

臺灣科學引文索引系統之設計與實作
Design and Implement A Taiwan Science Citation Index Journal System

林志垣(**Chih-Yuan Lin**) 郭建明(**Gen-Ming Guo**)

南台科技大學資訊管理系
Department of Information Management
Southern Taiwan University of Technology

摘要

臺灣科學引文索引(TSCI)系統是一個可以作為查詢、研究、以及評估臺灣科學研究的引文索引資訊系統。本論文描述 TSCI 期刊資料庫的設計、架構、特色、政策、規則、系統與應用。由於引文分析與文獻計量學已經被視為是一種系統化調查學術研究發展和趨勢的方法，本研究將應用引文分析暨文獻計量方法來設計與建構此一資料庫系統。另外，由於 SCI 所收錄的期刊大部份是英文科學類的期刊，因此 TSCI 可以補 SCI 引文索引之欠缺處，並讓研究者可以從臺灣的觀點，有更多的角度來探討臺灣科學的研究成果。

關鍵字：臺灣科學索引、中文期刊資料庫、引文分析、文獻計量

Abstract

The purpose of Taiwan Science Citation Index (TSCI) system is to design one Taiwan journal citation database platform which can be used to query, study and evaluate scientific research. This article describes the design, architecture, feature, policy, rule, system and application of TSCI journal database. Both citation analysis and bibliometrics had been thought as the mainstream methodology to study academic research development and trend in the past. These methodologies would be used to design and implement TSCI platform. The scientific journals indexed by Science Citation Index (SCI) are most published by English language. TSCI indexes Taiwan scientific journals published by traditional Chinese language. Therefore, TSCI journal database can provide scholars more resources to study the scientific research activities in Taiwan.

Keywords: TSCI, Chinese Journal Database, Citation Analysis, Bibliometry.

1. Background and Related Works

1.1 Background and Purpose

The tremendous increase in the number of serial publications, coupled with spiraling subscription costs and shrinking library budgets, are driving what many describe as the "crisis in scholarly publishing." (Scholarly, 2005). And the impact of every journal and peer-reviewed paper for scholars is variable all the time. In order to calculate the citation impact factor for every journal and provide citation index search for scientists, many countries began to build up their own journal citation database for the published journals in their own language. For instance, the Science Citation Index database was built up in USA for English peer-reviewed papers and the Chinese Science Citation Index database was setup in China for Chinese peer-reviewer articles. In this research project, the Taiwan Science Citation Index database system was built up for the Taiwan's peer-reviewed papers published in the tradition Chinese language.

This study will be focus on the journals which are classified into science field rather than social science or art fields in Taiwan. Some journals could be classified into TSSCI and TSCI at the same time. As for SCI and SSCI, there are similar situations. Take MIS Quarterly and Scientometrics for instances, both were classified into SCI and SSCI journal lists. There are three major goals which we want to achieve in this project. First, the raw data preparing, cleaning and correcting system would be designed and implemented. Second, the search function and basic statistics tools would be implemented. Third, the Journal Citation Report would be generated from this TSCI database system platform.

1.2 Related Works

Thomson Inc. (Thomson, 2007) built SCI, SSCI and A&HCI database (Web of Science DB) and Journal Citation Report (JCR) for the scholar community to search peer-reviewed papers and journal assessment. Google Scholar (2007) indexes research resources on the internet in order to provide scholars with papers search and article citation ranking in interdisciplinary subjects. Graduate University of Chinese Academy of Sciences started to build CSCD (Chinese Science Citation Database) since 1989. Nanjing University and The Hong Kong University of Science & Technology began to setup CSSCI (China Social Science Citation Index) database in 1997 (CSSCI, 2007). In Taiwan, the Social Science Research Center under the National Science Council (NSC) constructs the TSSCI (Taiwan Social Science Citation Index) database (TSSCI, 2007). And the department of Library in National Taiwan University built the THCI (Taiwan Humanities Citation Index) database (THCI, 2007). THCI project was sponsored by the Humanities Research Center of NSC. Therefore, Taiwan needs one TSCI database and science journal citation report to index the peer-reviewed papers which written in traditional Chinese language and

belonged to Science field. In this project, we design one TSCI system platform in order to assist with building TSCI database and provide search and analysis service for this database.

2. Material and Method

2.1 Material

The materials used in the project are all from journals published in Taiwan. The journals are classified in advance. Only the journals belonged to the science cluster would be selected as the source. For the TSCI policy and principle, they would be described as the following:

2.2 TSCI Policy

In this project, we made some policies to guide our directions. That is because the journals which were published in Taiwan have many different formats. Most of journals accept different languages' manuscripts including traditional Chinese language and English language. The metadata of each article have both Chinese and English versions too. In order to make the TSCI database consistency, we try to define some TSCI policies as following:

1. The collected journal must belong to the science field.
2. The society or publisher for collected journal must be located at Taiwan.
3. The references for each article are limited to journal, textbook and magazine.
4. The metadata of each article including in title, keyword, author name, address, journal name, volume, issue, page and reference would be indexed under Taiwan's copyright law.

2.3 TSCI Principle

As for the TSCI principle, we refer to related works and define our TSCI principle as follows. However, the principles may be modified, added and changed when we find that it is necessary to do so.

A. *The Disciplinary*

Multi-disciplinary were classified into Science field. They are life science, medicine, mechanical engineering, mathematics, earth science, physical science, chemistry, computer science and civil engineering (Table 1). If the journal belongs to one of above disciplinary, it would be classified into the Science field. Therefore, TSCI journal source would come from these disciplinary.

Table 1. The Disciplinary Category

Fields	Disciplinary
Science	1. Life Science

	2. Medicine
	3. Mech. Engineering
	4. Mathematics
	5. Earth Science
	6. Physical Science
	7. Chemistry
	8. Computer Science
	9. Civil Engineering
Social Science	N/A
Art and Humanities	N/A

B. The MetaData of Journal Article Source

Many metadata formats of journal article source exist in many journals. In order to make output consistency, the united format in the TSCI platform would be as the following:

1. Title

1.1 The journal full name would be used rather than the abbreviation name. The journal name could be changed after a period of time. In TSCI database, we would renew and unify journal name continuously. The purpose of this is to make data consistency and accuracy in the TSCI database and journal citation report.

1.2 Both Chinese and English title would be collected. If article was written in English and its Chinese title was not provided, we would collect English title only.

1.3 Simplified Chinese article would be collected in the format of traditional Chinese characters.

2. Author

2.1 The author name begins with Last name. As for the name sequence, the first is Last name and the second is Middle name. Third, First name would be the last one and behind the Middle name. And all the names would be in the upper case. The second and first name would be replaced with single one character. (Ex: BUSH GW and 王大明)

2.2 If author name is unavailable, the translator's name or organization would be added to database in the field of author.

3. Date

The date format would use the international standard date notation, YYYY-MM-DD, where YYYY is the year in the usual Gregorian calendar, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31. Leading "0"'s are always used to pad single digit days and months. For example, the fourth day of February in the year 2003 is written in the

standard notation as 2003-02-04.

4. Publisher

If there are multiple publishers, only the first one would be added to the database in the publisher field.

5. Page

The page format would be defined as the following example: pp11-23. Different formats from various sources would be transformed to this format.

C. The Cited References for Journal Article Source

1) The Indexing Reference Type

There are many reference content types including original paper, review paper, editorial, letter, comment, news, magazine, textbook, dissertation, website, patent and so on (Table 2). In the early beginning phase, we only handle the reference type of peer-reviewed paper, magazine and textbook. Both conference and dissertation are considered and evaluated to decide whether we would process those types or not in the future.

Table 2. The indexing reference type in TSCI Database

Reference Content Type	Indexing Type
Original paper	Yes
Review Paper	Yes
Letter	No
Comment	No
Editorial	Yes
News	No
Magazine	Yes
Textbook	Yes
Dissertation	Consider
Conference	Consider
Website	No
Patent	No
Technique Report	No

2) The Reference Format

There are also many different formats in the references of journal article. The processes for these formats in TSCI platform are also very similar with the solutions and the consistent format which we process the metadata of journal article source.

1. Title

1.1 The journal full name would be used rather than the abbreviation name. The

journal name could be changed after a period of time. In TSCI database, we would renew and unify journal name each year. The purpose of this is to make data consistency and accuracy in the TSCI database and journal citation report.

1.2 Both Chinese and English references and titles would be collected. If reference was written in English and author did not provide Chinese references, we would collect English references only.

1.3 Simplified Chinese article would be filtered because we focus on the journal published in Taiwan.

2. Author

2.1 The author name begins with Last name. As for the name sequence, the first is Last name and the second is Middle name. Third, First name would be the last one and behind the Middle name. And all the names would be in the upper case. The second and first name would be replaced with single one character. This format is used by Thomson ISI Inc. (2007) now. As the result, different format of author name would be transformed to this consistent format.

2.2 If author name is unavailable, the translator name or organization would be added to database in the field of author.

3. Date

We would only collect the publication year. The date property in references is not like the date format in the front page of each paper. We cannot get the detail information such as submit date and accept date. The website reference may have last access or available date. But we don't process website references currently.

4. Page

The page format would be defined as the following example: pp11-23. Different formats from various sources would be transformed to this format.

5. The Indexing Reference Type

There are many reference content types including original paper, review paper, editorial, letter, comment, news, magazine, textbook, dissertation, website, patent and so on. Currently, we only handle the reference type of peer-reviewed paper and textbook. The references which do not belong to original article, review paper or textbook would be filtered.

3 Methods

3.1 Bibliometrics and Citation Analysis

In this TSCI project, we would apply the Bibliometrics and Citation Analysis research methodologies to our works. Some of these mature concepts and theories would be used to analyze the results. The detail definitions and descriptions for bibliometrics and citation analysis would be described as following:

1. Bibliometrics

Bibliometrics (2007) is a type of research method used in library and information science. It utilizes quantitative analysis and statistics to describe patterns of publication within a given field or body of literature. Researchers may use this method to determine or evaluate the influence of a single author, for example, or to describe the relationship between two or more authors or papers.

One of the main areas in bibliometric research concerns the application of bibliometric laws. The three most commonly used laws in bibliometrics are: Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence.

A) Lotka's Law

Lotka's Law describes the frequency of publication by authors in a given field. It states that ". . . the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent". This means that out of all the authors in a given field, 60 percent will have just one publication, and 15 percent will have two publications ($1/2^2$ times .60). 7 percent of authors will have three publications ($1/3^2$ times .60), and so on. According to Lotka's Law of scientific productivity, only six percent of the authors in a field will produce more than 10 articles. Lotka's Law, when applied to large bodies of literature over a fairly long period of time, can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog.

B) Bradford's Law

Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can be divided into three parts, each containing the same number of articles: 1) a core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles, 2) a second zone, containing the same number of articles as the first, but a greater number of journals, and 3) a third zone, containing the same number of articles as the second, but a still greater number of journals. The mathematical relationship of the number of journals in the core to the first zone is a constant n and to the second zone the relationship is n^2 . Bradford expressed this relationship as $1:n:n^2$. Bradford formulated his law after studying a bibliography of geophysics, covering 326 journals in the field. He discovered that 9 journals contained 429 articles, 59 contained 499 articles, and 258 contained 404 articles. So it took 9 journals to contribute one-third of the articles, 5 times 9, or 45, to produce the next third, and 5 times 5 times 9, or 225, to produce the last third. As may be seen, Bradford's Law is not statistically accurate, strictly speaking. But it is still commonly

used as a general rule of thumb.

C) Zipf's Law

Zipf's Law is often used to predict the frequency of words within a text. The Law states that in a relatively lengthy text, if you "list the words occurring within that text in order of decreasing frequency, the rank of a word on that list multiplied by its frequency will equal a constant. The equation for this relationship is: $r \times f = k$ where r is the rank of the word, f is the frequency, and k is the constant. Zipf illustrated his law with an analysis of James Joyce's *Ulysses*. "He showed that the tenth most frequent word occurred 2,653 times, the hundredth most frequent word occurred 265 times, the two hundredth word occurred 133 times, and so on. Zipf found, then that the rank of the word multiplied by the frequency of the word equals a constant that is approximately 26,500". Zipf's Law, again, is not statistically perfect, but it is very useful for indexers.

2. Citation Analysis

Another major area of bibliometric research uses various methods of citation analysis in order to establish relationships between authors or their publications. Here is a definition of citation analysis, and definitions of co-citation coupling and bibliographic coupling, which are specific kinds of citation analysis (Bibliometrics, 2007).

A) Citation Analysis

When one author cites another author, a relationship is established. Citation analysis uses citations in scholarly works to establish links. Many different links can be ascertained, such as links between authors, between scholarly works, between journals, between fields, or even between countries. Citations both from and to a certain document may be studied. One very common use of citation analysis is to determine the impact of a single author on a given field by counting the number of times the author has been cited by others. One possible drawback of this approach is that authors may be citing the single author in a negative context.

B) Co-citation Coupling

Co-citation coupling is a method used to establish a subject similarity between two documents. If papers A and B are both cited by paper C, they may be said to be related to one another, even though they don't directly cite each other. If papers A and B are both cited by many other papers, they have a stronger relationship. The more papers they are cited by, the stronger their relationship is.

C) Bibliographic Coupling

Bibliographic coupling operates on a similar principle, but in a way it is the mirror image of co-citation coupling. Bibliographic coupling links two papers that cite the same articles, so that if papers A and B both cite paper C, they may be said to be

related, even though they don't directly cite each other. The more papers they both cite, the stronger their relationship is.

3.2 TSCI Database Design

Figure 1 is the Entity-Relationship model of TSCI database. There are several entities including journal, article, author analysis, unit analysis, country analysis and category analysis. The schema for each entity is illustrated as the following table. Table 3 lists the journal entity attributes and Table 4 lists the article entity attributes. Both tables are designed to store the raw data of journal and article in the TSCI database. Table 5 is the article citation entity table. Each journal has its own citing journal and cited journal. Every article has its own citing references and cited references. Table 6 is the journal citation report entity table. Citing reference can be seen as the input of article's citation and Cited reference can be seen as the output of article's citation. In order to do further value-added analysis, the meta data of author analysis, unit analysis, country analysis and category analysis (Table 7-10) are designed to provide required fields to store citation calculation results. The journal impact factor, author impact factor, unit impact factor and country impact factor would be calculated and saved. Through these impact factors, we can generate some further citation reports such as journal citation report, unit citation report and so on. The publication year could be a good dimension to drill down the data analysis in TSCI database.

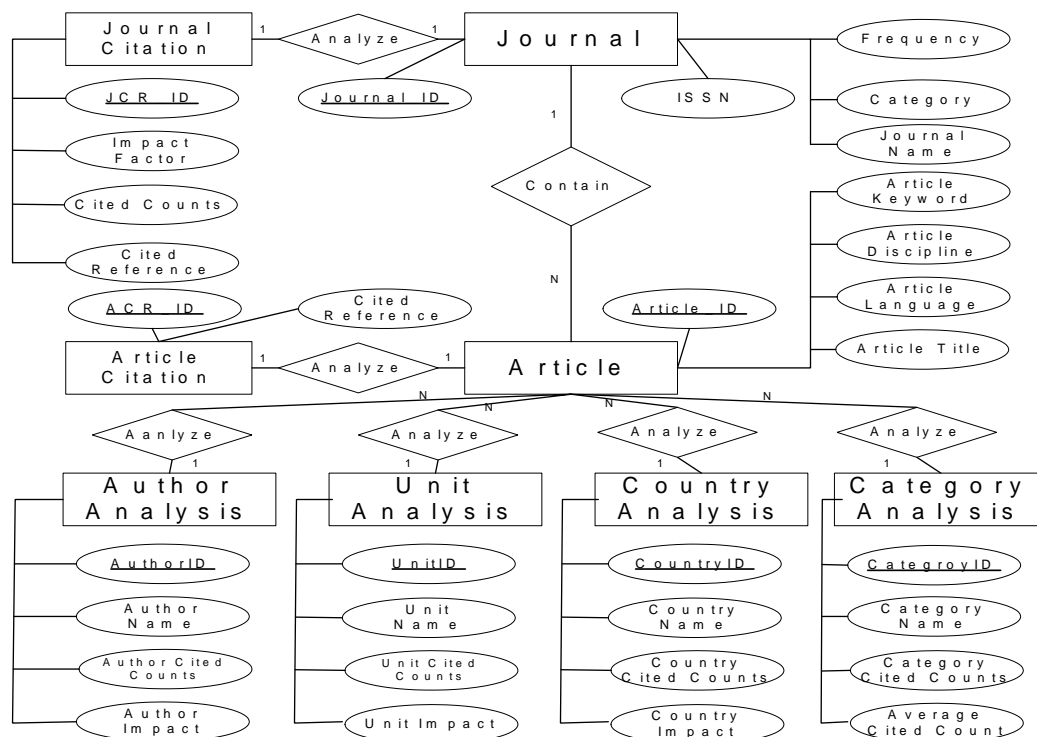


Figure 1. The Entity-Relation Model.

Table 3. Journal Entity Attributes

Field Label	Field Name	Data Type
Journal ID	Journal_ID	INT(10)
Journal Name	Journal_Name	VARCHAR(255)
ISSN or ISXN	ISSN	INT(10)
Subject Categories	Category	VARCHAR(255)
Publication Frequency	Publication_Frequency	INT(5)
Publication Date	Publication_Date	Date(7)
Publication Type	Publication_Type	VARCHAR(255)
Publisher	Publisher	VARCHAR(255)
Publisher Address	Publisher_Address	VARCHAR(255)
Publisher Website	Publisher_URL	VARCHAR(255)
National Library ID	NationalLibID	INT(10)
Journal Notes	Journal_Notes	VARCHAR(255)

Table 4. Article Entity Attributes

Field Label	Field Name	Data Type
Article_ID	Article_ID	INT(10)
Article_Title	Article_Title	VARCHAR(255)
Article_Language	Article_Language	VARCHAR(10)
Article_Discipline	Article_Discipline	VARCHAR(15)
Article_Vol	Article_Vol	INT(3)
Article_Issue	Article_Issue	INT(3)
Article_Page	Article_Page	INT(5)
Article_Date	Article_Date	DATE(8)
Article_Keyword	Article_Keyword	VARCHAR(255)
Article_Journal	Article_Journal	VARCHAR(100)
Article_Funding	Article_Funding	VARCHAR(100)
Article_Note	Article_Note	VARCHAR(150)

Table 5. Article Citation Entity Table

Field Label	Field Name	Data Type
Article ID	Article_ID	INT(10)
Citing Article ID	Article_Citing_ID	INT(10)
Citing Reference	Article_Citing_Reference	VARCHAR(255)
Cited Article ID	Article_Cited_ID	INT(10)
Cited Reference	Article_Cited_Reference	VARCHAR(255)
Cited Counts	Article_Cited_Counts	INT(10)
Notes	Notes	VARCHAR(255)

Table 6. Journal Citation Report Entity Table

Field Label	Field Name	Data Type
Journal ID	Journal_ID	INT(10)
Journal Name	Journal_Name	VARCHAR(255)
Journal Citing Counts	Journal_Citing_Counts	INT(10)
Journal Cited Counts	Journal_Cited_Counts	INT(10)
Journal Impact Factor	Journal_Impact Factor	FLOAT(10)
Immediacy Index	Journal_Immediacy	FLOAT(10)
Cited Half-Life Index	Journal_CitedHalfLife	FLOAT(10)
Citing Half-Life Index	Journal_CitingHalfLife	FLOAT(10)
Cited Year	Journal_CitedYear	DATE(8)
Citing Year	Journal_CitingYear	DATE(8)
JCR Notes	JCR_Notes	VARCHAR(255)

Table 7. Author Analysis Entity Table

Field Label	Field Name	Data Type
Author ID	AuthorID	INT(10)
Author Name	Author_Name	VARCHAR(255)
Author Publication Counts	Author_Pub_Counts	INT(10)
Author Cited Counts	Author_Cited_Counts	INT(10)
Author Impact Factor	Author_IF	FLOAT(10)
Author Publication Year	Author_Year	Date(8)
Notes	Notes	VARCHAR(255)

Table 8. Unit Analysis Entity Table

Field Label	Field Name	Data Type
Unit ID	UnitID	INT(10)
Unit Name	Unit_Name	VARCHAR(255)
Unit Publication Counts	Unit_Pub_Counts	INT(10)
Unit Cited Counts	Unit_Cited_Counts	INT(10)
Unit Impact Factor	Unit_IF	FLOAT(10)
Unit Publication Year	Unit_Year	Date(8)
Notes	Notes	VARCHAR(255)

Table 9. Country Analysis Entity Table

Field Label	Field Name	Data Type
Country ID	CountryID	INT(10)
Country Name	Country_Name	VARCHAR(255)
Country Publication Counts	Country_Pub_Counts	INT(10)
Country Cited Counts	Country_Cited_Counts	INT(10)
Country Impact Factor	Country_IF	FLOAT(10)
Country Publication Year	Country_Year	Date(8)
Notes	Notes	VARCHAR(255)

Table 10. Category Analysis Entity Table

Field Label	Field Name	Data Type
Category ID	CategoryID	INT(10)
Category Name	Category_Name	VARCHAR(255)
Category Publication Counts	Category_Pub_Counts	INT(10)
Category Cited Counts	Category_Cited_Counts	INT(10)
Category Average Cited Counts	Category_AVG_Counts	FLOAT(10)
Category Publication Year	Category_Year	Date(8)
Notes	Notes	VARCHAR(255)

3.3 TSCI System Design

The TSCI system platform is constructed via Microsoft Asp.Net and SQL Server 2005 on the Windows 2003 Server platform. The TSCI system platform is divided into three major sub-systems. They are journal indexing system, journal analysis system and journal citation system. The system architecture was illustrated as Figure 2 and Table 11.

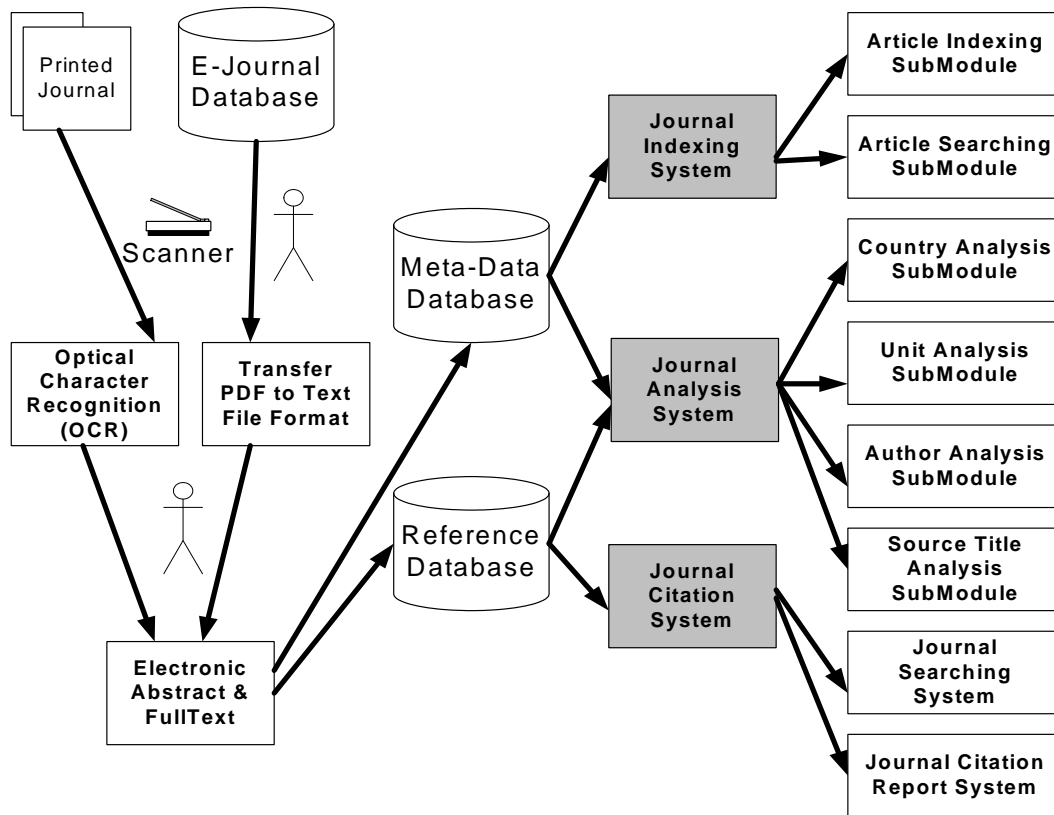


Figure 2. TSCI Database Platform Architecture

Table 11. The Building Block for the TSCI System Platform.

End-User / Scholar							
AIS	ASS	CAS	UAS	AAS	STAS	JSS	JCRS
Journal Indexing System		Journal Analysis System			Journal Citation System		
Journal Meta Database				Journal Reference Database			
Data Prepare		Data Clean			Data Filter		
Manual Work				PDF to Text Converter			
OCR (Optical Character Recognition)							
Printed Journal				Electronic Journal Database			
TSCI System Platform							

1) Journal Indexing System

There are two sub-modules in the journal indexing system. They are article indexing sub-module (AIS) and article searching sub-module (ASS). In the article indexing sub-module (AIS), articles would be indexed and stored in the meta-data database. These raw data would be obtained from printed journal or E-Journal database. After we process these raw data by Optical Character Recognition (OCR) or

Pdf-to-Word (Pdf2Word) software package, the electronic full-text becomes the source for Meta-Data database and Reference database.

In the article searching sub-module (ASS), two types of search functions were designed and implemented. They are basic and advanced search functions. For the basic search, we provide basic search for journal, author and article. The return fields are the basic properties in article and journal database tables such as author name, journal name, article title and so on. As for the advanced search, we provide richer search properties from journal, author and article database tables.

2) Journal Analysis System

In the journal analysis system, we design the offline analysis mode. There are four analysis sub-modules in journal analysis system. They are author analysis sub-modules (AAS), unit analysis sub-modules (UAS), country analysis sub-modules (CAS) and source title analysis sub-modules (STAS). For this the offline mode, TSCI platform would generate analysis reports based on its own classification policy and then adjust report by human being. These reports could become the electronic academic activity analysis reports.

Some other impact factors can be calculated via this system too. Three different kinds of value-added impact factors, including author impact factor (AIF), unit impact factor (UIF) and country impact factor (CIF), which can be calculated and extended from the raw data in the TSCI database. In Formula 3.1, which is the AIF calculation, F represents journal impact factor (JIF); M represents author member numbers; T represents article types including News, Review, Letter and Editorial Paper; and W represents the weight of article type. In Formulas 3.2 and 3.3, A represents author impact factor. U represents unit impact factor. Owing to the hierarchical and aggregate relationship among country, unit and author, we can firstly calculate AIF. Then UIF can be further obtained from AIF, and CIF can be further calculated from UIF recursively.

$$AIF = \sum_{i=1}^n F_i M_i^{-1} (W_x T_i) \dots\dots\dots(3.1)$$

$$UIF = \sum_j^k (\sum_i^n A_i)_j \dots\dots\dots(3.2)$$

$$CIF = \sum_i^n U_i \dots\dots\dots(3.3)$$

3) Journal Citation System

There are two sub-modules in the journal citation system. They are Journal Searching System (JSS) and Journal Citation Report System (JCRS). In the journal

searching system, we would provide much more advanced searching features such as cited and citing reference search. The journal citation report (JCR) in journal citation report system (JCRS) is the one of the most highly valued target in the TSCI platform. That is because JCR can be used to evaluate the impact factor of scientific journal published in Taiwan.

Due to the bi-directional citations between journals, we call this journal cited analysis and journal citing analysis. Similarly, it could be called journal linking in/out analysis. Four related factors would be used in this research: the *JCDF* (Journal Cited Factor), *SJCDF* (Subtract Self-Journal Cited Factor), *JCIF* (Journal Citing Factor) and *SJCIF* (Subtract Journal Citing Factor). In the $JCDF_k$ (Formula 3.4), DF_j is the abbreviation for Cited-Frequency, which counts the base journal's cited frequency (link-in) from “ n ” kinds of journals. In the citation matrix as in Table 12, DF_j is the cited counts in the horizontal direction. DA_k is abbreviated from Cited-Articles, which represents the amounts of articles published by the base journal. And, “ k ” is the target journal, whose citation factor is being calculated. “ n ” is the total number of journals. “ j ” is the base journal index ($j=1, \dots, n$). While the $SJCDF_k$ (Formula 3.5) is being calculated, the situation for $j=k$ must be skipped because its purpose is to eliminate noise coming from journal self-citation. The major difference between the $JCDF_k$ and $JCIF_k$ (Formula 3.4 and 3.6) is that the $JCIF_k$ is for journal co-citing counts and the other is for journal co-cited frequencies. For the direction in the citation matrix, $JCDF_k$ is in the horizontal direction and $JCIF_k$ is in the vertical direction. In $JCIF_k$, GF_j is abbreviated from Citing-Frequency, which counts the base journal's citing frequency (link-out) from “ n ” kinds of journals. In Table 10, GF_j is the citing counts in the vertical direction. GA_k is abbreviated from Citing-Articles, which represents the amounts of papers published by the journal “ k ”. The $SJCIF_k$ (Formula 3.7) is similar to the $SJCDF_k$ both subtracting self-journal citation from $JCIF$. Usually, every journal has a higher self-citation factor, which can be found in references (Rousseau R., 1999) (András S. 2006). That is why we develop $SJCIF$ and $SJCDF$ factor in order to subtract the citation bias which came from self-journal citation.

$$JCDF_k = \sum_{j=1}^n (DF_j / DA_k) \dots\dots\dots(3.4)$$

$$SJCDF_k = \sum_{j=1}^{n-1} (DF_j / DA_k) \quad \text{where } j \neq k \dots\dots\dots(3.5)$$

$$JCIF_k = \sum_{j=1}^n (GF_j / GA_k) \dots\dots\dots(3.6)$$

$$SJCIF_k = \sum_{j=1}^{n-1} (GF_j / GA_k) \quad \text{where } j \neq k \dots\dots\dots(3.7)$$

Table 12. Journal Cited Factor (JCDF) and Journal Citing Factor (JCIF).

	MS	MISQ	IEETSE	ISR	JMIS	OrgSCI	ASQ	I&M	DS	DSS	JCDF	SJCDF
MS	<u>15.4</u>	1.3	1.9	1.4	1.9	1.6	0.5	4.1	3.7	1.0	<u>3.3</u>	1.9
MISQ	13.5	<u>18.0</u>	2.2	12.0	17.8	1.1	0.0	40.3	9.6	4.3	<u>11.9</u>	11.2
IEETSE	0.2	0.0	<u>8.9</u>	0.1	0.3	0.0	0.0	0.4	0.1	0.3	<u>1.0</u>	0.2
ISR	6.8	5.7	0.8	<u>6.0</u>	4.6	0.5	0.0	7.7	2.0	1.8	<u>3.6</u>	3.3
JMIS	1.7	3.0	0.4	1.4	<u>7.1</u>	0.2	0.0	12.3	1.9	2.1	<u>3.0</u>	2.6
OrgSCI	5.0	3.5	11.1	1.4	3.1	<u>11.1</u>	2.9	4.4	2.1	0.6	<u>4.5</u>	3.8
ASQ	3.8	1.6	10.8	0.6	3.6	10.8	<u>4.9</u>	5.9	1.6	0.7	<u>4.4</u>	4.4
I&M	0.8	0.7	0.1	0.7	2.2	0.0	0.0	<u>12.2</u>	0.6	0.9	<u>1.8</u>	0.7
DS	4.8	2.5	0.7	3.1	4.1	0.5	0.0	10.7	<u>14.8</u>	1.7	<u>4.3</u>	3.1
DSS	1.0	0.5	0.1	0.7	1.1	0.0	0.0	3.2	0.6	<u>4.1</u>	<u>1.1</u>	0.8
JCIF	<u>5.3</u>	<u>3.7</u>	<u>3.7</u>	<u>2.7</u>	<u>4.6</u>	<u>2.6</u>	<u>0.8</u>	<u>10.1</u>	<u>3.7</u>	<u>1.7</u>		
SJCIF	4.2	2.1	3.1	2.4	4.3	1.6	0.4	9.9	2.5	1.5		

4. Result and Discussion

The Journal Impact Factor (JIF) is a measure of the frequency with which the "average article" in a journal has been cited in a given period of time. The impact factor for a journal is calculated based on a three-year period, and can be considered to be the average number of times published papers are cited up to two years after publication. For example, the journal impact factor 2005 for a journal would be calculated as follows: 1) A = the number of times articles published in 2003-4 were cited in indexed journals during 2005; 2) B = the number of articles, reviews, proceedings or notes published in 2003-4; 3) impact factor 2005 = A/B.

In the table 13, we show the journal impact factor for those highly cited journals published in Taiwan. The time span is from 2002 to 2004. The journal impact factor was calculated as above formula which was proposed by Thomson ISI Inc. The JIF-Remove-SC means that the journal impact factor was calculated after we remove the journal self citation. The articles in one journal often cite the articles published in the same journal. It will cause the bias to calculate journal impact factor. Therefore, the self-cited counts was deducted in this study in order to avoid this bias.

Most highly cited journals are from the medical field. That is because Taiwan has many medical journals. There are fewer engineering journals in Taiwan. Therefore, some famous journals such as "Journal of Information Science and Engineering" which is a SCI journal was poorly cited in TSCI database. The amount of relevant journals is one of the important variables to influence the cited counts and journal impact factor.

Table 13. The Journal Impact Factors List in TSCI database.

Chinese J.Nam e	Papers	Self Citation(SC)	Cited Counts	JIF	JIF Rem ove-SC
中華輔導學報	29	3	11	0.379	0.276
護理雜誌	267	0	71	0.266	0.266
台灣醫學	103	13	38	0.369	0.243
醫務管理期刊	91	0	22	0.242	0.242
安寧療護	72	0	16	0.222	0.222
新台北護理期刊	64	3	14	0.219	0.172
慈濟護理雜誌	153	0	26	0.170	0.170
應用心理研究	160	14	39	0.244	0.156
醫護科技學刊	81	4	15	0.185	0.136
榮總護理	140	0	19	0.136	0.136
臺灣醫學會雜誌	432	54	111	0.257	0.132
長庚護理	137	0	17	0.124	0.124
都市與計劃	59	6	12	0.203	0.102
中華職業醫學雜誌	100	5	14	0.140	0.090
林業研究季刊	93	0	8	0.086	0.086
台灣家庭醫學雜誌	61	2	7	0.115	0.082
台灣公共衛生雜誌	154	4	16	0.104	0.078
台灣醫界	339	0	21	0.062	0.062
醫學教育	150	30	39	0.260	0.060
衛生教育學報	53	8	11	0.208	0.057
臨床醫學	494	0	27	0.055	0.055
諮商與輔導	369	0	19	0.051	0.051
當代醫學	499	0	19	0.038	0.038
感染控制雜誌	133	23	28	0.211	0.038
物理治療	90	12	15	0.167	0.033
社區發展季刊	352	0	10	0.028	0.028
中華林學季刊	120	27	29	0.242	0.017
台耳醫誌	144	12	14	0.097	0.014
台灣農業化學與食品科學	180	11	13	0.072	0.011

5. Conclusion

In this study, one Taiwan Science Citation Index platform was designed and implemented. There are three subsystems including journal indexing system, journal analysis system and journal citation system in this TSCI platform. The indexing system was designed to provide raw data and basic utility. The analysis system was designed to add-value to raw data in database in order to proceed with many different statistics analysis. The Journal Citation Report was one of the most important results obtained from the journal citation report system. It could provide an assessment indicator to different journal.

This research project can contribute to five groups of audiences. They are University faculty, University student, Librarian, National Science Foundation committee member and Ministry of Education officer. As for university faculty, they can choose suitable Taiwan journals in order to submit their manuscripts. For university student, they could look for Chinese journal paper with high impact factor to digest them in order to save more study time. As for librarian, they have to subscribe core journals for their university. The TSCI Journal Citation Report may become a good reference for librarians to subscribe highly cited journals rather than all the Chinese journals. For National Science Foundation (NSF) committee members, they will evaluate scholars in order to grant funds to them. Through the Journal Citation Report of TSCI, NSF members may get more information about Taiwan's science journals and consider to sponsor those good journals and authors. For Ministry of Education officers, TSCI JCR can help them to evaluate Taiwan's science journals. Currently, they use TSSCI as the indicator to evaluate Taiwan's social science journals. However, there are many good Taiwan's science journals which are not indexed by TSSCI database.

Acknowledgement

This project was sponsored by National Science Council. The project serial number is NSC 95-2221-E-218-024.

References

- András S., Wolfgang G., Bart T., 2006, The weight of author self-citations. A fractional approach to self-citation counting, *Scientometrics*, 67:3, 503-514.
- Barbara, S. S., et al., 2003, Developing an Academic Medical Library Core Journal Collection in the Post-print Era: The Florida State University College of Medicine Medical Library Experience, *J Med Lib Assoc*, 91:3, 292-302.
- Bibliometrics, 2007, 1/24-last access,
Available: www.gslis.utexas.edu/~palmquis/courses/biblio.html.
- Bill, C. H. and Kent, A. W., 1997, Forum for MIS Scholars, *Communications of the ACM*, 40:11, 119-124.
- Bradford, S. C., 1948, London: Crosby Lockwood & Sons, *Documentation*.

- Bradford, S. C., 1934, Sources of Information on Specific Subjects, *Engineering*, 137, 85-86.
- Brookes, B. C., 1968, The Derivation and Application of the Bradford-Zipf Distribution, *Journal of Documentation*, 24, 247-265.
- CSSCI, 2007, 8/24-last access, Available: cssci.nju.edu.cn
- Google Scholar, 2007, 8/10-last access, Available: scholar.google.com.
- Hwang, S. Y., et al., 2001, A Survey of Journal Publications of MIS Department in Taiwan, *Journal of Information Management*, 9:1, 217-239.
- ISI Corp. 2007, 9/1-last access, *ISI Web of Knowledge*. Available: www.isinet.com.
- Janice, S. L., and John, D. M., 2002, Defining the Undergraduate Core Journal Collection, *The Serials Librarian*, 43:1, 45-59.
- Jeffrey, D. K., Kristin, H. G. and CyNthia, D., 1998, A Method for Building Core Journal Lists in Interdisciplinary Subject Areas, *Journal of Document*, 54:4, 477-488.
- Junping Q., et al., 2002, The Course, Status quo and Trend of Bibliometrics in China, *Journal of Information, Communication and Library Science*, 9:1, 1-11.
- Kobulnicky, P. J., 1977, Physics libraries and literature, *Encyclopedia of Library and Information Sciences*, 22, 202-221.
- Leslie K. W., 1996, Exciting Potential of Scholarly Electronic Journals, *CAUT Bulletin*, 43:7, 9-15.
- Rousseau R., 1999, Temporal differences in self-citation rates of scientific journals, *Scientometrics*, 44:3, 521-531.
- Scholarly Societies Project, 2007, 8/24, Last access, Available:www.scholarly-societies.org/crisis.html.
- THCI, 2007, 7/20-last access, *Taiwan Humanities Citation Index*, Available:www.hrc.ntu.edu.tw/thci.htm
- TSSCI, 2007, 1/10-last access, *Taiwan Social Science Citation Index*, Available:ssrc.sinica.edu.tw.
- TSCI, 2007, 9/12-last access, *Taiwan Science Citation Index*, Available:tsci.scholarworld.org.
- Weka, 2007, 8/23 last access, Data Mining Software, Available:www.cs.waikato.ac.nz/ml/weka/
- Zipf G. K., 1949, Human Behavior and the Principle of Least Efforts: An Introduction to Human Ecology, *Addison Wesley*.
- Zhiren Z., 2002, On the Characteristic Chinese Database - Chinese Social Science Citation Index, *Journal of Information, Communication and Library Science*, 9:1, 57-61.