PageRank algorithm.

3.4 Growth Potential Index

The growth potential index measures the speed at which a researcher improves his or her research performance. We measure the growth potential index based on the amount of time a researcher spends reaching his or her current research performance, thus enabling to the system to predict how quickly the researcher can become a leader.

$$GP = QP/n \quad (3)$$

QP represents the quantitative performance index and n denotes the elapsed time from the beginning of the study to the present.

4 Researcher Profiling

4.1 Design

During the development of our researcher analysis service, we define research profiling, which includes all analytical information about the researcher. Researcher profiling consists of a list of basic information about researchers such as names and affiliations. In addition, it identifies essential information derived from their research such as main research skills (as shown in Fig. 3), research area (Fig. 3), and the co-author network. Moreover, it includes indices provided by research performance models.

Fig. 3 shows an example of researcher profiling designed in this study. The profiling extracts basic information from bibliographies and citations such as names, affiliations, beginning years of research, paper types, author order, number of citations, and major technology used. Information such as the level of technology and co-author network is obtained from analysis models. All acquired information is used to calculate the four indices of researcher performance index models. Most profiling preferences include time information, which facilitates the observation.

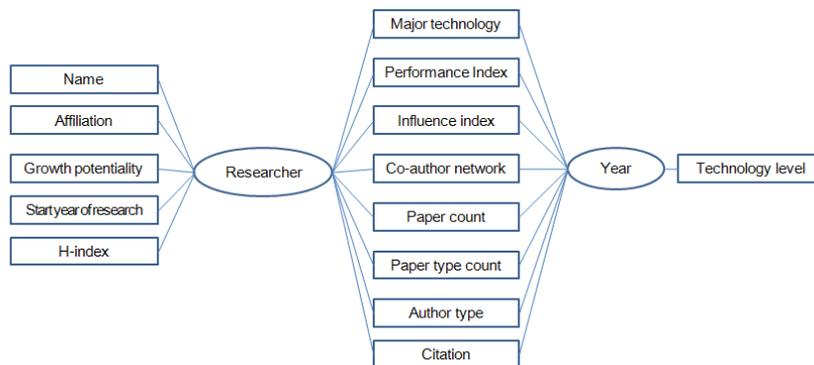


Fig. 3. Design of Researcher Profiling

4.2 Applications

Multiple aspects of research performance can be evaluated using the four indices of qualitative and quantitative performance, influence, and growth potential provided by research performance index models. Each index can be used to evaluate researcher tendencies and characteristics. For example, researchers can be divided into various types such as competent researchers who produce high quality research results, those who produce highly quantitative outcomes, and those who have immense research potential. In addition, yearly trends related to researcher profiling information can confirm changes in a researcher's performance based on flow patterns of falling, rising, and stagnancy. Moreover, the service can also be used for prescriptive analytics to reinforce researchers' strengths, correct research weaknesses, and

compare researchers of similar research capabilities and patterns. Finally, it can be used to analyze research performance patterns and to determine researcher's type.

5 Conclusion

This study described a researcher profiling method necessary to develop a researcher analysis service. Research profiling information consists of meta-information, performance information, and information of researchers acquired by analytical model. We evaluated researchers from multiple perspectives. We used a qualitative performance index, which indicates the quality of papers; a quantitative performance index, which indicates the number of papers published; an influence index, which refers to the amount of influence a researcher has among his or her peer researchers; and a growth potential performance index, which indicates degrees of change (i.e., levels of improvement) in research performance over time.

Our researcher profiling service represents an aggregation of all information that can be used to evaluate researcher and their research characteristics and patterns. This service is necessary to develop a service to assist researchers in improving their performance.

Acknowledgements

This work was supported by the IT R&D program of MSIP/KEIT. [2014-044-024-002, Developing On-line Open Platform to Provide Local-business Strategy Analysis and User-targeting Visual Advertisement Materials for Micro-enterprise Managers].

References

1. Rodrigo, C., Maria, B.: The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. In: *Journal of Informetrics*, vol 1 issue 3, pp193-203, Elsevier (2007)
2. S. Song, D. Jeong, J. Kim, M. Hwang, J. Gim, H. Jung, :Research Advising System based on Prescriptive Analytics, In Proc. of International Workshop on Data-Intensive Knowledge and Intelligence in conjunction with the 9th FTRA International Conference on Future Information Technology. (2014)
3. Jinyung Kim, Myungwon Hwang, Janwon Gim, Sa-Kwang Song, Do-Heon Jeong, Seungwoo Lee, Hanmin Jung.: Intelligent Analysis and Prediction Model of Researcher's Capability for Prescriptive Analytics. In: AIM 2014 (2014)
4. Jie, T., Jing, Z., Limin, Y., Juanzi, L., Li, Z., Zhong, S.: ArnetMiner: extraction and mining of academic social networks. The 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 990-998, ACM, New York, NY, USA (2008)
5. H-index, <http://en.wikipedia.org/wiki/H-index>
6. G-index, <http://en.wikipedia.org/wiki/G-index>
7. Page, L., Brin, S., Motwani, R., Winograd, T.: The pageRank Citation Ranking: Bringing Order to the Web, Technical Report, Stanford InfoLab. (1999)