CrossMark

ORIGINAL ARTICLE

Systematic Methodology for Excavating Sleeping Beauty Publications and Their Princes from Medical and Biological Engineering Studies

Tien-Chi Huang[1] · Chieh Hsu[1] · Zih-Jin Ciou[2]

Abstract From 2010 to 2014, the Journal of Medical and Biological Engineering (JMBE) published more than twice as many articles as it did during the previous 5-year period. This increase has raised significant information management and retrieval issues related to the emergence of "big data". Generally speaking, every publication has academic value and should be available to researchers. However, statistical assessments indicate that 95 % of studies are not cited by other publications. Some of these dormant articles could make significant contributions to medical and biological engineering research. Such studies are referred to as "sleeping beauties". The present study develops an information "awakening" framework to explain how to extract value from sleeping beauties. The proposed framework implements a web-based system through which researchers can propose effective awakening strategies, referred to as "princes". An analysis of experimental results found 14 sleeping beauties in JMBE, some of which were awakened by the two proposed princes, namely "keyword prince" and "H-index prince". The results are expected to provide journals with a clear way to promote published studies, and to make their findings more readily available to scholars.

Keywords Big data · Information awakening · H-index · Medical and biological engineering studies

✉ Tien-Chi Huang
  tchuang@gm.nutc.edu.tw; tchuang@nutc.edu.tw

1 Department of Information Management, National Taichung University of Science and Technology, Taichung 404, Taiwan, ROC

2 Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taichung 404, Taiwan, ROC

1 Introduction

There is a dramatic increase in the volume of scientific papers published recently year [1]. For example, the academic database for the Journal of Medical and Biological Engineering (JMBE) features a total of 662 articles published in the past 15 years, including 131 studies published from 2000 to 2004, 171 published from 2005 to 2009, and 345 studies published from 2010 to 2014 (Fig. 1). The rate of publication in JMBE has thus more than doubled in the two consecutive 5-year periods from 2000 to 2009. This rate of increase is also found in other academic journals. However, 95 % of JMBE articles during this period failed to be cited in other respected journals. Such articles are referred to as "sleeping beauties", since they can potentially provide value to researchers if they were "awakened".
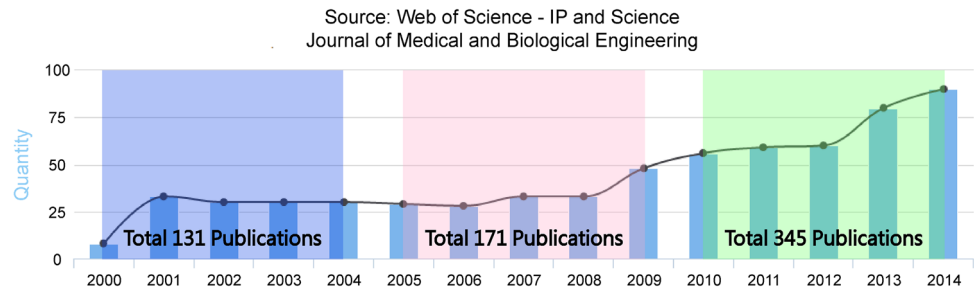
"Sleeping" information may go unnoticed during a long sleeping state, but then suddenly come to attention because of certain triggers. Such publications are referred to as "premature discoveries" [2], "resisted discoveries" [3], "delayed recognition" [4], and sleeping beauties [5]. The classic example of this citation phenomenon is called Mendel's syndrome, named after Gregor Mendel [6], whose discovery in plant genetics was so unprecedented that it took 34 years for the scientific community to catch up with it [5]. Wang et al. [7] noted that sleeping beauties is very common in science, and only a small fraction of sleeping beauties are eventually awakened [8]. A methodology for identifying sleeping beauties in the field of medical and biological engineering is the aim of the present study. A sleeping beauty must be awakened by a "prince", a study that brings the dormant study to wide attention; however, little research has focused on identifying such princes.

🍀 Springer

Fig. 1 Journal of Medical and Biological Engineering publications from 2000 to 2014

Source: Web of Science - IP and Science
Journal of Medical and Biological Engineering

Total 131 Publications   Total 171 Publications   Total 345 Publications

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

This research examines the efficacy of various types of princes in identifying sleeping beauties in JMBE. Princes and sleeping beauties were matched through a web-based information system based on Hirsch index (H-index) analysis techniques, which are widely used to identify the academic prominence of authors, studies, and journals. This study seeks to devise a systematic methodology for identifying potential sleeping publications in a journal, and to match them with princes that can effectively awaken them, thus making them available to other researchers and readers.

2 Literature Review

2.1 Big Data

Global digital data volumes are now calculated in terms of zettabytes [equivalent to one billion terabytes (TB)] [9], with Google alone processing at least 24 PB (petabytes, equivalent to 1024 TB) of data every day [10]. In the past, such data could be only analyzed by sampling [11], but today's technology can quickly analyze these data and extract useful information.

Laney identified three characteristics of big data, namely volume, velocity, and variety [12, 13]. Volume refers to the size of the data, with "big" defined differently by various industries and for different data types [14]. Velocity refers to the rate of data growth, currently driven by the increasing popularity of intelligent electronic devices and increasing network bandwidth [15, 16]. This study focuses on analyzing data in academic journals that publish massive amounts of content every year. For example, JMBE has more than doubled the number of articles published over the past decade. Academic publishing now faces big data issues similar to those experienced by other data-intensive industries. Variety refers to data structure types: structured, semi-structured, and unstructured. Structured data refers to electronic forms or data stored in associated databases; semi-structured data are texts with identifiable rules, such as extensible markup language (XML). Unstructured data refers to articles, pictures, audio files,

video files, and other files without a fixed structure [15, 17]. This study focuses on the collection and analysis of citation counts, authors, H-index of authors, and keywords in JMBE, which are a kind of structured data [18].

In the 19th century, a US naval officer named Maury analyzed large volumes of ships' log data, including the temperature, velocity, and direction of winds and currents to plan more efficient trans-Atlantic routes [19–21]. Today, similar approaches are used to analyze large amounts of data, for example from medical records to help doctors determine optimal medical interventions [22]. Similarly, the global retail giant Wal-Mart analyzed sales records and discovered that hurricane warnings prompt consumers to not only purchase flashlights but also strawberry-filled cakes and beer [23], allowing the company to optimize product displays and purchasing to maximize sales revenue. Baumgartner analyzed more than 7000 articles from the Journal of Marketing to identify certain types of articles that were particularly influential in analyzing consumer behavior [24]. Radicchi and Castellano analyzed over 30,000 articles to assess the relationship between number of articles, citation frequency, and H-index [25]. Van Raan analyzed approximately 20,000,000 articles from 1980 to 2004 to identify sleeping beauties, which he defined as articles that fail to receive attention at the time of publication, but receive recognition later on [5]. However, Van Raan failed to identify the cause of this phenomenon. The present study applies data analysis techniques to identify sleeping beauties in JMBE and seeks to identify princes to awake the sleeping articles and bring them to the attention of the research community.

2.2 Information Awakening

Every published article has a certain degree of academic value, but timing is a critical issue in finding a receptive audience for scientific findings. Some "scientifically premature" studies are published before the research community is ready to accept or understand the findings [26]. These studies cannot be verified with current knowledge and technology and thus their academic value remains hidden as "premature discoveries" [2] or "resisted

discoveries" [3]. Such findings are away further advances in technology and general understanding, and thus suffer from "delayed recognition" [4]. In 2004, Van Raan referred to articles that did not receive attention at time of publication as sleeping beauties that could be awakened by a prince who would bring the article to general attention. Van Raan defined multiple sleeping states, including "deep sleep" and "less deep sleep", and further explored their relationship with "awakening intensity". In one classic case, an article published by Roman in 1986 [27] remained unnoticed until 1995 when cited by Polchinski [28]. Roman was subsequently cited numerous times, and thus Polchinski played the role of a prince who awakened Roman's sleeping beauty [5]. This could be regarded as one of cases of "Mendel's syndrome" [29].

Information awakening has received increasing attention in many academic fields [7]. The journal Atmospheric Environment conducted a SCOPUS literature analysis to indirectly discover sleeping beauties [30]. The Journal of Consumer Psychology analyzed over 7000 articles in the Journal of Marketing, Journal of Marketing Research, and Journal of Consumer Research to examine this phenomenon in the consumer research domain [31]. Journals focused on research in the field of data analysis have sought to identify common characteristics of sleeping beauties [8, 32] and to predict the appearance of corresponding princes [33, 34].

A recent study explored the characteristics of information awakening [29], specifically focusing on the factors contributing to the awakening of a sleeping beauty. This study hence develops an information awakening framework that can be applied to any academic database. Moreover, a prince is defined from two different viewpoints, namely a macroscopic sense and a narrow sense. The effectiveness of the proposed framework is verified experimentally, with the expectation that the principles can be applied to find and awaken sleeping publications in other journals.

2.3 H-index

Academic databases contain huge volumes of research reports, and scholars are devoting increasing attention to the problem of maximizing the impact of individual articles given the recent sustained increase in publication output [35]. Garfield's impact factor is widely used to evaluate journal influence [36, 37], but this indicator cannot evaluate the academic impact and contribution of individual scholars. The H-index has been extensively verified as a useful indicator of the impact of individual scholars in a given academic field [37]. The H-index is defined as follows:

A scientist has an index h if h of his or her Np papers have at least h citations each and the other (Np-h) papers have no more than h citations each.

The invention of the H-index has led to increased research focus on popular topics [38–42]. As the volume of research output has increased, the H-index has been widely applied to assess different impact indicators in research groups [43, 44], journals [45, 46], and research topics [47]. The indicator is also used by many influential academic search platforms (such as Web of Science, Scopus, and Google Scholar) [48–50].

Despite significant prior research into sleeping beauties and the application of the H-index [51], few studies have focused on individuals and actions taken to awaken sleeping publications. The present study analyzes sleeping publications in a specific publication database and uses the H-index as a key indicator to identify those who are successful in awakening sleeping publications.

2.4 Data Collection

This study analyzes all articles published in JMBE from 2000 to 2014 for article title, author, year of publication, keywords, and number of citations.

Citation count data were sourced through queries of Thomson Reuter's Web of Science (WoS) in January, 2015. JMBE began publication in 2000, and has published a total of 647 articles over the subsequent 15 years, featuring 1032 citations and 2406 keywords. However, WoS queries can only access JMBE articles published after 2008, when JMBE was first included in the Science Citation Index (SCI). After 2008, JMBE published a total of 426 articles with 971 citations and 1727 keywords.

This study seeks to design a set of methodologies for the analysis of academic databases that meet the characteristics of big data proposed by Laney in 2001 [12]. In terms of volume, the output of academic journals has increased steadily [52]. In terms of velocity, the output of JMBE has accelerated rapidly in the past 5 years [53]. Finally, in terms of variety, keyword analysis shows that JMBE includes articles in various fields.

3 Methodology

3.1 Information Awakening Framework

Although many previous studies have examined the "sleeping beauty" phenomenon, they mostly focused on theoretical aspects and data analysis. This research devises a methodological framework, named the information awakening framework (IAF), to discover sleeping beauties,

and introduces the use of big data analysis techniques to improve the accessibility of such articles. This study adopts the big data management and analytics processes proposed by Gandomi and Haider [54]. Data management fills the Processor role; Analytics is classified into two roles: Analyzer and Decision-maker. The Analyzer role performs overall data modeling and analysis, and the Decision-maker role conducts data interpretation combined with data visualization [52, 54]. According to Hashem, structured big data should be preprocessed before analysis [55], after which the data can be used for observation and decision-making [56].

Figure 2 depicts the proposed roles in IAF and their respective working items and processes. Figure 3 illustrates the three roles used in the process of information awakening, and the potential contribution of awakening to decision-makers.

Three processing roles are used: Data processor, Analyzer, and Decision-maker. To identify incidents of information awakening, the data processor first uses the proposed information awakening system to filter large amounts of raw data from academic databases based on year, subject, author, keywords, citation count, H-index of author, cited article, year of citation, subject of cited article, author of cited article, and H-index of author of cited article.

Next, the analyzer determines whether the filtered data is suitable for subsequent analysis. The data processor searches other academic databases and supplements missing information. In the subsequent analysis stage, the "sleeping" concept proposed by Van Raan [5] is used to find sleeping articles, which are either in "deep sleep" or "less deep sleep". By analyzing the year in which the sleeping papers are cited by other articles, it can be determined whether there are awakened sleeping papers (i.e., sleeping beauties). An analysis of citations will also reveal the identity of the specific prince or princes responsible for the initial awakening. Finally, the system visualizes the analytical results to help decision-makers (e.g., scholars or journal proprietors) clearly understand development trends for particular journals. In addition,

during the data analysis process, Decision-maker can update the analysis results to the academic database, and provide a visualization of the analysis results to allow scholars to easily grasp research trends and self-contribution via popular keywords and H-index (shown as a dotted line in Fig. 3). Moreover, these outcomes provide journal operators with a clear understanding of developmental trends in various research fields, and may help to match sleeping articles with potential princes.

3.2 Princes

An analysis of a variety of articles in deep sleep and less deep sleep in JMBE revealed two types of prince, namely a "keyword prince" and an "H-index prince".

The keyword prince is formed by keywords derived in response to research trends that characterize an entire era. For example, a study about osteogenesis had been sleeping since being published in 2008. In 2012, osteogenesis emerged as a trending topic, gathering attention in other studies, attention which soon awakened the sleeping article, which began to attract citations. Thus, articles in an area of research for which key technologies or techniques have yet to be developed may sleep until such prerequisites are realized, at which point a keyword prince will awaken previously published research that received little attention at the time [29].

The H-index prince mainly uses the H-index of the primary author as the evaluation indicator because the primary author is usually the one with highest contribution and impact [57]. Authors with a high H-index are generally more influential, and thus by citing a sleeping article, such authors can attract new attention to previously overlooked research.

3.3 System Architecture

To verify the proposed IAF, this study developed a web-based information awakening system (http://163.17.136.212/IA/JMBE). The system is implemented on the Apache web server in PHP, using web front-end technology jQuery and the MySQL database. Data are drawn from two academic

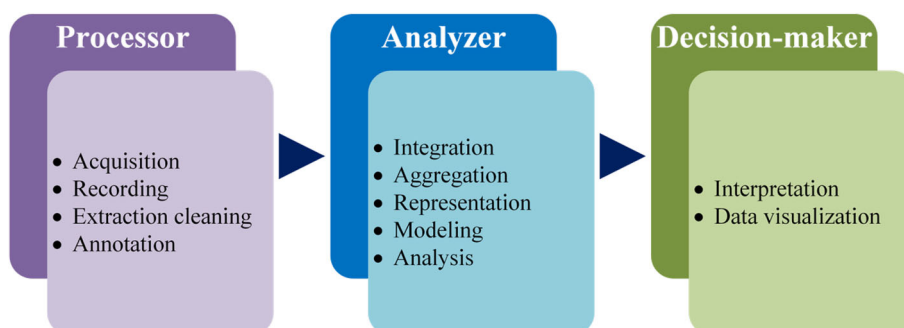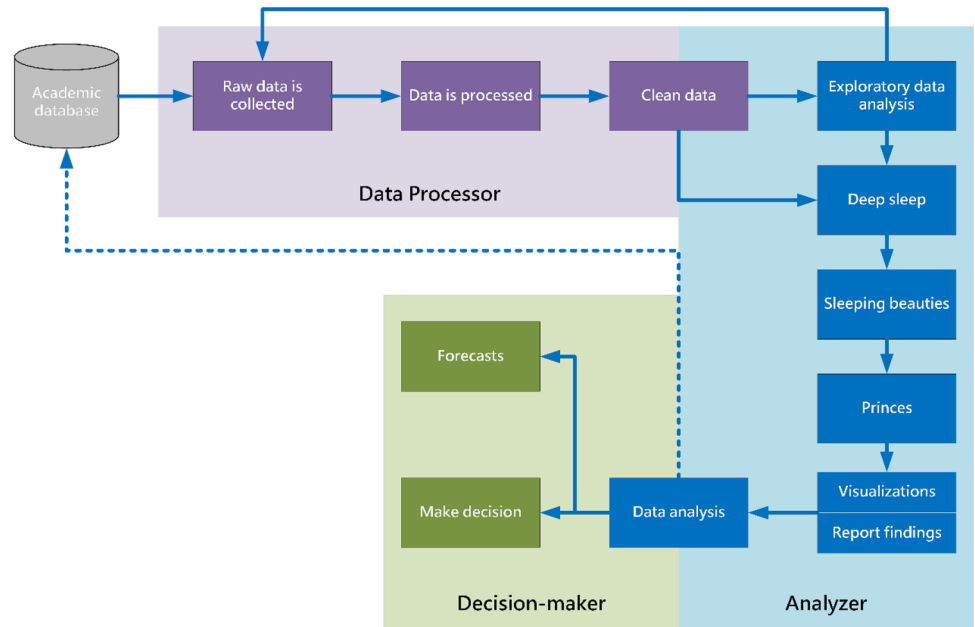**Fig. 2** Three proposed roles and their working items

**Fig. 3** Proposed information awakening framework

databases, namely JMBE and WoS. As shown in Fig. 4, the system integrates individual modules corresponding to the different roles: Crawler, Data Clean, Data Analysis, and Operation. The system stores JMBE publications and the data analysis results, which are presented as a web page and can be customized according to user specifications. Detailed module processes are depicted in Fig. 5.

3.3.1 Phase 1: Exploratory Data Analysis

The Crawler module is executed automatically on the 1st of every month to collect data from the JMBE database

and to send out HTTP requests. Regular expressions are then used to analyze data required for website source code. Crawler collects article titles, years, and authors and forwards them to the Data Clean module to remove undesired text. The cleaned data is then sent to the Data Analysis module, which conducts a statistical analysis to determine the number of authors and year of publication. The results are stored in the information awakening database. Once data processing is complete, the Crawler module is reactivated to collect article-relevant data (e.g., keywords, cited articles, author of cited articles, and number of citations).

Fig. 4 Information awakening system architecture
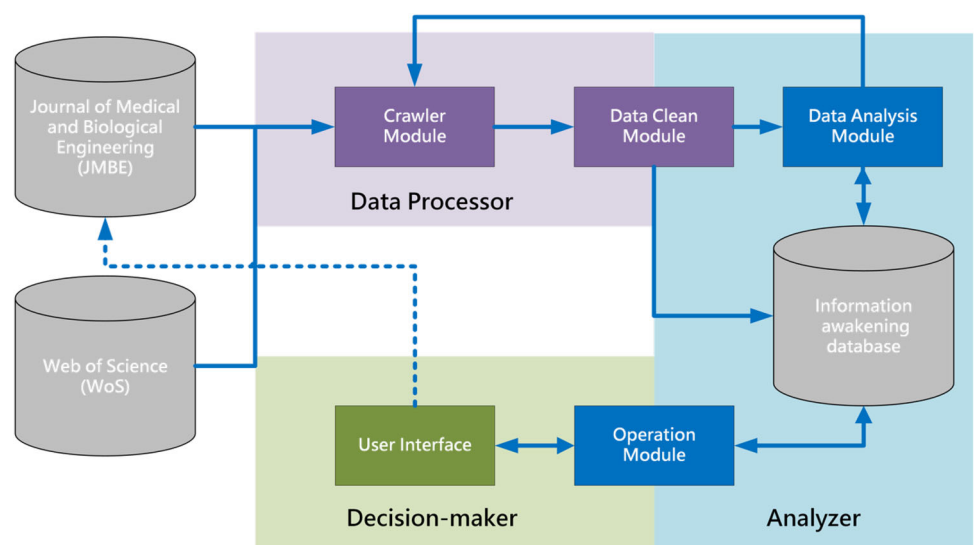
Phase 1 Exploratory Data Analysis

```
Start
go to Search Journal of Medical and Biological Engineering web page;//Crawler Module operations
While(Cannot find page of journal data){
    Input Crawler_Condition;
    If(Crawler_Condition ∈ Page){
        Crawler Page data(title, author, year);
        Store the Page data into Database;//Data Clean Module operations
    }
}
Count published articles in each year;//Data Analysis Module operations
End
```
Data processor

Phase 2 Sleeping Beauties Analysis

```
Start
go to Web of Science//Crawler Module operations
Search title data stored in the Database;
While(load title of Database data Until to end){
    Input Crawler_Condition;
    If(Crawler_Condition ∈ Page){
        Crawler Page data(keywords, citation);
        Processes segmentation on collected keywords//Data Clean Module operations
        Update the Page data into Database;
        If(citation > 0){//Crawler Module operations
            Crawler cite Page data(title, author, keywords, year);
            Processes segmentation on collected first author;//Data Clean Module operations
            Store the cite Page data into Database;
        }
    }
}
```
Data processor

```
Analysis all JMBE data
If(every year citation > 2){//Data Analysis Module operations
    Those data are Awake.
    Update the Awake data in the Database;
}else{
    If(every year citation <= 2 or every year citation > 1){
        Those data are light sleep.
        Update the light sleep data into Database;
    }else{
        Those data are deep sleep.
        Update the deep sleep data into Database;
    }
}
End
```
Analyzer

Phase 3 Prince Analysis

```
Start
go to Web of Science//Crawler Module operations
While(load cite Page's author of Database data Until to end){
    Search cite Page's first author;
    Crawler Page data(H-index);
    Update the H-index data into Database;//Data Clean Module operations
}
```
Data processor

```
Analysis sleeping beauties
countif(keywords); //Data Analysis Module operations
If(Awaken data had been slept){
    Those data are sleeping beauties.
    If(The target sleeping beauty was awaken by a specific"keyword prince"){
        Update keyword prince into Database;
    }else If(The target sleeping beauty was awaken by a specific "H-index prince"){
        Update H-index prince into Database;
    }else{
        Those data was awaken by Unknown.
    }
}
End
```
Analyzer

Phase 4 Information awakening and decision making

```
start
Using a jQuery plugin to transform following data to Figures or Tables on the web page: //Operation Module operations
1. Published articles in each year
2. Article details table
3. Hot keywords table
4. Sleeping beauties details table
End
```
Decision-marker

Fig. 5 Pseudo code of four analysis phases

3.3.2 Phase 2: Sleeping Beauty Analysis

We design the Crawler module to extract information from WoS database. The WoS database provides a complete collection of academic resources with high reference value and relevant data, including number of citations, the cited article, and cited authors. Therefore, this system uses the WoS database as a supplemental database for the analysis

of JMBE data. Crawler also adopts the Phase 1 parsing method and sends parsed data to the Data Clean module for processing. The Data Clean module first segments the collected keywords, and then uses a composite key to store the data in the information awakening database along with other article-related data. The primary author and other authors are extracted and stored in different database columns. Following a second processing using the Data Clean module, the Data Analysis module analyzes the data in the information awakening database to identify potential sleeping beauties. In addition, to analyze articles that are in the deep sleep or less deep sleep state, it also gathers JMBE articles that fit the sleeping beauty criteria and stores all analytical results to the information awakening database. Since the data collected so far is insufficient for the next-stage Prince Analysis, it cannot be used to identify the H-index prince defined in this study. Therefore, the Crawler module is used to obtain the H-index of the primary author of the cited article.

3.3.3 Phase 3: Prince Analysis

In the last run of the Crawler module, the data source is still the WoS database but Crawler retrieves articles that cite sleeping beauties along with the H-index of the primary author of these articles. The retrieved H-index values are stored in a specific database column via the Data Clean module. When the H-index data collection procedure is completed, the Data Analysis module's Prince Analysis conducts two different analyses on the discovered sleeping beauties.

First, article keywords are sorted by year to determine the total number of sleeping beauties awakened by popular keywords each year. This verifies that the sleeping beauties are awakened by "keyword princes". Next, the H-index of primary authors who cite sleeping beauty articles is analyzed to determine how many sleeping beauties are awakened by each H-index prince. Finally all analysis results are stored in the information awakening database for usage by Decision-maker.

3.3.4 Phase 4: Information Awakening and Decision-Making

Decision-maker queries the database via the Operation module; and query results are transformed into tables and drawings through a jQuery-based dynamic chart program. These clear representations of research trends allow Decision-maker to easily determine the relevance and value of specific articles. Useful data can be imported to Decision-maker's own academic database for future use.

4 Results and Discussion

4.1 Sleeping Publications and Sleeping Beauties

From 2008 to 2014, JMBE published a total of 426 articles. The proposed information awakening system found that 356 of these articles were in deep sleep and that 50 were in less deep sleep. The proportion of sleeping articles increased in 2013 and 2014, possibly because these more recent articles had not yet had sufficient time to be cited.

Of the 426 articles, 14 were sleeping beauties, which are presented both in terms of year of publication and year of awakening. Of these 14 sleeping beauties, 3 were awakened by a keyword prince (Table 1 and Fig. 6) and 7 were awakened by an H-index prince (Fig. 7). Only one was awakened by both types of prince.Top-5 keywords for 2008–2014

4.2 Keyword Prince

Table 1 shows the top five keywords from 2008 to 2014. With the exception of "Tissue engineering," the four most popular keywords all awakened these 3 sleeping beauties. That is, 80 % of the top keywords played an awakening role. This verifies the efficacy of the keyword prince in awakening such articles. For example, an article titled "Application of Near Infrared Spectroscopy and Imaging for Motor Rehabilitation in Stroke Patients" was published in 2009 with the keyword "Stroke". However, the article was only awakened in 2012 when "Stroke" became a popular keyword in research.

This approach takes a holistic view of the awakened information. Such articles are "ahead of their time" at the

Table 1 Top-5 keywords for 2008–2014

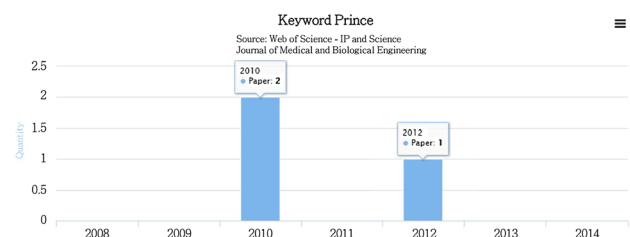| Hot Keywords	Quantity
Stroke	10
Tissue engineering	8
Brain computer interface (BCI)	8
Heart rate variability (HRV)	8
Electrocardiogram (ECG)	7

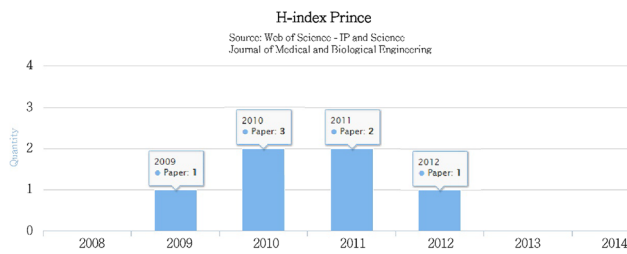**Fig. 6** Papers awakened by keyword prince by year

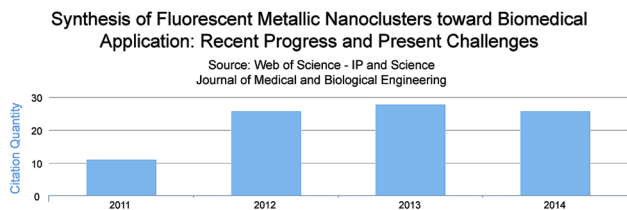**Fig. 7** Papers awakened by H-index prince by year

Fig. 8 Number of citations following awakening

time of publication, and recognition only comes when the research community as a whole is primed to accept and/or understand the sleeping research. A journal that regularly features such articles could indicate that its editors and contributors are relatively far-sighted and pioneering.

4.3 H-index Prince

The implications in Fig. 7 suggest that citation by influential authors increases attention on specific articles, thus increasing the likelihood of them being awakened [8].

An article entitled "Synthesis of Fluorescent Metallic Nanoclusters toward Biomedical Application: Recent Progress and Present Challenges" was published in 2009, but had slept until 2011 when it was cited in another article whose primary author had an H-index of 31. Over the following 4 years, the original article was cited a total of 89 times, suggesting that, once awakened, an article will attract a steady stream of citations (Fig. 8).

4.4 Contributions

Although several previous studies have examined the phenomenon of information awakening (Tables 2, 3), no standardized data analysis process or systematic methodology has emerged for the discovery and analysis of sleeping beauties. This study focuses on developing an automated analytical framework that can be applied to various journals. The framework is divided into three stages, namely data collection, processing, and analysis. This automated process reduces the impact of human subjectivity, and makes the following four significant contributions:

Table 2 Quantity analysis of articles in deep sleep and light sleep and sleeping beauties in each year

Year	No. of papers in deep sleep	No. of papers in light sleep	No. of published sleeping beauties	No. of awakened sleeping beauties
2014	85	4	0	0
2013	72	5	1	1
2012	47	9	2	8
2011	42	15	2	3
2010	40	11	5	2
2009	41	4	3	0
2008	29	2	1	0
Total	356	50	14	14

Table 3 Review of literature associated with sleeping beauties

Reference	Objectives/description
Braun et al. [16]	Explores the relationship between princes and sleeping beauties in a given domain. It also describes the conditions required for a prince to awaken a sleeping beauty. Journals with a high impact factor have a higher chance of being cited
Wang et al. [7]	Focuses on conditions required to awaken sleeping beauties
Van Raan [5]	Analyzes the sleeping beauty phenomenon and defines the sleeping cycle in terms of deep sleep and less deep sleep

1. Creates an explicit analytical framework for analyzing various types of journal through WoS.
2. Provides specific information on sleeping articles in JMBE.
3. Provides a clear understanding of keyword usage to increase exposure and citations.
4. Provides insight into the sleeping beauty phenomenon in journals, to effectively assist decision-makers in further journal development.

5 Conclusion

This study designed and implemented an IAF to identify sleeping beauties and their princes. These princes are found to take two forms: keyword princes awaken articles when certain keywords become popular, and H-index princes are influential scholars who cite otherwise unnoticed articles, thus attracting the interest of others.

Many sleeping beauties are still dormant, awaiting discovery, and the proposed system provides a means of

potentially identifying such resources, thus possibly accelerating the awakening process.

This study suffers from three key limitations. It uses the WoS database to search JMBE archives, and thus the data collected are limited to articles stored in the SCI. Secondly, self-citations may be of questionable value in the awakening process, and authors could potentially game the system by citing their own sleeping articles, thus serving as their own H-index prince. As a result, certain articles of questionable value may be identified as sleeping beauties. However, self-citations could potentially be legitimate and provide useful information, so the proposed system still takes them into account.

Finally, only the H-index of the primary author is used for evaluation, mainly because the primary author is normally responsible for the research and manuscript. However, this may underplay the potentially significant roles played by co-other authors.

Although this study clearly defines the two types of princes who awaken sleeping beauties, only 11 of the 14 sleeping beauties found in the JMBE database between 2008 and 2014 were awakened by princes, leaving 3 sleeping beauties awakened for reasons yet to be determined. One possible reason is that, from the perspective of big data, the awakening of these sleeping beauties may be due to the correlation proposed by Wright [58]. Mayer-Schönberger also stated that in the era of big data, correlation is more important than causation [14], and there may be a certain degree of correlation between the sleeping articles and the articles which awaken them. Future research could use such a correlation as the foundation for the design of new methodologies and analysis techniques.

Although the IAF is only applied to JMBE, data from other journals is structured similarly, including article name, author names, number of citations, H-index of authors, and keywords. Thus, the proposed system can be applied to other journals in the future. Moreover, the system can be used to analyze keywords across different journal articles to provide a more accurate analysis of current research trends.

Future studies can also focus on the role played by the keyword prince. Although the current study analyzes articles awakened by keyword princes, it cannot provide a comprehensive explanation for why certain sleeping beauties are awakened, while others remain dormant. Therefore, future work should adopt text mining techniques to overcome the existing limitations of author-selected keywords to collect and analyze keywords automatically extracted from the article title and abstract, thus increasing the accuracy of analytical results for keyword princes.

References

1. Voronin, Y., Myrzahmetov, A., & Bernstein, A. (2011). Access to scientific publications: The scientist's perspective. *PLoS ONE, 6*, e27868.
2. Wyatt, H. V. (1975). Knowledge and prematurity: The journey from transformation to DNA. *Perspectives in Biology and Medicine, 18*, 149–156.
3. Barber, B. (1963). Resistance by scientists to scientific discovery. *American Journal of Clinical Hypnosis, 5*, 326–335.
4. Cole, S. (1970). Professional standing and the reception of scientific discoveries. *American Journal of Sociology, 76*, 286–306.
5. Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics, 59*, 467–472.
6. Mendel, G. (1866). Versuche über Pflanzenhybriden, Verhandlungen des naturforschenden Vereines in Brünn.
7. Wang, J., Ma, F., Chen, M., & Rao, Y. (2012). Why and how can "sleeping beauties" be awakened? *The Electronic Library, 30*, 5–18.
8. Li, J., Shi, D., Zhao, S. X., & Fred, Y. Y. (2014). A study of the "heartbeat spectra" for "sleeping beauties". *Journal of Informetrics, 8*, 493–502.
9. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. (2012). Available via http://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf.
10. Davenport, T. H., Barth, P., & Bean, R. (2013). How 'big data' is different. *MIT Sloan Management Review, 54*, 21–25.
11. Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't.* New York: Penguin.
12. D Data Management: Controlling Data Volume, Velocity and Variety. (2012). Available via http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
13. Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: analytics for enterprise class hadoop and streaming data.* New York: McGraw-Hill Osborne Media.
14. Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Boston: Houghton Mifflin Harcourt.
15. Edosio U. Z. (2014). Big data paradigm-analysis, application, and challenges. *13th Research seminar series workshop.*
16. Braun, T., Glänzel, W., & Schubert, A. (2010). On sleeping beauties, princes and other tales of citation distributions…. *Research Evaluation, 19*, 195–202.
17. McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review, 90*, 60–66.
18. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144.
19. Maury, M. F. (2003). *The physical geography of the sea and its meteorology.* New York: Dover.
20. Beaty, J. J. (1966). *Seeker of seaways: A life of Matthew Fontaine Maury pioneer oceanographer.* Milano: Pantheon Books.
21. Lewis, C. L. (1927). *Matthew Fontaine Maury: The pathfinder of the seas.* Annapolis: Naval Institute.
22. Costa, F. F. (2014). Big data in biomedicine. *Drug Discovery Today, 19*, 433–440.
23. Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data, 1*, 51–59.

24. Baumgartner, H. (2010). Bibliometric reflections on the history of consumer research. *Journal of Consumer Psychology, 20*, 233–238.

25. Radicchi, F., & Castellano, C. (2013). Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics, 97*, 627–637.

26. Stent, G. S. (1972). Prematurity and uniqueness in scientific discovery. *Scientific American, 227*, 84–93.

27. Romans, L. J. (1986). Massive N = 2a supergravity in ten dimensions. *Physics Letters B, 169*, 374–380.

28. Polchinski, J. (1995). Dirichlet branes and ramond–ramond charges. *Physical Review Letters, 75*, 4724.

29. Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*, 359–375.

30. Brimblecombe, P., & Grossi, C. M. (2009). The bibliometrics of atmospheric environment. *Atmospheric Environment, 43*, 9–12.

31. Baumgartner, H. (2010). Bibliometric reflections on the history of consumer research. *Journal of Consumer Psychology, 20*, 233–238.

32. Van Dalen, H. P. (2005). Signals in science-on the importance of signaling in gaining attention in science. *Scientometrics, 64*, 209–233.

33. Li, J., & Ye, F. Y. (2012). The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics, 92*, 795–799.

34. Burrell, Q. L. (2005). Are "sleeping beauties" to be expected? *Scientometrics, 65*, 381–389.

35. Xu, F., Liu, W. B., & Mingers, J. (2015). New journal classification methods based on the global H-index. *Information Processing and Management, 51*, 50–61.

36. Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*, 471–479.

37. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 16569–16572.

38. Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics, 3*, 273–289.

39. Liu, Y., & Yang, Y. (2014). Empirical study of L-Sequence: the basic H-index sequence for cumulative publications with consideration of the yearly citation performance. *Journal of Informetrics, 8*, 478–485.

40. Rousseau, R. (2006). New developments related to the Hirsch index.

41. Ausloos, M. (2015). Assessing the true role of coauthors in the H-index measure of an author scientific impact. *Physica A: Statistical Mechanics and its Applications, 422*, 136–142.

42. Ball, P. (2005). Index aims for fair ranking of scientists. *Nature, 436*, 900.

43. Van Raan, A. F. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67*, 491–502.

44. dos Santos Rubem, A. P., & de Moura, A. L. (2015). Comparative analysis of some individual bibliometric indices when applied to groups of researchers. *Scientometrics, 102*, 1019–1035.

45. Du, H., Li, B., Brown, M. A., Mao, G., Rameezdeen, R., & Chen, H. (2014). Expanding and shifting trends in carbon market research: A quantitative bibliometric study. *Journal of Cleaner Production, 103*, 104–111.

46. Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics, 69*, 169–173.

47. Banks, M. G. (2006). An extension of the Hirsch index: Indexing scientific topics and compounds. *Scientometrics, 69*, 161–168.

48. Bar-Ilan, J. (2008). Which H-index?-A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*, 257–271.

49. Meho, L. I., & Rogers, Y. (2008). Citation counting, citation ranking, and H-index of human–computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology, 59*, 1711–1726.

50. Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*, 2105–2125.

51. Franceschini, F., & Maisano, D. (2010). The citation triad: an overview of a scientist's publication output based on Ferrers diagrams. *Journal of Informetrics, 4*, 503–511.

52. Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences, 275*, 314–347.

53. Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: issues and challenges moving forward. *Proceedings IEEE Hawaii International Conference, 46*, 995–1004.

54. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144.

55. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: review and open research issues. *Information Systems, 47*, 98–115.

56. Sheikh, N. (2013). *Implementing analytics: A blueprint for design, development, and adoption*. San Francisco: Morgan Kaufmann.

57. Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. *PLoS Biology, 5*, e18.

58. Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.