

Performance Analysis of IES Journals using Text Processing Robots in PERL

by

Jiao Yu

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
May 7, 2012

Keywords: Internet robots, Text processing, PERL, Excel, Impact Factor

Copyright 2012 by Jiao Yu

Approved by

B. M. Wilamowski, Chair, Professor, Electrical and Computer Engineering
John Y. Hung, Professor, Electrical and Computer Engineering
Thaddeus Roppel, Associate Professor, Electrical and Computer Engineering

Abstract

In the past, many approaches to measure the quality of journals are developed, e.g. 2-year Impact Factor (IF), 5-year IF, Eigenfactor Score, etc. Most of them are related to the number of citations of published papers. Unfortunately, the citation analysis is no easy task and almost impossible using manual examination of references. [1] This must be done by developing special computer tools for extracting data from various locations. Also, if only citations are of interest then this information is already preprocessed on different web sites such as GoogleScholar, PublishOrPerish, or WebOfKnowlege. However, if for example, someone wants to analyze the performance of editors, associate editors, and reviewers, then the problem is much more complicated than to treat the journal as a whole. These would require development of specialized computer tools for automatic data processing. The method proposed in this thesis is targeted at answering advanced performance analysis as listed before. A text processing robot is developed here using PERL, with the aid of its powerful regular expressions and Excel processing packages. In conjunction with the Internet Robot developed by [2], a large amount of valuable information can be extracted about performance of editors, associate editors, and reviewers.

Acknowledgments

I would like to express my sincere thanks to my advisor, Prof. B. M. Wilamowski, who constantly provided valuable guidance and detailed help during my master's study. He taught me not only the specific way to solve the problems in my thesis, more importantly, he inspired me how to think innovatively in a fresh and different way. Also, his attitude towards research and life has benefited me a lot.

Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
List of Figures.....	vi
List of Tables.....	vii
List of Abbreviations.....	viii
Chapter 1 Traditional measures of Journal Quality.....	1
1.1 2-year Impact Factor.....	1
1.2 ES (Eigenfactor Score) and AIS (Article Influence Score).....	3
Chapter 2 Fundamentals of the Internet and Text Processing Robots.....	4
2.1 Introduction.....	4
2.2 Perl scripting language.....	7
Chapter 3 Evaluation of the performance of the Editorial Boards.....	10
3.1 Innovative measures of journal quality (EIC, AE, SS).....	10
3.2 Citation Based Evaluation.....	10
3.2.1 Evaluation of Editorial Boards.....	10
3.2.2 Evaluation of Special Sections.....	16
3.3 Time based Evaluation.....	20
3.3.1 Extract Submission Date, First Decision Date and Acceptance Date.....	20
3.3.2 Computation of Passed Days Between Two Dates.....	23

3.4	Results	24
3.4.1	Quality of the Review Process	24
3.4.2	Citation Analysis for Special Sections	31
3.4.3	Timely Performance of the Review Process	34
Chapter 4	Implementation of the Text Processing Robot.....	37
4.1	Integrate Data of Interest	37
4.2	Get the Publication Issue	41
4.3	Time Averaged Citation Number for Papers	44
4.4	Averaging Citations for AEs.....	44
4.5	Average Citations for SS	46
4.6	Average Time Analysis.....	48
Chapter 5	Conclusion and Future work	50
References	52
APPENDIX A:	combine_data.pl	55
APPENDIX B:	analyze.pl	60
APPENDIX C:	aveCitations_AE	64

List of Figures

1	One Year Impact Factors Trends for IES journals.....	2
2	Original IEEE Xplore webpage	5
3	Formatted output file from the Internet Robot.....	6
4	The Publish or Perish software	6
5	Fragment of the raw “TII_citation.xls”	12
6	Fragment of “TII_ManuscriptReceived.xls”	12
7	Combined data with both Editor Info and Citation Info	12
8	One Year Impact Factors Trends for IES journals.....	15
9	IEEE Xplore Webpage of TII Volume 7, Issue 4	17
10	“Table of Content” from TII Volume 7, Issue 4.....	18
11	Snapshot of “TII_citation.xls” with paper type information	19
12	Flow chart of the algorithm to extract the “Acceptance Date” for a paper.....	22
13	Average time between submission and the first decision for TIE and TII	34
14	Average time between submission and the final decision for TIE and TII	35
15	Average time between acceptance and the publication for TIE and TII.....	35
16	Average time between submission and publication for TIE and TII.....	36
17	Snapshot of “TII_citation.xls” with all the data needed	43
18	Snapshot of “TII_citation.xls” with time gaps information	49

List of Tables

Table 1 Impact factor calculations for IES Journals	2
Table 2 An example of average citation number computation	14
Table 3 Citation Analysis for Paper processed by Different Associate Editors in TIE	25
Table 4 Citation Analysis for Papers processed by Different Associate Editors in TII	28
Table 5 Citation Analysis for Paper processed by Different EICs in TIE (Grouped by years) ..	30
Table 6 Citation Analysis for Paper processed by Different EICs in TIE (Grouped by EICs) ..	30
Table 7 Citation Analysis for Papers processed by Different EICs in TII (Grouped by years) .	31
Table 8 Citation Analysis for Papers processed by Different EICs in TII (Grouped by EICs) ..	31
Table 9 Citation Analysis for Special Section Papers Published in TII	32
Table 10 Citation Analysis for Special Section Papers Published in TIE	33

List of Abbreviations

AE	Associate Editor
EIC	Editor in Chief
IF	Impact Factor
IES	Industrial Electronics Society
TIE	<i>IEEE Trans. on Industrial Electronics</i>
TII	<i>IEEE Trans. on Industrial Informatics</i>
IEM	<i>IEEE Industrial Electronics Magazine</i>

Chapter 1

Traditional measures of Journal Quality

There are various well-established metrics to evaluate journal quality based on the citations of papers published in this journal. In this section, several notable traditional measures of journal quality are reviewed and compared, and new insightful measures are proposed.

1.1 2-year Impact Factor

2-year Impact Factor, often abbreviated IF [3] , is probably the most popular measure of journal performance. It reflects the average citation number to articles published in a journal in the 2 preceding years. The higher IF a journal has, the more important and influential it is considered within its field. 2-year IF of a journal in a given year, for example 2011, can be calculated as follows:

A = number of citations of articles published in 2009 and 2010 during 2011.

B = the total number of articles published in 2009 and 2010 by that journal.

2011 impact factor=A/B.

Table 1 shows data for IF calculations for three IES (Industrial Electronics Society) journals: *IEEE Trans. on Industrial Electronics (TIE)*, *IEEE Trans. on Industrial Informatics (TII)*, and *IEEE Industrial Electronics Magazine (IEM)*.

Table 1 Impact factor calculations for IES Journals

	TIE	TII	IEM
Number of citations to 2008 papers	2121	62	31
Number of citations to 2009 papers	1220	48	29
Number of citations to 2008&2009 papers	3341	110	60
Number of papers published in 2008	454	28	15
Number of papers published in 2009	505	39	17
Number of papers published in 2008& 2009	959	67	32
IF Impact Factor	3.48	1.64	1.87

Similar to 2-year IF, other scales such as 5-year IF, JII (Journal Immediacy Index) are also used in some cases [4]. Obviously the 5 year IF changes much slower, so it is more difficult to predict trends. JII is the ratio of number of citations to number of papers in the current year, which can be used for fast prediction of trends. JII accounts for the incapability of IF to incorporate information of the current year publication. The following figure shows an example of JII for IES journals.

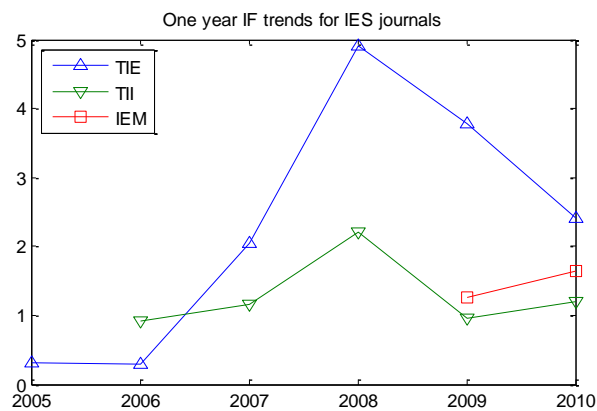


Fig. 1. One Year Impact Factors Trends for IES journals

Even though IF is simple to calculate and straightforward in meaning, the validity of IF has been much debated and more advanced measurements are proposed.

1.2 ES (Eigenfactor Score) and AIS (Article Influence Score)

More recently, another measure ES [5, 6] was developed to rate a scientific journal according to its citations, with citations from highly ranked journals weighted more than those from poorly ranked journals. ES is considered more representative and robust than IF, which counts purely the citation total without differentiating the significance of these citations. The computation of the ES requires an iterative approach because during computation journal rankings are changing, and this is affecting the score. However, the ES often gives misleading information because journals with a larger number of published papers automatically are receiving a higher ES. This problem was corrected by the introduction of AIS (Article Influence Score) [7] where the ES is normalized by number of papers published. ES and AIS are calculated by eigenfactor.org, and can be viewed freely there.

Chapter 2

Fundamentals of the Internet and Text Processing Robots

2.1 Introduction

Two kinds of robots, the Internet Robot [8, 9] and the text processing robot, are utilized in this thesis to perform complicated evaluations of the performance of journals editors, associate editors and special sections. The Internet Robot is a PERL program [2] which extracts and processes data from the IEEE website, and generates output files with a structured template. So what the Internet Robot basically does is to transform the representation of information on the web to a convenient form for the users. In this thesis, we will use the output of the Internet Robot to extract the publication time information of papers, which is a prerequisite of further analysis of timely performance of journals. The following figures 2 and 3 are a comparison of the original IEEE website and the processed output file from the Internet Robot.

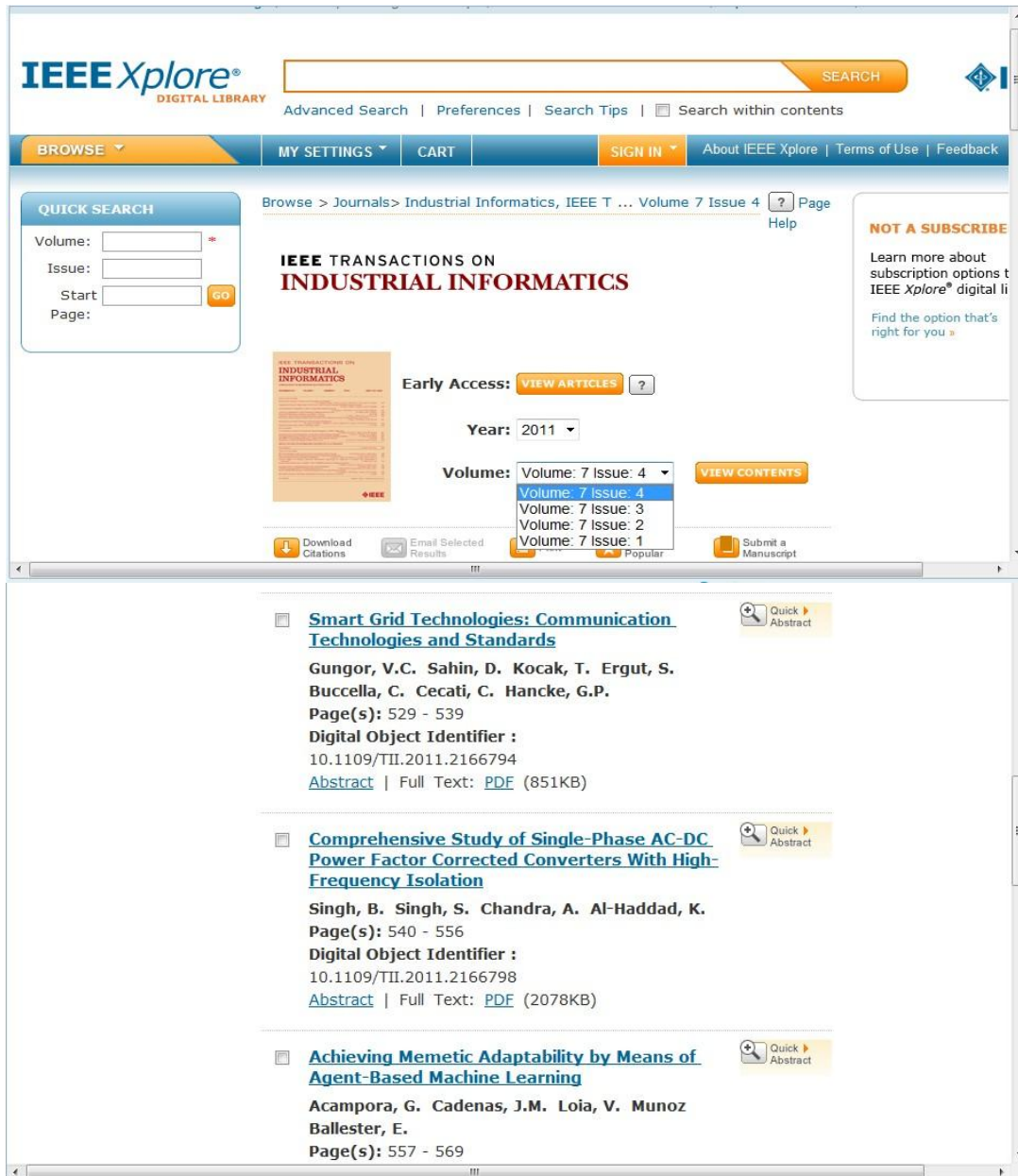


Fig. 2. Original IEEE Xplore webpage



IEEE Transactions on Industrial Informatics

Volume 7, Number 4, Nov 2011

[Access to the journal on IEEE XPLORE](#)

[IE Transactions Home Page](#)

- 7.4.1 V.C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, G.P. Hancke, "Smart Grid Technologies: Communication Technologies and Standards," *IEEE Trans. on Industrial Informatics*, vol. 7, no. 4, pp. , Nov 2011. [Abstract Link](#) [Full Text](#)
- 7.4.2 B. Singh, S. Singh, A. Chandra, K. Al-Haddad, "Comprehensive Study of Single-Phase AC-DC Power Factor Corrected Converters With High-Frequency Isolation," *IEEE Trans. on Industrial Informatics*, vol. 7, no. 4, pp. 540-556, Nov 2011. [Abstract Link](#) [Full Text](#)
- 7.4.3 G. Acampora, J.M. Cadenas, V. Loia, E. Munoz Ballester, "Achieving Memetic Adaptability by Means of Agent-Based Machine Learning," *IEEE Trans. on Industrial Informatics*, vol. 7, no. 4, pp. 557-569, Nov 2011. [Abstract Link](#) [Full Text](#)
- 7.4.4 M. Metzger, G. Polakow, "A Survey on Applications of Agent Technology in Industrial Process Control," *IEEE Trans. on Industrial Informatics*, vol. 7, no. 4, pp. 570-581, Nov 2011. [Abstract Link](#) [Full Text](#)

Fig. 3. Formatted output file from the Internet Robot for the Society web page

The screenshot shows the 'Harzing's Publish or Perish' software interface. The main window displays a 'Journal impact analysis' for 'Transactions on Industrial Informatics'. The search criteria include the journal title and a publication year range from 2010 to 2010. The results section shows a table with columns for Cites, Per year, Rank, Authors, Title, Year, Publication, and Publisher. The table lists 35 entries, each with a checkmark in the first column. The software interface also includes a sidebar with navigation options like 'Citation analysis', 'Program maintenance', and 'Help resources'. The status bar at the bottom indicates the version is 3.2.4150 and the date is Monday, December 19, 2011.

Fig 4. The "PublishOrPerish" software based on Google Scholar

The Text Processing Robot, also written in PERL, mainly serves to process, extract and combine useful information from different Excel files. These Excel files are obtained from mainly two sources. One major source is the MC (manuscriptCentral). [1] The MC system for paper collection and review can keep relatively good track of the submission and review process. There is information for each article, how many days has passed since the first decision, how long was the manuscript in revision with authors, and when the final decision was made. Users can log on the MC system and download such information in the format of excel files. The other important source of data related to citation number is from the “PublishOrPerish” software, as shown in Fig 4, which can also generate Excel files recording the citation number, title, authors and more information of papers published on a particular journal in a certain year. As we can see, the Excel files obtained from the MC system and the “PublishOrPerish” software contain separate information we are interested in, and how to match the titles of papers to integrate all the useful information is no trivial task, considering the huge amount of data to be processed. In this context, the Text Processing Robot is developed to efficiently and accurately handle this task.

2.2 Perl scripting language

PERL stands for “Practical Extraction and Report Language”, which was created by Larry Wall in the mid-1980s to make report processing easier. Since then, continuous changes and revisions have been made to improve it. PERL is an efficient language related to string processing. Other than string processing, the PERL language is also a very efficient platform to develop software run over the internet [10, 11]: such as the internet SPICE [14] or online neural network trainer [12, 13]. These attempts were precursors of the recently grown trend of the

cloud computing. PERL can be also very useful for data mining [15, 16] and for development of internet robots.

One main feature of PERL is its well-known regular expression support, which is so powerful and versatile that it has actually set a new standard for the regular expressions and is now emulated in many other programs and languages. String matching, searching and replacing are made especially easy as to just one statement. Another very attracting feature of PERL is its huge resource of free modules which are written by many different contributors and can be found at *cpan.org*. The installation of modules can be managed by PPM (Perl Package Manager), and users can just use the command “ppm install *PackageName*” in the Command Line Prompt on Windows to download and install a package. In Perl codes, the use of modules requires as simply as only one declaration “*use ModuleName*” at the beginning of the Perl code.

To more efficiently process Excel files, a specialized package targeted at handling Excel files “*Spreadsheet::ParseExcel::SaveParser*” is used in this thesis. There are a variety of functions available in this package to perform almost all the basic read/write tasks, such as opening a file, getting row/column range, reading/writing a cell, saving a file, and etc. The following code segment is given as an example:

```
use Spreadsheet::ParseExcel;
use Spreadsheet::ParseExcel::SaveParser;

$parser= Spreadsheet::ParseExcel::SaveParser->new();
$test=$parser->Parse('test.xls');
if ( !defined $test) {
    die $parser->error(), ".\n";
}
$worksheet1=$test->worksheet(1);
$row=0;
$column=0;
$cell=$worksheet1->get_cell($row,$column);
```



```
$cell_content=$cell->unformatted();  
$worksheet1->AddCell($row, $column+1, $cell_content);  
$test->SaveAs('newfile.xls');
```

In this example code, the “*test.xls*” is first read in by the program in line 4. Lines 5-7 are aimed to check errors in the file opening process, if the file is not opened correctly, then the program will abort. In line 8 the second worksheet is selected by calling the function *worksheet()*, with the index of the worksheet as input. (Note that the index of worksheets starts from 0 instead of 1). Next the cell A1 from this worksheet is read by calling the function *get_cell()*, specifying the row and column number of the cell as the two inputs, also note here that the row and column numbers start from 0. The value read from A1 is stored in a variable *\$cell_content*, who is then written to cell A2 by the function *AddCell()*. Finally, the modified excel file is saved in a file named “*newfile.xls*”. Through this demonstration, we can see the convenience and power of using packages. We don't bother to know the internal mechanics of excel files, but only need to manipulate the interface APIs provided by the corresponding package.

Chapter 3.

Evaluation of the performance of the Editorial Boards

3.1 Innovative measures of journal quality (EIC, AE, SS)

As shown in Chapter 1, traditional measures only can evaluate a journal's overall performance, but if we want to quantify specifically one editor's contribution to the journal, then new approaches must be proposed. In this section, we are going to present several innovative measures of journal quality, more specifically, the performance of EIC (Editor in Chief), AE (Associative Editor), and SS (Special Sections). There are actually two new kinds of evaluation methods studied in this thesis, based on citation and time respectively. The next two subsections 3.2 and 3.3 are going to explain in details the meaning and process of conducting these two kinds of evaluation.

3.2 Citation Based Evaluation

For citation based evaluation, we want to obtain the data reflecting how well the papers are cited that are selected by a certain EIC/AE, or in a SS with a particular topic. This kind of information will help us evaluate the insight and judgement of EICs and AEs, or how interesting and impactful is a topic for SS.

3.2.1 Evaluation of Editorial Boards

There is an indirect measure of Editor in Chief or Associate Editor performance by analyzing the acceptance rate for each EIC/AE. This information can be extracted from MC data, but the results could be misleading. For example, one AE may receive only very good manuscripts so his acceptance rate is very high, and another AE may receive for processing lower quality manuscripts, so naturally his acceptance rate would be low. Therefore the acceptance rate may not be the only measure to evaluate performance of AEs.

The more objective measure of EIC/AE quality work would be to link papers which she/he has accepted to the citations of these papers. In other words, apply the same measure which is being used to evaluate journal ranking. Unfortunately this information is not easily accessible. Part of the information about who has processed the manuscript is in the MC database, and other information about citations of manuscripts can be found in Google Scholar, "Publish or Perish", or in the data generated by Thomson Reuters. It was a challenge to extract and to combine this information.

To conduct citation based evaluation for Editorial Boards, first the citation information need to be combined with the editor information for every paper. To better illustrate how to integrate data from two excel files, the following Figure 5 and 6 show the raw data from "TII_citation.xls" and "TII_ManuscriptReceived.xls" respectively, and Figure 7 shows the combined data. In this thesis, the integrated data is directly saved in "Journal_citation.xls".

	A	B	C	D	E	F	G
1	Cites	Authors	Title	Year	ArticleURL	CitesURL	Author Institution
2	34	G Acciani,	Application of neural network	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Politecnico di Bari
3	31	FL Lian, JH	Network architecture and cor	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	National Taiwan Un
4	22	J Ferreira,	Combining operational flexibil	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Faculdade de Ciênc
5	70	AP Kaloge	Vertical integration of enterpr	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto
6	18	M Barranc	An active star topology for im	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto
7	16	G Cena...	On the properties of the flexit	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto
8	85	JLM Lastr	Semantic web services in fac	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Tampere University
9	48	T Skeie, S	Timeliness of real-time IP coi	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Sannic
10	10	Q Morante	Pervasive grid for large-scale	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Sannic
11	3	C Ryan, D	Interactive consistency on a t	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Limeric

Fig. 5. Fragment of the raw “TII_citation.xls” from *PublishOrPerish*

	Manuscript Title	Manuscript Type	Author Names	Submission Date	Decision Date	Decision	Author Institution	EIC Full Name	Editor Full Name
133	On the Use of Wireless Networks at Low Level of Factory Automation Systems	Regular Paper	Vituri, Stefano; De Pellegrini, Francesco;	Jan 09, 2006 11:39:10 AM	Feb 19, 2006 1:49:36 PM	Accept	IEIIT_CNR University of Padova	Zurawski, Richard	Colombo, Armando
134	An Internet-Based Distributed Manufacturing System Utilizing a Recurrent Network Net Controller for	Regular Paper	Lin, Wilfred	Jan 13, 2006 11:31:23 PM	Mar 29, 2006 9:29:40 AM	Major Revision	Hong Kong Polytechnic University	Zurawski, Richard	Kasabov, Nikola
135	Applying and Adapting the IEC 61346 Standard to Industrial Automation	Regular Paper	García, Rodrigo; Galle, Esther	Jan 26, 2006 10:02:09 AM	Mar 22, 2006 12:31:50 PM	Accept	EPFL	Zurawski, Richard	Martinez Lastra, Jose L.
136	Time-Triggered Smart Transducer	Regular Paper	Elmenreich, Wilfried	Jan 09, 2006 1:47:28 PM	Mar 07, 2006 4:59:39 PM	Accept	Vienna University of	Zurawski, Richard	Yu, Xinghuo
137	An active star topology for improving fault confinement in CAN networks	2010 SS paper	Barranco, Manuel; Proenza, Julián;	Jan 23, 2006 10:38:54 AM	Mar 17, 2006 2:02:01 PM	Minor Revision	University of Porto	Zurawski, Richard	Sauter, Thilo
138	On The Properties of The Flexible Time Division Multiple Access Technique	2010 SS paper	Cena, Gianluca; Valenzano, Adriano	Jan 19, 2006 4:16:29 PM	Mar 17, 2006 2:03:00 PM	Minor Revision	University of Porto	Zurawski, Richard	Sauter, Thilo
139	Combining Operational Flexibility and Dependability in FTT-CAN	2010 SS paper	Ferreira, Joaquim; Almeida, Luis; Fonseca, Jose;	Jan 29, 2006 5:08:28 AM	Mar 17, 2006 1:35:45 PM	Minor Revision	Faculdade de Ciências da Universidade de Lisboa	Zurawski, Richard	Sauter, Thilo
140									

Fig. 6. Fragment of “TII_ManuscriptReceived.xls” from *Manuscript Center*

	A	B	C	D	E	F	G	H	I
1	Cites	Authors	Title	Year	ArticleURL	CitesURL	Author Institution	EIC Full Name	Editor Full Name
2	34	G Acciani,	Application of neural network	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Politecnico di Bari	Zurawski, Richar	Vaccaro, Alfredo
3	31	FL Lian, JH	Network architecture and cor	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	National Taiwan Un	Zurawski, Richard	
4	22	J Ferreira,	Combining operational flexibil	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Faculdade de Ciênc	Zurawski, Richar	Sauter, Thilo
5	70	AP Kaloge	Vertical integration of enterpr	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto	Zurawski, Richar	Sauter, Thilo
6	18	M Barranc	An active star topology for im	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto	Zurawski, Richar	Sauter, Thilo
7	16	G Cena...	On the properties of the flexit	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Porto	Zurawski, Richar	Sauter, Thilo
8	85	JLM Lastr	Semantic web services in fac	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	Tampere University	Zurawski, Richar	Colombo, Armando
9	48	T Skeie, S	Timeliness of real-time IP coi	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Sannic	Zurawski, Richar	Naedele, Martin
10	10	Q Morante	Pervasive grid for large-scale	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Sannic	Zurawski, Richar	Gomes, Luis
11	3	C Ryan, D	Interactive consistency on a t	2006	http://ieeexplore.ieee.org/xpl/articleDetail.jsp?arnumber=4011111	http://scholar.google.com/scholar?hl=pt&btnG=Search	University of Limeric	Zurawski, Richar	Naedele, Martin

Fig. 7. Combined data with both Editor Info (Column H, I) from *Manuscript Center* and Citation Info (Column A) from *PublishOrPerish*

The matching process is based on paper title, but note that the same paper title may take different formats in the two excel files, such as cases and spacing. Also, some paper titles contain non alphabetic characters which cannot be recognized and used in PERL regular expression. Therefore, it is necessary to filter out those symbols and change titles' format to a consistent one before doing any comparison. The following sub routine is written to achieve this goal.

```
sub match
  {$string1=$_[0];
   $string2=$_[1];
   $string1 =~ s/(\W+)/ /;
   $string1 =~ s/(\W+)$//;
   $string2 =~ s/(\W+)/ /;
   $string2 =~ s/(\W+)$//;
   if (lc($string1) eq lc($string2)){
     return 1;}
   else{return 0;}
 }
```

In the above code snippet, two strings are passed to the sub routine as arguments, and their values are assigned to two local variables *\$string1* and *\$string2* in the first two lines. The next four lines are using PERL regular expression to search the non character symbols in the two strings and replace them with a single space. “\W” is one of the mega characters in PERL syntax, which refers to all the non alphabetic characters. In the forth line, the “\$” sign following (\W+) means matching at the end of the string, we are trying to eliminate any non word characters at the end of the string in this line. The “if” conditional statement compares the lower case of both the two strings, so the title matching process is case insensitive. Eventually, the function will return boolean value 1 if the two processed strings are the same, otherwise it will return 0.

After the integrated data is generated as shown in Fig 6, the average citation number for a certain EIC or AE can then be calculated. It is worth mentioning, that the meaning of ‘average’ is

twofold here, the obvious aspect is the average over number of papers processed by the same EIC/AE. The second aspect is less explicit, it refers to the average citation number over publication time for each paper, which need to be preprocessed before computing average over number of papers. The time unit used for time averaged citation number computation in this thesis is a quarter of a year. For example, assume Table 2 is a summary of all the papers “Editor 1” has selected for publication in TII, the next paragraph will show how to calculate the average citation number for “Editor 1”.

Table 2. An example of average citation number computation

Editor	Paper Title	Citation	Publication Date	Current Date	Time Averaged Citation (per quarter year)
Editor1	Paper 1	33	Feb 10, 2010	Dec 19, 2011	33/8
Editor1	Paper 2	14	May 10, 2011	Dec 19, 2011	14/3
Editor1	Paper 3	27	Nov 10, 2010	Dec 19, 2011	27/5

The last column in the above table is the time averaged citation number for each paper, in the unit of “citation number per quarter year”. For Paper 1, the time period between publication date and current date is 22 months and 9 days, which would be counted as 8 quarters, so its time averaged citation number would be 33/8. Using the same logic we can compute the time averaged citation number for every paper. With such information at hand, the final average citation number for this editor can be calculated as the average of the last column.

However, a question arises here: How do we get the publication date for each paper? This information is neither in “Journal_Citation.xls” or “Journal_ManuscriptReceived.xls”. As stated in Chapter 2, we will refer to the output from the Internet Robot to obtain the publication issue numbers for papers, which is also done by automatic title matching method as mentioned above.

	A	B	C	D	G	H	M	N	P
1	Cites	Authors	Title	Year	ArticleURL	CitesURL	EIC Full Name	Editor Full Name	Issue
2	2	S Fischme	Time-aware	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Buttazzo, Giorgio	4
3	1	MD Ma, D	Fault detect	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Ho, Weng Khuen	1
4	6	Y Zhang, F	Decentraliz	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Moyne, James	1
5	4	CA Boano	The impac	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Lu, Chenyang	3
6	1	G Cengic	On Formal	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Campos, Javier	2
7	2	H Zeng, M	Using Stat	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Cena, Gianluca	4
8	0	A Rowe, K	Rate-Harm	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Chen, Jian-Jia	3
9	5	D Cancila	Toward coi	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Hansson, Hans	2
10	0	Y Zheng, F	Inspiring In	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Proth, Jean-Marie	2
11	1	DP Filev, F	An Industri	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Lee, Jay	4
12	1	G Wang, M	Optimal S	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	simonot-lion, francoise	4
13	2	D Wang, J	Data-Drive	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Ho, Weng Khuen	1
14	7	MH Kim, S	Kalman pr	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Ho, Weng Khuen	1
15	1	J Choi, H	(A process	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Hu, Xiaobo Sharon	3
16	0	MA Gora	A Flexible	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Carloni, Luca	4
17	4	G Gaderer	Improving f	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Fonseca, Jose	2
18	0	L Wang...	An efficien	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Hu, Xiaobo Sharon	3
19	2	M Bertogn	Limited pre	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Nolte, Thomas	4
20	1	BB Frague	Address-In	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Buttazzo, Giorgio	4
21	2	Y Wu, G	EParameter	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Kuo, Tei-Wei	4
22	4	G Gamba	Retransmi	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Proenza, Julián	3
23	0	H Yan, W	Multi-Zone	2010	http://ieeexplore.org/study/cite	http://scholar.google.com/citations?view_op=view_citation	Zurawski, Richard	Dunia, Ricardo	2

Fig. 8 Snapshot of “TII_citation.xls” with data from 3 sources: Citations (Column A) from *PublishOrPerish*, Editors Information (Column M, N) from *Manuscript Center*, and issue number (Column P) from *the output webpages of the Internet Robot*.

Fig 8 is an example of “TII_citation.xls” after getting the publication issue number for every paper. Having the issue number information for each paper, then we are able to infer the publication date information for different journals. The journal TII has 4 issues per year, and they are published in February, May, August, November respectively; TIE has 12 issues per year and they are published every month. In this thesis, we assume the exact date of publication for

every issue falls on the 10th of the publication month. So if a paper is published in TII in Issue 2, 2011, then its publication date is assumed to be May 10th, 2011.

3.2.2 Evaluation of Special Sections

The principle to perform citation based evaluation for Special Sections is basically the same as that for Editorial Boards, however, the procedure involves more efforts because there is no direct way to obtain the paper type information. In other words, there is no easy way of identifying which paper belongs to which SS, or whether it is a regular paper. To the best of my knowledge, the only reliable source of such data is from the IEEE Xplore website. For every issue published, there is a link “Table of Content” to a PDF file, which states the type of every paper in this issue, regular paper or SS paper, and the title of the SS. Fig 9 and 10 is an example of such a link and its pointed PDF file.

Browse > Journals> Industrial Informatics, IEEE T ... Volume 7 Issue 4 Page Help

QUICK SEARCH

Volume: *
 Issue:
 Start Page: **GO**

IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS

Early Access: **VIEW ARTICLES** ?

Year: 2011 ▼

Volume: Volume: 7 Issue: 4 ▼ **VIEW CONTENTS**

Download Citations | Email Selected Results | Print | View Popular | Submit a Manuscript

Select All | Deselect All

Table of Contents

Page(s): C1
 Digital Object Identifier : 10.1109/TII.2011.2173011
 Full Text: [PDF](#) (41KB)

NOT A SUBSCRIBER?
 Learn more about subscription options to the IEEE Xplore® digital library.
 Find the option that's right for you »

Fig. 9. IEEE Xplore Webpage of TII Volume 7, Issue 4. The first entry of its contents is “Table of Contents” as shown at the bottom of this figure.

IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS

A PUBLICATION OF THE IEEE INDUSTRIAL ELECTRONICS SOCIETY

NOVEMBER 2011

VOLUME 7

NUMBER 4

ITIICH

(ISSN 1551-3203)

STATE-OF-THE-ART PAPERS

Smart Grid Technologies: Communication Technologies and Standards	529
..... V. C. Güngör, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati, and G. P. Hancke	
Comprehensive Study of Single-Phase AC-DC Power Factor Corrected Converters With High-Frequency Isolation	540
..... B. Singh, S. Singh, A. Chandra, and K. Al-Haddad	
Achieving Memetic Adaptability by Means of Agent-Based Machine Learning	557
..... G. Acampora, J. M. Cadenas, V. Loia, and E. Muñoz Ballester	
A Survey on Applications of Agent Technology in Industrial Process Control	570
..... M. Metzger and G. Polaków	
Fractional Order Systems in Industrial Automation—A Survey	582
..... M. O. Efe	
Comparative Study of Derivative Free Optimization Algorithms	592
..... N. Pham, A. Malinowski, and T. Bartczak	
Power-Aware System Design of Wireless Sensor Networks: Power Estimation and Power Profiling Strategies (Invited Paper)	601
..... J. Haase, J. M. Molina, and D. Dietrich	
Managing Process Model Complexity Via Abstract Syntax Modifications	614
..... M. La Rosa, P. Wöhed, J. Mendling, A. H. M. ter Hofstede, H. A. Reijers, and W. M. P. van der Aalst	
Enterprise Systems: State-of-the-Art and Future Trends	630
..... L. D. Xu	

REGULAR PAPERS

A Comprehensive Solution for Deterministic Replay Debugging of SoftPLC Applications	641
..... H. Prühofer, R. Schatz, C. Wirth, and H. Mössenböck	
Expressing Coarse-Grain Dependencies Among Tasks in Shared Memory Programs	652
..... P. Larsen, S. Karlsson, and J. Madsen	
Uncertainty-Robust Design of Interval Type-2 Fuzzy Logic Controller for Delta Parallel Robot	661
..... O. Linda and M. Manic	
The GMPLS Controlled Optical Networks as Industry Communication Platform	671
..... J. Korniak	
Performance of a Real-Time EtherCAT Master Under Linux	679
..... M. Cereia, I. C. Bertolotti, and S. Scanzio	

SPECIAL SECTION ON INFORMATION TECHNOLOGY IN AUTOMATION

Guest Editorial	688
..... L. Lo Bello and G. Frey	

SPECIAL SECTION PAPERS

Real-Time Modeling for Direct Load Control in Cyber-Physical Power Systems	689
..... T. Facchinetti and M. L. Della Vedova	
How to Access Factory Floor Information Using Internet Technologies and Gateways	699
..... T. Sauter and M. Lobashov	
A Novel Standard for Footwear Industrial Machineries	713
..... G. Danese, S. Dulio, M. Giachero, F. Leporati, and N. Nazzicari	
Software Support for Building Automation Requirements Engineering—An Application of Semantic Web Technologies in Automation	723
..... S. Runde and A. Fay	
Semantic-Based Enhancement of ISO/IEC 14543-3 EIB/KNX Standard for Building Automation	731
..... M. Ruffa, F. Scioscia, E. Di Sciascio, and G. Loseto	
Functional Analysis of Manufacturing Execution System Distribution	740
..... A. Bratukhin and T. Sauter	
A Dual Programming Model for Distributed Real-Time Java	750
..... P. Basanta-Vál, M. García-Valls, and I. Estévez-Ayres	
Service-Oriented Infrastructure to Support the Deployment of Evolvable Production Systems	759
..... G. Cândido, A. W. Colombo, J. Barata, and F. Jammes	
IEC 61499 as Enabler of Distributed and Intelligent Automation: State-of-the-Art Review	768
..... V. Vyatkin	

2011 INDEX	Available online at http://ieeexplore.ieee.org
------------------	---



Fig. 10. “Table of Content” from TII Volume 7, Issue 4. From this page information about Special Sections are extracted, such as SS name, paper types, etc.

The paper type information is looked up in the “Table of Content” PDF files and added to the “Journal_citation.xls” excel files manually. This manual process is feasible due to the small number of papers falling in Special Sections. Fig 11 is a snapshot of the file “TII_citation.xls” after adding the paper type information. As far as now, it contains data from 4 sources: *PublishOrPerish*, *Manuscript Center*, *the output webpages of the Internet Robot*, *the IEEE Xplore Table of Content*.

	A	B	C	D	G	H	I	M	N	P
1	Cites	Authors	Title	Year	ArticleURL	CitesURL	Paper Type	EIC Full Name	Editor Full Name	Issue
2	2	S Fischme	Time-aware	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Real-Time Applications and Tools De	Zurawski, Richard	Buttazzo, Giorgio	4
3	1	MD Ma, D	Fault detect	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Industrial Control	Zurawski, Richard	Ho, Weng Khuen	1
4	6	Y Zhang, F	Decentraliz	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Industrial Control	Zurawski, Richard	Moyne, James	1
5	4	CA Boano,	The impac	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R2	Zurawski, Richard	Lu, Chenyang	3
6	1	G Cengic,	On Formal	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Formal Methods in Manufacturing	Zurawski, Richard	Campos, Javier	2
7	2	H Zeng, M	Using Stat	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R2	Zurawski, Richard	Cena, Gianluca	4
8	0	A Rowe, K	Rate-Harm	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Power-Aware Computing	Zurawski, Richard	Chen, Jian-Jia	3
9	5	D Cancila,	Toward col	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R	Zurawski, Richard	Hansson, Hans	2
10	0	Y Zheng, F	Inspiring In	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R	Zurawski, Richard	Proth, Jean-Marie	2
11	1	DP Filev, F	An Industri	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R2	Zurawski, Richard	Lee, Jay	4
12	1	G Wang, M	Optimal S	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R2	Zurawski, Richard	simonot-lion, francoise	4
13	2	D Wang, J	Data-Drive	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Industrial Control	Zurawski, Richard	Ho, Weng Khuen	1
14	7	MH Kim, S	Kalman pr	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Industrial Control	Zurawski, Richard	Ho, Weng Khuen	1
15	1	J Choi, H	(A process	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Power-Aware Computing	Zurawski, Richard	Hu, Xiaobo Sharon	3
16	0	MA Gora,	A Flexible	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R2	Zurawski, Richard	Carloni, Luca	4
17	4	G Gaderer	Improving f	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R	Zurawski, Richard	Fonseca, Jose	2
18	0	L Wang...	An efficien	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Power-Aware Computing	Zurawski, Richard	Hu, Xiaobo Sharon	3
19	2	M Bertogn	Limited pre	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Real-Time and (Networked) Embeddi	Zurawski, Richard	Nolte, Thomas	4
20	1	BB Frague	Address-In	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Real-Time Applications and Tools De	Zurawski, Richard	Buttazzo, Giorgio	4
21	2	Y Wu, G E	Parameter	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Real-Time Applications and Tools De	Zurawski, Richard	Kuo, Tei-Wei	4
22	4	G Gamba,	Retransmi	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Industrial Communication Systems	Zurawski, Richard	Proenza, Julián	3
23	0	H Yan, W	Multi-Zone	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	R	Zurawski, Richard	Dunia, Ricardo	2
24	0	D Shin, S	Energy-Op	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Power-Aware Computing	Zurawski, Richard	Mosse, Daniel	3
25	5	D Herrero-	Modeling c	2010	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5501111	http://scholar.google.com/citations?view_op=view_citation&hl=en&user=93131130000000000000&citation_for_view=93131130000000000000:5501111	SS on Formal Methods in Manufacturing	Zurawski, Richard	Campos, Javier	2

Fig. 11. Snapshot of “TII_citation.xls” with data from 4 sources: Citations (Column A) from *PublishOrPerish*, Editors Information (Column M, N) from *Manuscript Center*, Issue Number (Column P) from *the output webpages of the Internet Robot*, and SS Paper Type information (Column I) from *the IEEE Xplore Table of Content*.

The computation of average citation number for every SS is the same as that for Editorial Board. But note here, the publication date for papers within the same SS is the same, so the computation process can be simplified a little.

3.3 Time based Evaluation

For time based evaluation, we want to measure the responsiveness of journal review process. In this thesis, three timing factors are computed and analyzed: the average processing time from paper submission to first decision, from paper submission to final decision, and from acceptance to publication. It may seem natural to think that shorter review time indicates higher efficiency of the Editorial Boards. However, the fact is more complicated, considering large journals will attract more paper submissions thus consuming more review time; some writers may take more time to revise the papers than others thus prolonging their papers review time; Journals with sufficient high-quality papers supply may have a large pool of already accepted papers waiting to be published, so their acceptance to publication time will be greater than other journals. In all, we have to bear these factors in mind when evaluating journals according to their time performance.

3.3.1 Extract Submission Date, First Decision Date and Acceptance Date

Time based evaluation requires paper title matching within a single Excel file produced by the MC database system, with the name format to be “Journal_ManuscriptReceived.xls”. An example is shown in Fig 6, which is a fragment of “TII_ManuscriptReceived.xls”. From Fig 6 we can see, the “Decision” field for a paper may take different values of “Accepted”, “Major Revision”, “Minor Revision”, “Rejected”. That’s because a paper may go through several

revisions before being finally accepted, so the same paper may have several entries in the excel file.

In order to get a paper's submission date and first decision date, we need to scan from the top of the file "journal_ManuscriptReceived.xls" until the first entry of the paper is found. The submission date and decision date fields of this entry are the information we need. But, because we are not sure whether the paper is accepted or not during its first decision, so the acceptance date of the paper need to be further determined. If the decision state in the first entry of the paper is "accepted", which means the paper was accepted the first time it was submitted without any revision, then its acceptance date is simply the value of the "Decision Date" field; otherwise, the scan has to be continued until the entry of the paper with "acceptance" decision is found.

However, to make things more complicated, there exists data inconsistency in the MC database; ideally a paper from "Journal_ciation.xls" is already accepted and published, however, there may not exist an entry in "Journal_ManuscriptReceived.xls" indicating it is accepted. In this case, the "decision date" field of the last entry of the paper is used to approximate the acceptance date information. Fig 12 is the flow chart of the algorithm to find a paper's acceptance date.

As for the publication date, it is already discussed and resolved in 3.2.

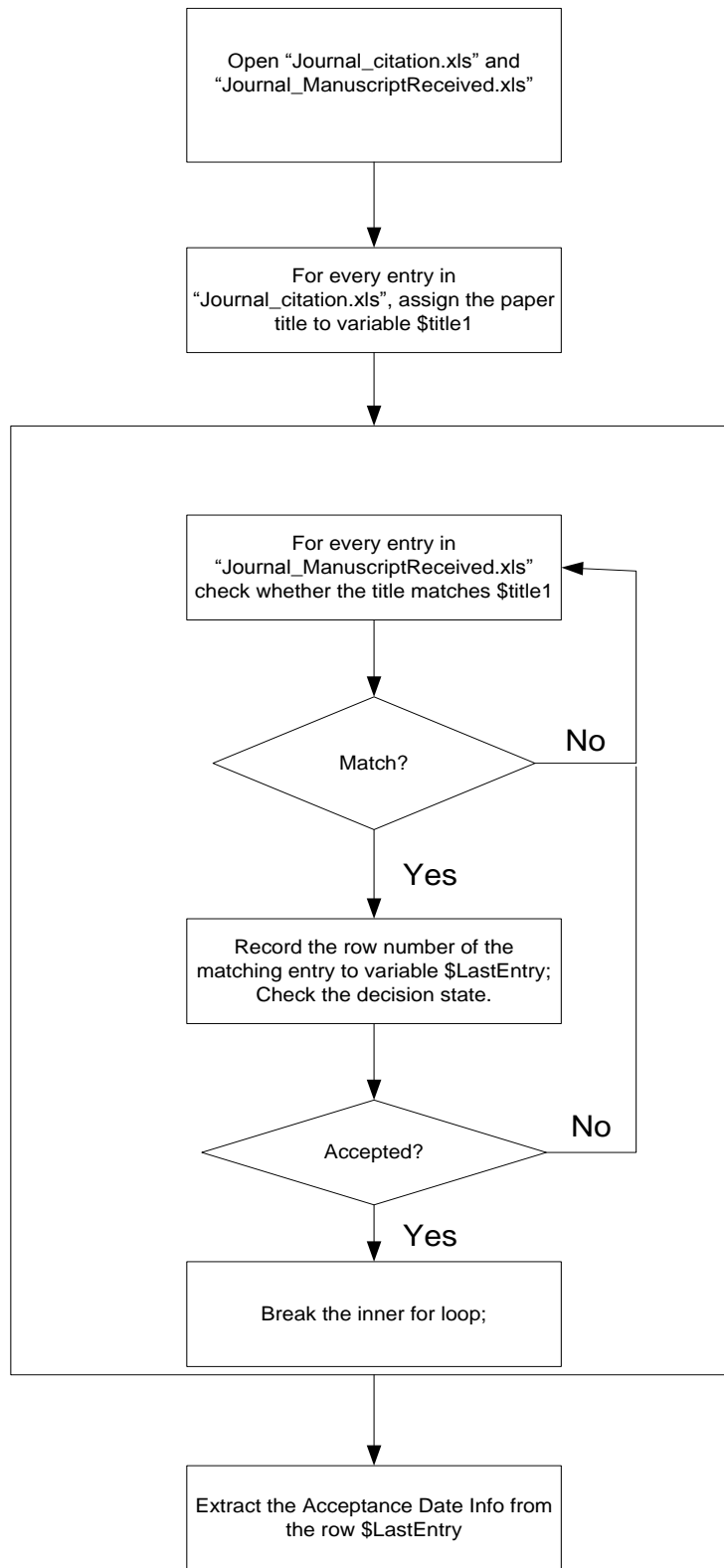


Fig.12. Flow chart of the algorithm to extract the “Acceptance Date” for a paper

3.3.2 Computation of Passed Days Between Two Dates

After the data of Submission Date, First Decision Date and Acceptance Date for papers are obtained and saved in the file “Journal_citation.xls”, we are ready to compute the collapsed days between them for every paper. A sub routine *get_days* is written to compute how many days have passed between two dates, with the input format to be “Month Date, Year”. This sub routine takes advantage of the *hash* data structure of PERL to maintain the numeric index of every month according to their name abbreviations. And, an array is used to store the length of every month from Jan to Dec. The syntax of declaring and initializing the hash and array is as following:

```
my @month_length=(31,28,31,30,31,30,31,31,30,31,30,31);  
my %month_order=(Jan=>0, Feb=>1, Mar=>2, Apr=>3, May=>4, Jun=>5, Jul=>6,  
Aug=>7, Sep=>8, Oct=>9, Nov=>10, Dec=>11);
```

Using the hash is very convenient, we can simply use the syntax `$month_order{Month Abbreviation}` to get the index of that month. For example, `$month_order{Jan}` will give the value of 0, which can be further used to index the array and get the length of Jan--31 days.

This sub first analyzes the two input dates to get the starting month, date, year and ending month, date, year. Then the total months between the two date is computed. For example, if the two inputs are “Jan 07, 2010” and “Mar 18, 2011”, then there are 14 months between them. And the total days between the two dates are computed as

$$Total\ days = \sum (lengths\ of\ months\ in\ between) + end\ day - start\ day$$

After we have got the data of review time for every paper, then the average data can be easily computed for every journal.

3.4 Results

The above sections introduced the concept, meaning and procedure to perform several innovative journal evaluations. In this section, the results will be shown in figures and tables.

3.4.1 Citation Performance of the Editorial Boards

Tables 3 and 4 present normalized citations of papers processed by Associate Editors in TIE and TII. In Tables 3-4, column 1 shows a random number assigned to each AE instead of their real names because of privacy issues; column 2 shows total number of papers selected for publication by a given AE; column 3 lists total citations of the papers; column 4 presents the sum of average citations over time (cites / per quarter year) of these papers; and the last column shows the average citations over time and over paper numbers (cites / per paper and per year).

Tables 5-8 present citations analysis for EICs in TIE and TII. Tables 5 and 7 are grouped by years, citation data for EICs in different years are listed in the tables. Except for the first column being “Year”, the other columns fall in the same sequence as in Tables 3-4. From Tables 5 and 7, a trend is shown that old publications tend to have higher average citations than new publications, which is especially obvious from EIC #1’s yearly average citations in Table 7. Multiple reasons may contribute to this phenomena, including authors’ preference to cite well-known papers rather than new papers, easy access to well-cited papers on Google Scholar, etc. Tables 6 and 8 take out the “*year*” column, and show aggregate citation data for EICs across all the years from 2006 to 2011 for TIE and TII.

Tables 3, 5 and 6 present data for the AEs and EICs of the IEEE Trans. on Industrial Electronics, while Tables 4, 7 and 8 present data for the AEs and EICs of the IEEE Trans. on

Industrial Informatics. Because the TIE is about 7 times as large as TII, each EIC/AE is processing a larger number of papers than their partners in TII. Also, TIE has a larger Impact Factor and a larger number of EICs/AEs which can be ranked.

The information provided in Tables 3-8 is definitely a better measure of the Editorial Boards performance than commonly used measures such as the acceptance rate, review time, etc. Of course the review time is also important, but it is not as important as a proper evaluation of chances of manuscript citations.

Table 3 Citation Analysis for Paper processed by Different Associate Editors in TIE

AE num	# of Papers	# of cit.	Citations /quarter	Citations /pap/year
AE# 050	1	12	6.00	24.00
AE# 009	24	1709	104.21	17.37
AE# 029	8	378	31.78	15.89
AE# 054	34	1701	121.33	14.27
AE# 024	8	309	28.05	14.02
AE# 001	11	312	36.77	13.37
AE# 037	11	433	34.12	12.41
AE# 041	5	205	15.32	12.26
AE# 088	28	762	85.38	12.20
AE# 031	1	44	2.93	11.73
AE# 043	5	152	14.51	11.61
AE# 076	16	470	45.22	11.30
AE# 008	7	224	19.35	11.06
AE# 063	27	541	72.60	10.76
AE# 086	7	121	18.63	10.65
AE# 094	2	89	5.24	10.47
AE# 061	21	423	53.77	10.24
AE# 010	15	292	37.63	10.04
AE# 057	7	129	17.46	9.98
AE# 051	25	567	61.85	9.90
AE# 044	7	97	17.25	9.85

AE# 052	4	95	9.73	9.73
AE# 002	16	361	38.30	9.58
AE# 012	20	680	45.33	9.07
AE# 102	2	17	4.25	8.50
AE# 084	19	314	39.93	8.41
AE# 046	8	137	16.71	8.36
AE# 073	11	285	22.79	8.29
AE# 064	9	139	18.50	8.22
AE# 069	2	21	4.00	8.00
AE# 042	2	68	3.87	7.73
AE# 027	31	627	59.67	7.70
AE# 090	9	204	17.18	7.64
AE# 096	11	118	20.40	7.42
AE# 055	14	228	25.94	7.41
AE# 058	5	45	9.20	7.36
AE# 070	13	254	23.68	7.29
AE# 095	16	279	28.71	7.18
AE# 062	23	318	40.89	7.11
AE# 033	20	318	35.10	7.02
AE# 066	17	284	29.67	6.98
AE# 038	31	445	53.31	6.88
AE# 087	13	283	22.13	6.81
AE# 018	3	65	5.07	6.76
AE# 078	15	303	25.24	6.73
AE# 007	13	288	21.85	6.72
AE# 019	13	280	21.76	6.70
AE# 003	7	145	11.66	6.67
AE# 098	13	269	21.48	6.61
AE# 059	12	166	19.81	6.60
AE# 015	3	61	4.81	6.42
AE# 083	3	21	4.77	6.36
AE# 077	7	117	10.95	6.26
AE# 092	4	88	6.20	6.20
AE# 099	7	151	10.79	6.16
AE# 013	14	158	21.37	6.11
AE# 040	5	74	7.55	6.04
AE# 045	4	12	6.00	6.00
AE# 049	1	6	1.50	6.00
AE# 075	4	61	6.00	6.00
AE# 103	2	8	3.00	6.00
AE# 060	15	183	22.22	5.92
AE# 080	5	131	7.36	5.89
AE# 004	15	221	22.05	5.88

AE# 100	10	77	14.62	5.85
AE# 026	5	181	7.22	5.78
AE# 039	18	290	25.63	5.70
AE# 035	12	164	16.99	5.66
AE# 068	7	59	9.69	5.54
AE# 020	8	120	11.00	5.50
AE# 056	11	123	14.77	5.37
AE# 022	1	20	1.33	5.33
AE# 017	4	32	5.26	5.26
AE# 085	3	55	3.88	5.17
AE# 005	14	159	17.50	5.00
AE# 011	3	11	3.58	4.78
AE# 053	22	260	26.24	4.77
AE# 081	1	19	1.19	4.75
AE# 067	1	20	1.18	4.71
AE# 093	2	5	2.33	4.67
AE# 079	2	36	2.30	4.60
AE# 091	4	56	4.42	4.42
AE# 048	3	43	3.00	3.99
AE# 089	3	20	2.93	3.91
AE# 071	3	41	2.93	3.90
AE# 032	5	13	4.83	3.87
AE# 097	10	113	8.80	3.52
AE# 030	4	14	3.33	3.33
AE# 028	5	32	4.00	3.20
AE# 047	6	45	4.52	3.02
AE# 072	1	10	0.75	3.00
AE# 074	3	22	2.21	2.95
AE# 006	1	5	0.67	2.67
AE# 016	3	47	2.00	2.67
AE# 034	1	8	0.67	2.67
AE# 025	4	12	2.03	2.03
AE# 023	1	2	0.50	2.00
AE# 101	2	5	1.00	2.00
AE# 014	11	13	4.83	1.76
AE# 065	1	5	0.42	1.67
AE# 036	3	5	0.82	1.10
AE# 021	4	4	1.00	1.00
AE# 082	1	0	0.00	0.00

Table 4 Citation Analysis for Papers processed by Different Associate Editors in TII

AE num	# of Papers	# of cit.	Citations /quarter	Citations /pap/year
AE #07	3	253	14.32	19.09
AE #50	6	168	16.84	11.23
AE #15	3	101	6.86	9.14
AE #55	5	143	10.69	8.55
AE #32	1	27	1.80	7.20
AE #58	6	201	10.64	7.04
AE #31	2	55	3.44	6.88
AE #43	7	164	10.21	5.83
AE #49	2	37	2.33	4.65
AE #53	2	32	2.29	4.57
AE #37	6	45	6.81	4.54
AE #35	3	58	3.26	4.34
AE #10	1	16	1.07	4.27
AE #01	1	3	1.00	4.00
AE #27	1	4	1.00	4.00
AE #30	1	2	1.00	4.00
AE #48	1	2	1.00	4.00
AE #26	1	14	0.93	3.73
AE #52	2	28	1.87	3.73
AE #57	2	5	1.75	3.50
AE #05	11	75	9.06	3.30
AE #23	4	23	3.28	3.28
AE #18	13	92	10.42	3.21
AE #14	1	4	0.80	3.20
AE #25	1	7	0.78	3.11
AE #24	2	8	1.50	3.00
AE #02	5	24	3.69	2.95
AE #03	8	41	5.62	2.81
AE #34	2	7	1.40	2.80
AE #36	10	40	6.86	2.74
AE #28	1	10	0.67	2.67
AE #21	9	42	5.58	2.48
AE #16	2	6	1.20	2.40
AE #44	1	5	0.56	2.22
AE #41	2	6	1.02	2.04
AE #56	4	31	2.03	2.03
AE #09	1	5	0.45	1.82
AE #45	8	22	3.60	1.80

AE #39	4	7	1.75	1.75
AE #42	1	6	0.43	1.71
AE #04	2	2	0.67	1.33
AE #46	1	3	0.33	1.33
AE #54	3	2	1.00	1.33
AE #22	1	3	0.30	1.20
AE #33	3	2	0.50	0.67
AE #51	2	2	0.25	0.50
AE #38	2	3	0.20	0.40
AE #19	3	1	0.25	0.33
AE #06	1	0	0.00	0.00
AE #08	1	0	0.00	0.00
AE #11	1	0	0.00	0.00
AE #12	1	0	0.00	0.00
AE #13	1	0	0.00	0.00
AE #17	1	0	0.00	0.00
AE #20	1	0	0.00	0.00
AE #29	1	0	0.00	0.00
AE #40	1	0	0.00	0.00
AE #47	1	0	0.00	0.00

Table 5 Citation Analysis for Paper processed by Different EICs in TIE
(Grouped by years)

year	EIC num	# of paper	citation	cites/quarter	cites/paper/year
2006	EIC #1	70	3641	219.2	12.52
2007	EIC #1	200	6109	433.7	8.67
2008	EIC #1	299	6573	575.04	7.69
2009	EIC #2	1	9	1.285714	5.14
2009	EIC #1	379	4742	626.47	6.61
2010	EIC #3	1	2	0.33	1.32
2010	EIC #4	93	343	86.5	3.72
2010	EIC #5	7	34	4	2.28
2010	EIC #2	3	19	3.06	4.08
2010	EIC #1	207	1648	366.05	7.07
2011	EIC #3	23	50	28	4.86
2011	EIC #4	75	198	81.5	4.34
2011	EIC #5	22	32	13.5	2.45
2011	EIC #2	23	68	34	5.91
2011	EIC #6	7	11	5.5	3.14
2011	EIC #1	18	96	24	5.33

Table 6 Citation Analysis for Paper processed by Different EICs in TIE
(Grouped by EICs)

EIC num	# of paper	Citation	cites/quarter	cites/paper/year
EIC #1	1173	22809	2244.46	47.91
EIC #2	27	96	38.34	15.13
EIC #3	24	52	28.33	6.18
EIC #4	168	541	168	8.06
EIC #5	29	66	17.5	4.74
EIC #6	7	11	5.5	3.14

Table 7 Citation Analysis for Papers processed by Different EICs in TII
(Grouped by years)

year	EIC num	# of paper	citation	cites/quarter	cites/paper/year
2006	EIC #1	23	635	35.27	6.13
2007	EIC #1	17	316	21.9	5.15
2008	EIC #1	21	286	27.2	5.18
2009	EIC #1	35	241	32.62	3.72
2010	EIC #1	55	107	25.71	1.86
2011	EIC #1	11	5	2.5	0.90
2011	EIC #2	9	4	2	0.88

Table 8 Citation Analysis for Paper processed by Different EICs in TII
(Grouped by EICs)

EIC num	# of paper	Citation	cites/quarter	cites/paper/year
EIC #1	162	1590	145.2	22.97
EIC #2	9	4	2	0.88

3.4.2 Citation Analysis for Special Sections

There is also a significant citations difference depending on the topic of Special Sections. In the case of most Special Sections, citations are slightly higher than citations to regular papers. However there are some cases where citations to SS papers are significantly lower, and this may

provide a valuable feedback to the editorial board. Table 5 and 6 show the name, publication time and average citations for TII and TIE respectively.

TABLE 9 Citation Analysis for Special Section Papers Published in TII

SS Name	printed	citations /pap/year
SS on Wireless Technologies in Facto	2007/2	7.99
SS on Wireless Technologies in Facto	2007/3	6.99
SS on Factory Communication System	2005/3	5.79
SS on Communication in Automation	2006/2	4.87
SS on Real-Time Systems part 2	2009/1	3.96
SS on Real-Time and (Networked) En	2009/3	3.18
SS on Communication in Automation	2008/2	2.35
SS on In-Vehicle Embedded Systems	2009/4	2.11
SS on Industrial Control	2010/1	2.09
SS on Communication in Automation	2009/2	1.97
SS on Industrial Communication Syst	2010/3	1.88
SS on Real-Time Applications and To	2010/4	1.85
SS on Formal Methods in Manufactur	2010/2	1.70
SS on Real-Time and (Networked) En	2010/4	1.21
SS on Real-Time Systems Part 1	2008/4	1.03
SS on Power-Aware Computing	2010/3	0.68

Table 10 Citation Analysis for Special Section Papers Published in TIE

SS Name	printed	citations /pap/year
SS on Automotive	2008/6	23.64
SS on Digital Ecosystems and Cyber	2011/6	15.82
SS on Predictive Control of Power El	2009/6	11.74
SS on Photovoltaic Power Processin	2008/7	11.63
SS on Industrial Wireless Sensor Net	2009/10	10.77
SS on FPGA	2007/4	10.02
SS on Advances in Electrical Machin	2008/12	9.70
SS on FPGA	2008/4	9.60
SS on Voltage and Current Control o	2009/2	9.41
SS on Application and Control of Do	2009/10	9.20
SS on Multilevel Inverters	2010/8	9.03
SS on Active and Hybrid Filters to En	2009/8	8.27
SS on Efficient and reliable photovo	2009/11	6.84
SS on Renewable Energy Systems	2011/1	6.68
SS on Education	2007/6	6.45
SS on Energy Harvesting	2010/3	6.37
SS on Predictive Control of Power El	2008/12	6.25
SS on Hardawre-in-the-loop	2010/4	6.17
SS on Advances in Electrical Machin	2010/1	6.12
SS on Fuel cells power	2010/6	5.89
SS on Applications of Embedded Sys	2011/3	5.88
SS on Plug-in hybrid electric vehicle	2010/2	5.86
SS on Sliding Mode Control in Indust	2008/11	5.70
SS on Radio Frequency Identificatio	2009/7	5.23
SS on Thermal Issues in Electrical Ma	2008/10	5.15
SS on Education	2008/6	4.79
SS on Advances in Motion Control	2009/10	4.63
SS on Automotive Power and Energy	2010/3	4.59
SS on Education	2009/12	4.58
SS on Advances in Electrical Machin	2009/11	4.15
SS on Sliding Mode Control in Indust	2009/9	4.11
SS on Diagnostics of EMPED	2011/5	3.66
SS on Energy Storage System	2010/12	3.44
SS on Electronic Devices and System	2011/7	3.33
SS on Building Automation, Control	2010/11	3.17
SS on Switched Reluctance Machine	2010/9	2.80
SS on Wireless Technology	2010/5	2.55
SS on Advances in Microelectromec	2009/4	2.54
SS on Education	2010/10	2.42

3.4.3 Timely Performance of the Review Process

Fig 13 shows average time between manuscript submission and the first decision for TIE and TII. One may notice that this time in TIE was significantly shorter in 2008, and it is staying in the range of 10 to 11 weeks. In TII this time oscillates about 11 weeks. Fig 14 shows average time from submission to the final decisions. Fig 15 shows average time between acceptance and the publication and Fig 16 shows average times between submissions to the publication date. Figs 15 and 16 show a significant delay in publications in TIE because relatively large backlog of accepted papers. On the other hand in TII (see Fig 15) the time between acceptance and printing was below 50 days in 2008. This means that there were not enough accepted manuscripts to submit them on time for printing because IEEE usually needs final manuscripts about 90 days before publication date.

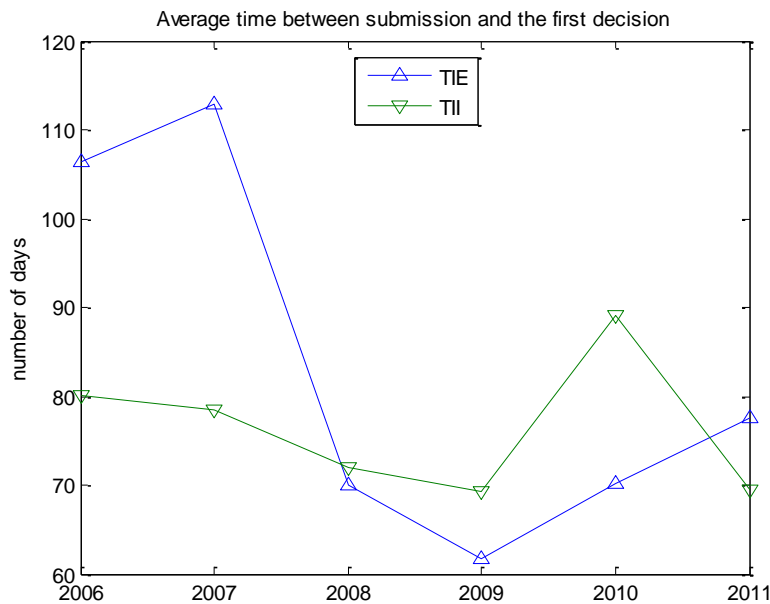


Fig. 13. Average time between submission and the first decision for TIE and TII.

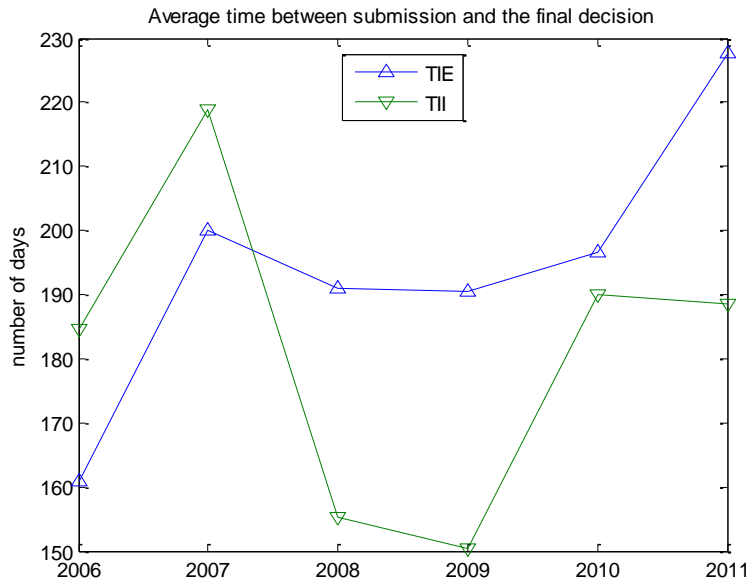


Fig. 14. Average time between submission and the final decision for TIE and TII.

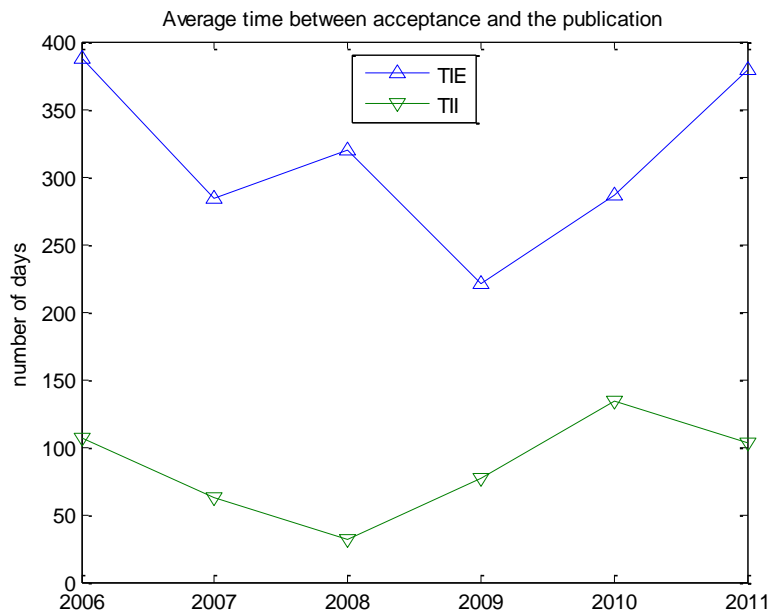


Fig. 15. Average time between acceptance and the publication for TIE and TII

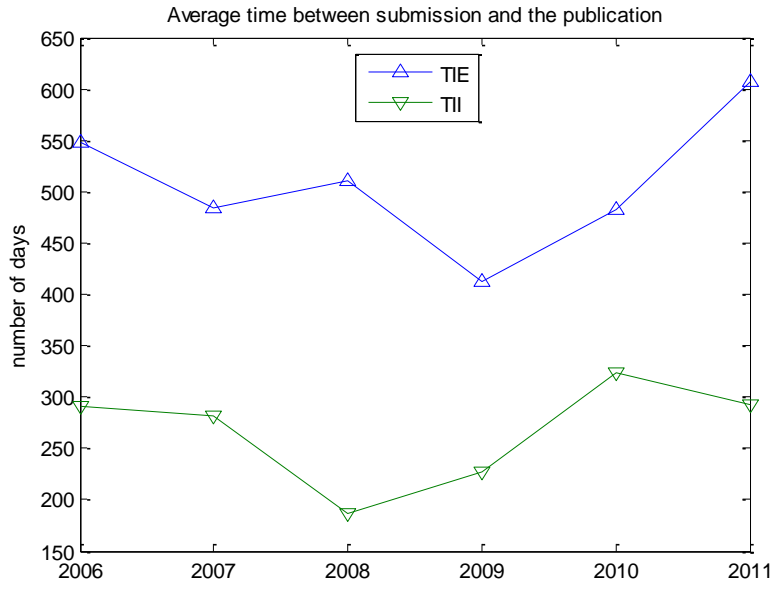


Fig. 16. Average time between submission and publication for TIE and TII

Chapter 4.

Implementation of the Text Processing Robot

Chapter 3 gives us an overview of the concept and procedure to perform several new evaluations of journal performance, such as citation analysis for EICs and AEs, and time based analysis for journals. In this chapter, we are going to delve into more details of how the text processing robot works, by looking at the main routine and several important sub routines.

4.1 Integrate Data of Interest

As stated in Chapter 3, the basis of all the new evaluation methods is to combine the useful data in two excel files into an integrated one. For every paper in “journal_citation.xls”, we are trying to extract the matching data such as submission date, first decision date, final decision date, EIC name, AE name, etc, from the other file “journal_ManuscriptReceived.xls”. As the latter file keeps record of the paper review process, it may contain multiple entries of the same paper if the paper is revised and resubmitted. Thus the data of submission date and first decision date should be extracted from the first matching entry in “journal_ManuscriptReceived.xls”, while all other data such as final decision date should be extracted from the last matching entry. The following code functions to combine the two excel files according to the above rules.

```
use Spreadsheet::ParseExcel;  
use Spreadsheet::ParseExcel::SaveParser;  
use Spreadsheet::WriteExcel;
```

```

my $parser= Spreadsheet::ParseExcel::SaveParser->new();
my $TII_citation=$parser->Parse('TII_citation.xls');
my $editor_info=$parser->Parse('TII_ManuscriptReceived.xls');

if ( !defined $TII_citation) {
    die $parser->error(), ".\n";
}
if ( !defined $editor_info) {
    die $parser->error(), ".\n";
}

```

The above code first declares 3 packages to be used, *Spreadsheet::ParseExcel*, *Spreadsheet::ParseExcel::SaveParser*, *Spreadsheet::WriteExcel*, which are related to Excel files reading and writing. Then the two excel files to be merged are opened, and the file handles are *\$TII_citation* and *\$editor_info*. After files are opened, it is necessary to check whether they are opened correctly, that's what the following two “if” statements do.

```

my $Page2_2=$editor_info->worksheet(1);
my ( $row_min1, $row_max1 ) = $Page2_2->row_range();
for $worksheet ( $TII_citation->worksheets() ){
    my ( $row_min, $row_max ) = $worksheet->row_range();
    for my $row ( 1..$row_max) {
        my $cell_title=$worksheet->get_cell($row,2);#get the paper title from 'TII_citation.xls'
        my $title=$cell_title->unformatted();
        my $LastMatchRow=0;
        my $FirstEntry=0;
        for my $row1 ( $row_min1..$row_max1) {
            #to cope with some paper with no acceptance entry
            my $cell_title1=$Page2_2->get_cell($row1,1);
            if(!defined $cell_title1)
                {next;}
            my $title_match=$cell_title1->unformatted();
            if (match($title,$title_match))
            {
                if ($FirstEntry==0)
                {$FirstEntry=1;
                my $cell_SubDate=$Page2_2->get_cell($row1,4);
                if (defined $cell_SubDate){
                    my $SubDate=$cell_SubDate->value();

```

```

        $worksheet->AddCell($row,9,$SubDate);
    }
    my $cell_FirstDecisionDate=$Page2_2->get_cell($row1,5);
    if(defined $cell_FirstDecisionDate){
        my $FirstDecisionDate=$cell_FirstDecisionDate->value();
        $worksheet->AddCell($row,10,$FirstDecisionDate);}
    }
}
$LastMatchRow=$row1;
my $cell_Decision=$Page2_2->get_cell($row1,6);
if(!defined $cell_Decision){next;}
my $Decision=$cell_Decision->unformatted();
if ($Decision=~m/Accept/) {last;}
}
}
if($LastMatchRow!=0)
{add_info($row,$LastMatchRow);}
}
}

```

The above code first selects the second worksheet *\$Page2_2* from *\$editor_info* since it contains the paper review records that we are interested in, while the first worksheet is a chart summary of the paper submission numbers and acceptance rate generated by the MC database system. Then the program enters an outer “for” loop which iterates through all the worksheets in *\$TH_citation*, with each worksheet summarizing the citations of papers published in a different year. The outer “for” loop contains 2 more inner “for” loops, with the middle one iterating through every paper listed in *\$TH_citation*, and the innermost one iterating through the worksheet *\$Page2_2*.

In the middle “for” loop, first the paper title are read from the cell (*\$row*, 2) in “Journal_citation.xls”, then two variables are declared and initialized to 0. The variable *\$LastMatchRow* is meant to record the row number of the last title matching row number in the file “journal_Manuscript.xls”, which should be the “acceptance” entry for the paper. This row is going to be used to extract data such as “final decision date”. But remember in cases where paper

lost its “acceptance” entry due to database incompleteness, the last matching row is used even if the decision state of the paper is not “accepted”. However, for information such as “submission date” and “first decision date”, the target entry is the first matching entry instead of last matching entry. The second variable *\$FirstEntry* is a flag to indicate whether it is the first time of finding a matching entry in the file. If it is, then data of “submission date” and “first decision date” is extracted and added to the worksheets of “Journal_citation.xls”, in cells (\$row, 9) and (\$row, 10) respectively.

The inner “for” loop is to search through the second worksheet of “Journal_ManuscriptReceived.xls” to find matching entries. This part has already been discussed in section 3.3.1, which also gives the flow chart of the algorithm to find the “final decision date”. The sub routine “*match*” used here to do title matching is also discussed before in 3.2.1, so no more explanation will be given here.

At last, a sub routine “*add_info*” is called to add data from the last matching row in “Journal_ManuscriptReceived.xls” to “Journal_citation.xls”. The added data includes *final decision date, author institution, EIC full name and AE full name*. Note here in the sub routine, also in the code above, the cells to be read are first checked empty or not. Because if the cell is empty, the call to the method *\$cell->value()* is illegal and will cause an error.

```
sub add_info
  {my $row=$_[0];
  my $row1=$_[1];

  my $cell_DecisionDate=$Page2_2->get_cell($row1,5);
  if(defined $cell_DecisionDate){
  my $DecisionDate=$cell_DecisionDate->value();
  $worksheet->AddCell($row,11,$DecisionDate);}

  my $cell_Ins=$Page2_2->get_cell($row1,7);
```



```

if(defined $cell_Ins){
my $Ins=$cell_Ins->unformatted();
$worksheet->AddCell($row,12,$Ins);}

my $cell_EIC=$Page2_2->get_cell($row1,8);
if(defined $cell_EIC){
my $EIC=$cell_EIC->unformatted();
$worksheet->AddCell($row,13,$EIC);}

my $cell_Editor=$Page2_2->get_cell($row1,9);
if(defined $cell_Editor){
my $Editor=$cell_Editor->unformatted();
$worksheet->AddCell($row,14,$Editor);}
}

```

4.2 Get the Publication Issue

Since the publication date is not contained in the MC database, we have to find other ways to obtain the publication date for papers. In this thesis, we choose to look up the output html files of the Internet robot introduced in Chapter 1. The following sub routine serves to extract the publication date information for all the papers in “journal_citation.xls”. Two input arguments are passed to this sub routine, which are the year of the paper being published and the title of the paper.

```

sub get_pubissue
{
my $year=$_[0];
$year=$_[0]-2004; #TII starts from year 2004
my $file="e:/website_manage/TIIpub/".$year.".s.htm";
open(H,$file) || die "couldn't open the file";
my @lines=<H>;
my $total_line=@lines;
my $title=$_[1];
$title=~s/(\\W+)/g;#remove some strange characters such as "-"
$title=~s/(\\W+)$//;

```

In the above code snippet, first the directory and name of the html file to be searched is assigned to the variable *\$file*. According to the naming rule of the Internet Robot, the volume number is used to name the html file that record the information of papers in a given publication year. For example, TII starts from the year 2005, so publications in the year 2011 will fall into volume 7, and 2011's html file is named "7s.htm". After opening the corresponding html file, all its content is copied to an array variable *@lines*, and the length of the array variable is assigned to *\$total_line*.

```

for(my $i=1;$i<$total_line;$i++){
  if($lines[$i]=~m/<td valign="*top"*>/i){
    my @array1=split(/&nbsp;"/,$lines[$i]);
    my $title_match=$array1[1];

    if($year==7){
      my @array2=split(/<\a>/,$title_match);
      $title_match=$array2[0];
    }
    else{
      my @array2=split(/<i>/,$title_match);
      $title_match=$array2[0];
    }
    $title_match=~s/(\W+)/g;#remove some strange characters
    $title_match=~s/(\W+)$/;

    if($title=~m/$title_match/i){
      my $volume,$issue,$order)=( $lines[$i]=~m/(\d+)\.(\d+)\.(\d+)/);
      return $issue;
    }
  }

}
return 0;
}

```

The above code seems messy because it is dealing with the syntax of the html file. It tries to first locate the lines containing titles of papers and then extract the titles from those lines. One example of such a html line is like following:

`<td valign="top">5.1.2 </td><td> Junyoung Heo, Jiman Hong, Yookun Cho, "EARQ: Energy Aware Routing for Real-Time and Reliable Communication in Wireless Industrial Sensor Networks`

After the paper title is extracted, comparison of the title with the 2nd input is performed. If successful, the paper’s issue number is searched and extracted in the same line. If no matching title is found in the html file, the sub routine will return 0.

By far, the data needed to perform both citation based and time based analysis is complete, and a figure of “journal_citation” at this stage is shown below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Cites	Authors	Title	Year	Source	Publisher	ArticleURL	CitesURL	Paper Type	Submission	FirstDec	DFinalDec	C Author	Inst	EIC Full	N:Editor Full	Issue
2	34	G Acciani,	Applicator	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Mar 28, 20	May 06, 2	(May 06, 2	(Politecnico	Zurawski,	I Vaccaro,	#3	
3	31	FL Lian, J	Network ar	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 04, 2	Jul 20, 20	Jul 20, 20	C National T	Zurawski,	Richard	1	
4	22	J Ferreira,	Combining	2006	...	IEEE	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 2	Jan 04, 20	Mar 17, 20	Faculdade	Zurawski,	I Sauter,	Th 2	
5	70	AP Kaloge	Vertical int	2006	...	IEEE	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 2	Jan 04, 20	Mar 29, 20	University	Zurawski,	I Sauter,	Th 2	
6	18	M Barranc	An active s	2006	...	IEEE	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 2	Jan 04, 20	Mar 29, 20	University	Zurawski,	I Sauter,	Th 2	
7	16	G Cena...	On the pro	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 2	Jan 04, 20	Mar 29, 20	University	Zurawski,	I Sauter,	Th 2	
8	85	JLM Lastr	Semantic v	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 25, 2	Oct 05, 20	Oct 05, 20	Tampere U	Zurawski,	I Colombo,	#1	
9	48	T Skeie,	S Timeliness	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Aug 15, 2	Oct 10, 20	Jan 01, 2006	8:50:02	Zurawski,	I Naedele,	#1	
10	10	Q Morante	Pervasive c	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Dec 12, 2	Feb 19, 20	May 01, 2	(University	Zurawski,	I Gomes,	L#3	
11	3	C Ryan, D	Interactive	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	I	Mar 14, 20	May 01, 2	Aug 08, 2	C University	Zurawski,	I Naedele,	#4	
12	25	T Chang, F	Web-base	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 09, 2	Jul 20, 20	Oct 05, 20	NJIT	Zurawski,	I Ho, Weng	#1	
13	14	R Obermai	Reuse of C	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Jan 09, 20	Mar 24, 20	Jun 16, 20	Vienna Un	Zurawski,	I Vogel-Heu	#4	
14	7	M Long...	Energy-effi	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Jun 28, 20	Sep 14, 2	Dec 14, 2	C Auburn Un	Zurawski,	I Naedele,	#1	
15	2	G Chen, Z	Coordinati	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	I	Jan 23, 20	Mar 22, 20	Jul 23, 20	C Nanyang T	Zurawski,	I Pham, Duc	#4	
16	75	ZW Li...	Two-stage	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Oct 11, 20	Dec 23, 2	May 24, 2	(Xidian Univ	Zurawski,	I Gomes,	L#4	
17	3	HCW Lau,	Developme	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Dec 15, 2	Feb 19, 20	Sep 02, 2	C The Hong I	Zurawski,	I Vaccaro,	#4	
18	10	J Kolbusz,	Network tr	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Sep 23, 2	Nov 30, 20	Jun 14, 20	Auburn Un	Zurawski,	I Malinowsk	#4	
19	3	RG Garcia	Applying a	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Jun 30, 20	Dec 28, 2	Mar 22, 20	EPFL	Zurawski,	I Martinez	L#3	
20	80	F De Pelle	On the us	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	R	May 20, 2	Dec 14, 2	Feb 19, 20	IEIIT_CN	Zurawski,	I Colombo,	#2	
21	52	TT Quan,	Automatic	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 20, 2	Oct 26, 20	Feb 23, 20	Nanyang T	Zurawski,	I Martinez	L#3	
22	16	IM Delame	Service-ori	2006	...	Informati	http://ieeexplore.ieee.org	http://scholar.google.com	I	Sep 02, 2	Mar 07, 20	Jun 08, 20	Tampere U	Zurawski,	I Dillon, Tha	#4	
23	10	W Elmenr	Time-trigg	2006	Industrial I	IEEE explore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 04, 2	Jun 20, 20	Mar 07, 20	Vienna Un	Zurawski,	I Yu, Xinghu	#3	

Fig. 17. Snapshot of “TII_citation.xls” with all the data needed: Citations (Column A), Paper Type (Column I), Submission Date (Column J), First Decision Date (Column K), Final Decision Date (Column L), EIC Full Name as in Column N, AE Full Name as in Column O, and Issue Number as in Column P.

4.3 Time Averaged Citation Number for Papers

As mentioned in Chapter 3, the citations for papers need to be first averaged over time before the average citation for EICs and AEs can be computed. Two sub routines are needed to calculate time averaged citations for papers, “*get_days()*” and “*cite_ave()*”. The algorithm of “*get_days()*” is already discussed before in Chapter 3, so no more explanations will be given here. The complete code of “*get_days()*” is in appendix. The sub routine “*cite_ave()*” requires two input arguments, publication date and the citation number of the paper, and it assumes the current date is "Oct 03,2011". The forth line of the sub routine calls “*get_days()*” to get the number of passed days between the paper’s publication date and current date, and then it approximates the quarter years by rounding up the passed days over 120. At last, the time averaged citations is computed and returned.

```
sub cite_ave()  
  {my $pub_date=$_[0];  
  my $cites=$_[1];  
  my $current="Oct 03,2011";  
  my $past_time=get_days($pub_date,$current);  
  my $past_quarter=ceil($past_time/120);  
  my $cite_ave=$cites/$past_quarter;  
  return $cite_ave;  
}
```

4.4 Averaging Citations for AEs

After the time averaged citations are computed for every paper, it is easy to compute the average citations of papers selected by different AEs. To simplify the code, every worksheet is first sorted by the column of AEs so that papers processed by the same AE will be adjacent to each other. A sub routine is written to do the calculation, which requires two input arguments, the column number of data to be averaged and the column number of AEs. And the final

averaged results will be written to a text file with the format "AE name; total citation; Paper Number; Averaged citations;\n".

```
sub ave_editor
  {my $col_data=$_[0];
  my $col_editor=$_[1];
  my $cell_editor=$sheet2->get_cell(1,$col_editor);
  my $editor=$cell_editor->unformatted();
  my $cell_data=$sheet2->get_cell(1,$col_data);
  my $data=$cell_data->unformatted();
  my $paperNumber=1;
  my $save=0;
  open (F,">>data.txt")|| die "couldn't open data.txt!\n";
```

The above code first reads in the two inputs, column number of the data to be averaged and the AEs, then reads the two cells in the first row to initialize two variables *\$editor* and *\$data*. *\$editor* is used to store the name of the AE, and *\$data* is used to store the total citations of papers processed by this AE. Next, a text file “*data.txt*” is opened and is going to be used to store the results in the following code.

```
for my $row (2..$row_max1){
  my $cell_editorNext=$sheet2->get_cell($row,$col_editor);
  if (!defined $cell_editorNext){last;}
  my $editor_next=$cell_editorNext->unformatted();
  my $cell_dataNext=$sheet2->get_cell($row,$col_data);
  if (!defined $cell_dataNext){next;}
  my $data_next=$cell_dataNext->unformatted();
  if ($editor eq $editor_next){
    $data+=$data_next;
    $paperNumber++;}
  else {
    if($paperNumber!=0){$save=$data/$paperNumber;}
    print F "$editor; $data; $paperNumber; $save;\n";
    $data=$data_next;
    $paperNumber=1;
    $editor=$editor_next;
  }
}
print F "$editor; $data; $paperNumber; $save;\n";
close F;
```

}

The above code examines whether the next row has the same AE with the previous row, if it does, then the data of interest in this row should be added to the total data; Otherwise, it indicates that all the papers processed by the previous AE has been counted, the result for this AE need to be written to “*data.txt*”. Also, if a new AE is encountered, the two variables *\$editor* and *\$data* should be reinitialized. The last two lines of code are used to record the results for the last AE on the sorted worksheet.

4.5 Average Citations for SS

To compute the average citations for Special Sections, the method used in 4.4 is totally applicable in this situation. But the method above has the deficiency of having to sort every worksheet in the file first before being able to call the sub routine to compute average citations. In this section, an alternative sub routine is provided without the need of any pre sorting work, at the price of slightly degraded efficiency of execution.

```
#This sub takes no argument, and returns several arrays of data regarding citations for every SS on $sheet2  
sub getSScitation()  
  {my @SSname=(); # to store the names of SSs  
  my @SScitation=(); #to store the total raw citations of SSs  
  my @papernum=(); # to store the total paper number of SSs  
  my @time=(); # to store the publication time of SSs  
  my @to_now=(); # to store the passed time (unit: year ) from publication to current date  
  for my $row (1..$row_max1)  
  {my $cell_PaperType=$sheet2->get_cell($row,10);  
  if (!defined $cell_PaperType){next;}  
  my $PaperType=$cell_PaperType->unformatted();  
  if($PaperType!~m/^SS/){next;} # if it is a regular paper, jump to the next row  
  #get citation  
  my $cell_citation=$sheet2->get_cell($row,0);
```

```

my $citation=$cell_citation->value();
#get issue
my $cell_issue=$sheet2->get_cell($row,17);
my $issue=$cell_issue->value();
#get year
my $cell_year=$sheet2->get_cell($row,3);
my $year=$cell_year->value();
#get publish--now time period
my $pubtime=get_pubdate($year,$issue);
my $period=get_days($pubtime,"Oct 05, 2011");
my $p_year=$period/365; # period in year, ex, 1.5 years;
$p_year=sprintf("%.2f",$p_year); #format the floating number $p_year

```

In the above code, every row in *\$sheet2* is examined to see whether the paper belongs to a SS or just a regular paper. If the paper in a given row is a regular paper, then the rest of the for loop will be skipped and next row will be examined until a SS paper is encountered. Then the data of interest of the SS paper is extracted, such as citations, issue number, publication year and publication to current time period. The algorithm used next is as such: for every SS paper encountered, its SS name is looked up in the array *@SSname*. If there is such an element in *@SSname*, it indicates that at least a paper in the same SS has been previously counted, and the citation number of the current paper need to be added to the total citations of the SS, also the paper number of the SS should increment by 1. Otherwise, a new SS is discovered, and its information such as name, initial citations and paper number should be added to corresponding arrays. At last, the 5 arrays are returned.

```

my $num_SS=@SSname;
my $flag=0;#flag whether the above SS name is already contained in @SSname
for my $i(0..($num_SS-1)){
  if ($SSname[$i] eq $PaperType){
    $flag=1;
    $SScitation[$i]+=$citation;
    $papernum[$i]++;
    last;
  }
}

```

```

if ($flag==0){ #new SSname, need to add to the two arrays
  push(@SSname,$PaperType);
  push(@SScitation,$citation);
  push(@papernum,1);
  push(@time,$year."/".$issue);
  push(@to_now,$p_year);
}
}
return (@time,@SSname,@papernum,@SScitation,@to_now);
}

```

4.6 Average Time Analysis

For time based evaluations proposed in chapter 3, first three time gaps between final decision date and submission date, publication date and final decision date, first decision date and submission date need to be computed for every paper, then an average is computed for every year for the journal. The sub routine “*get_days()*” can again be used to calculate the passed days between two dates, thus solving the above problem. The following figure shows the resulting excel file after getting such data. After calculating the desired data, the built in average function of Microsoft Excel is used to compute the average time periods in days for the 3 columns: “*Dec_Sub*”, “*Pub_Dec*” and “*FirstDec_Sub*”.

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Title	Year	Source	Publisher	ArticleURL	CitesURL	Paper Type	Submission	FirstDec	FinalDec	Author Inst	EIC Full N	Editor Full	Issue	Dec_Sub	Pub_Dec	FirstDec_Sub	
2	Application	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Mar 28, 20	May 06, 21	May 06, 21	Politecnico di Milano	Zurawski, I	Vaccaro, F	3	39	96	39	
3	Network ar	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 04, 21	Jul 20, 20	Jul 20, 20	National Technical University of Athens	Zurawski, R	Richard	1	77	205	77	
4	Combining	2006	...	IEEE	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 20	Jan 04, 20	Mar 17, 20	Faculdade de Engenharia	Zurawski, I	Sauter, Th	2	99	54	27
5	Vertical int	2006	...	IEEE	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 20	Jan 04, 20	Mar 29, 20	University of Zielona Gora	Zurawski, I	Sauter, Th	2	111	42	27
6	An active c	2006	...	IEEE	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 20	Jan 04, 20	Mar 29, 20	University of Zielona Gora	Zurawski, I	Sauter, Th	2	111	42	27
7	On the pro	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	SS on Cor	Dec 08, 20	Jan 04, 20	Mar 29, 20	University of Zielona Gora	Zurawski, I	Sauter, Th	2	111	42	27
8	Semantic	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 25, 21	Oct 05, 20	Oct 05, 20	Tampere University	Zurawski, I	Colombo, J	1	133	128	133
9	Timeliness	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Aug 15, 20	Oct 10, 20	Jan 01, 2006	8:50:02	Zurawski, I	Naedele, H	1	139	40	56	
10	Pervasive	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Dec 12, 20	Feb 19, 20	May 01, 21	University of Zielona Gora	Zurawski, I	Gomes, L	3	140	101	69	
11	Interactive	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Mar 14, 20	May 01, 21	Aug 08, 20	University of Zielona Gora	Zurawski, I	Naedele, H	4	147	94	48
12	Web-base	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 09, 21	Jul 20, 20	Oct 05, 20	NJIT	Zurawski, I	Ho, Weng	1	149	128	72	
13	Reuse of C	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Jan 09, 20	Mar 24, 20	Jun 16, 20	Vienna University of Applied Sciences	Zurawski, I	Vogel-Heu	4	158	147	74	
14	Energy-effi	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Jun 28, 20	Sep 14, 20	Dec 14, 20	Auburn University	Zurawski, I	Naedele, H	1	169	58	78	
15	Coordinati	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Jan 23, 20	Mar 22, 20	Jul 23, 20	Nanyang Technological University	Zurawski, I	Pham, Duc	4	181	110	58
16	Two-stage	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Oct 11, 20	Dec 23, 20	May 24, 21	Xidian University	Zurawski, I	Gomes, L	4	225	170	73	
17	Developme	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Dec 15, 20	Feb 19, 20	Sep 02, 20	The Hong Kong Polytechnic University	Zurawski, I	Vaccaro, F	4	261	69	66	
18	Network tr	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Sep 23, 20	Nov 30, 20	Jun 14, 20	Auburn University	Zurawski, I	Malinowski	4	264	149	68	
19	Applying a	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	Jun 30, 20	Dec 28, 20	Mar 22, 20	EPFL	Zurawski, I	Martinez L	3	265	141	181	
20	On the use	2006	Industrial I	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	R	May 20, 21	Dec 14, 20	Feb 19, 20	IEIT_CNR	Zurawski, I	Colombo, J	2	275	80	208	
21	Automatic	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	P	May 20, 21	Oct 26, 20	Feb 23, 20	Nanyang Technological University	Zurawski, I	Martinez L	3	279	168	159
22	Service-on	2006	...	Informal	ieeexplore	http://ieeexplore.ieee.org	http://scholar.google.com	I	Sep 02, 20	Mar 07, 20	Jun 08, 20	Tampere University	Zurawski, I	Dillon, Tha	4	279	155	186

Fig. 18. Snapshot of “TII_citation.xls” with time gaps information. (Column Q shows days from *Submission Date* to *Final Decision Date*, Column R shows days from *Publication Date* to *Final Decision Date*, Column S shows days from *First Decision Date* to *Submission Date*)

Chapter 5

Conclusion and Future work

The new methods of journal performance evaluation proposed in this thesis provide a more detailed view towards the work of EICs and AEs, which complement the traditional methods which always treat the entire journal as a whole. And to the best of my knowledge it is also the first to consider the time performance of journals. The text processing robot, which successfully accomplishes the task of data integration and processing, is a preferable solution to implement the new evaluations. Provided with the necessary data files, the text processing robot can automatically combine data of interest into one file, and do the desired computation and analysis to the integrated data, thus yielding results to do the new evaluations of journals.

Most advantages of the text processing robot, such as simplicity, fast speed and accuracy are thanks to the inherent features of its implementing language, Perl Script. As has been shown throughout the thesis, Perl is a more powerful language in text processing compared with other popular languages such as C++. Its built in regular expression syntax and many free but powerful packages are great tools for programmers. In addition to text processing, Perl is also popular and widely use in other areas such as network programming (CGI), database management, etc.

It is obvious that good papers have a good chance for citations, but there are other things that can also affect citations. For example, a paper with very good ideas will not be cited if it is not found and read. Therefore there are several other elements that can be investigated about

their influence on journal citations. Specifically, the following aspects are interesting future research topics:

(1) The influence of titles and abstracts on the citations of papers, for example, will papers with titles/abstracts containing more keywords be better cited? Will the length of titles/abstracts affect citations?

(2) The manuscript should be within the scope of the journal. It is important because papers out of the journal scope have reduced chances to be found and cited. Some future work may be devoted to quantify the fitness of scope of a paper to the journal, and its relationship to citations. One way to verify the scope is to check if the manuscript is linked with previously published papers in the journal.

(3) A comparison of existing techniques with some comments about their efficiencies are always interesting to readers. It would be helpful for authors to know whether the number of related work explained in a paper will affect its citations or not.

References

- [1] Jiao Yu, P. Gnanachchelvi, B. M. Wilamowski, "Performance Analysis of IES Journals using Internet and Text Processing Robots", Proc. of the 27th Annual Conference of the IEEE Industrial Electronics Society, pp. 4612-4618, Melbourne Australia, Nov 7-10, 2011.
- [2] Randal L. Schwartz , Brian D Foy, Tom Phoenix, Learning Perl O'Reilly Media, Inc 2011 (sixth editions).
- [3] Althouse BM, West JD, Bergstrom TC, Bergstrom CT. "Differences in impact factor across fields and over time", Department of Economics, University of California, Santa Barbara. Departmental Working Papers. Paper 2008-4-23, April 23, 2008.
- [4] Christ Tomer, "A statistical assessment of two measures of citation: The impact factor and the immediacy index", Information Processing and Management, Volume 22, Issue 3, pp. 251-258, 1986.
- [5] Bergstrom CT. "Eigenfactor: measuring the value and prestige of scholarly journals", C&RL News 2007;68: No. 5.
- [6] Carl T. Bergstrom and Jevin D. West, "Assessing citations with the Eigenfactor™ Metrics", Neurology 2008;71;1850-1851.
- [7] Dou Xiqian, Qi Yanli, "A Brief Analysis of Eigenfactor Score and Article Influence Score", Journal of Academic Libraries, June 2009.
- [8] Aleksander Malinowski and Bogdan Wilamowski " Paper Collection and Evaluation Through the Internet", Proc. of the 27th Annual Conference of the IEEE Industrial Electronics Society, pp. 1868-1873, Denver CO, Nov 29-Dec 2, 2001.
- [9] Nam Pham and B. M. Wilamowski "IEEE article data extraction from internet", 13-th IEEE Intelligent Engineering Systems Conference, INES 2009, Barbados, April 16-18,2009, pp. 251-256.
- [10] Bogdan M. Wilamowski "Design of network based software", 24th IEEE International Conference on Advanced Information Networking and Applications 2010, April 20-23, 2010, Perth, Australia, pp.4-10.
- [11] Nam Pham, B. M. Wilamowski and Aleksander Malinowski,"Running Software over Internet" Industrial Electronics Handbook, vol. 4 –Industrial Industrial Communication Systems, 2nd Edition, chapter 63, pp. 63-1 to 63-11, CRC Press 2011.

- [12] M. Manic, B. M. Wilamowski, and A. Malinowski “Internet Based Neural Network Online Simulation Tools” Proc. of the 28th Annual Conference of the IEEE Industrial Electronics Society, pp. 2870-2874, Sevilla, Spain, Nov 5-8, 2002.
- [13] Nam Pham, Hao Yu, B. M. Wilamowski, “Neural network trainer through computer networks”, 24th IEEE International Conference on Advanced Information Networking and Applications 2010, pp. 1203-1209, 2010.
- [14] Bogdan Wilamowski, Aleksander Malinowski, and John Regnier, “Internet as a New Graphical User Interface for the SPICE Circuit Simulator”, IEEE Transactions on Industrial Electronics, vol. 48. No. 6, pp. 1266 –1268, Dec. 2001.
- [15] Nam Pham and B. M. Wilamowski, "Automatic Data Mining on Internet by Using PERL” Industrial Electronics Handbook, vol. 4 –Industrial Industrial Communication Systems, 2nd Edition, chapter 65, pp. 65-1 to 65-9, CRC Press 2011.
- [16] S. Neeli, K. Govindasamy, B.M. Wilamowski, and A. Malinowski, “Automated Data Mining from Web Servers Using Perl Script” 12th INES 2008 -International Conference on Intelligent Engineering Systems, Miami, Florida, USA, February 25-29, 2008, pp. 191-196.

APPENDICES

PERL CODE OF TEXT PROCESSING ROBOT FOR
NEW EVALUATIONS OF JOURNALS

APPENDIX A: combine_data.pl

```
#####This program aims to integrate data of interest from 3 sources: journal_citation.xls,
#####journal_ManuscriptReceived.xls, and output from the Internet Robots.
#####After running the program, “journal_citation.xls” should contain extra data:
#####Submission Date, First Decision Date, Final Decision Date, Author Institution,
#####EIC name, AE name, Issue number

use Spreadsheet::ParseExcel;
use Spreadsheet::ParseExcel::SaveParser;
use Spreadsheet::WriteExcel;

my $parser= Spreadsheet::ParseExcel::SaveParser->new();
my $TII_citation=$parser->Parse('TII_citation.xls');
my $editor_info=$parser->Parse('TII_ManuscriptReceived.xls');

if ( !defined $TII_citation) {
    die $parser->error(), ".\n";
}
if ( !defined $editor_info) {
    die $parser->error(), ".\n";
}
my $Page2_2=$editor_info->worksheet(1);
my ( $row_min1, $row_max1 ) = $Page2_2->row_range();
for $worksheet ( $TII_citation->worksheets() ){
    my ( $row_min, $row_max ) = $worksheet->row_range();
    for my $row ( 1..$row_max ) {
        my $cell_title=$worksheet->get_cell($row,2);#get the paper title from 'TII_citation.xls'
        my $title=$cell_title->unformatted();
        my $cell_year=$worksheet->get_cell($row,3);#get the publish year from “TII_citation.xls”
        my $year=$cell_year->value();
        my $issue= get_pubissue($year, $title); #get the publication issue number
        $worksheet->AddCell($row,15,$issue); #write the issue number to “TII_citation.xls”
        my $LastMatchRow=0;
        my $FirstEntry=0;
        for my $row1 ( $row_min1..$row_max1 ) {
            #to cope with some paper with no acceptance entry
```

```

my $cell_title1=$Page2_2->get_cell($row1,1);
if(!defined $cell_title1)
{next;}
my $title_match=$cell_title1->unformatted();
if (match($title,$title_match))
{
  if ($FirstEntry==0)
  {$FirstEntry=1;
  my $cell_SubDate=$Page2_2->get_cell($row1,4);
  if (defined $cell_SubDate){
    my $SubDate=$cell_SubDate->value();
    $worksheet->AddCell($row,9,$SubDate);
  }
  my $cell_FirstDecisionDate=$Page2_2->get_cell($row1,5);
  if(defined $cell_FirstDecisionDate){
    my $FirstDecisionDate=$cell_FirstDecisionDate->value();
    $worksheet->AddCell($row,10,$FirstDecisionDate);}
  }
  $LastMatchRow=$row1;
  my $cell_Decision=$Page2_2->get_cell($row1,6);
  if(!defined $cell_Decision){next;}
  my $Decision=$cell_Decision->unformatted();
  if ($Decision=~m/Accept/) {last;}
}
}
if($LastMatchRow!=0)
{add_info($row,$LastMatchRow);}
}
}

```

```

***** all subroutines*****
*****This sub takes two arguments, row# in "Citation.xls" and row# in "Manuscript.xls",
*****and add info to "Citation.xls"

```

```

sub add_info
{my $row=$_[0];
  my $row1=$_[1];

  my $cell_DecisionDate=$Page2_2->get_cell($row1,5);
  if(defined $cell_DecisionDate){
    my $DecisionDate=$cell_DecisionDate->value();
    $worksheet->AddCell($row,11,$DecisionDate);}

  my $cell_Ins=$Page2_2->get_cell($row1,7);
  if(defined $cell_Ins){
    my $Ins=$cell_Ins->unformatted();
  }
}

```



```

$worksheet->AddCell($row,12,$Ins);}

my $cell_EIC=$Page2_2->get_cell($row1,8);
if(defined $cell_EIC){
my $EIC=$cell_EIC->unformatted();
$worksheet->AddCell($row,13,$EIC);}

my $cell_Editor=$Page2_2->get_cell($row1,9);
if(defined $cell_Editor){
my $Editor=$cell_Editor->unformatted();
$worksheet->AddCell($row,14,$Editor);}
}

```

*****sub "match" takes two strings as input, removes multiple space and strange characters,
*****then compares whether the two string are equal or not (case insensitive)

```

sub match
{
$string1=$_[0];
$string2=$_[1];
$string1=~s/(\W+)/ /;
$string1=~s/(\W+)$//;
$string2=~s/(\W+)/ /;
$string2=~s/(\W+)$//;
if (lc($string1) eq lc($string2)){
return 1;}
else{return 0;}
}

```

*****This sub takes two arguments, publish year and paper title, returns paper publish
*****issue number

*****used differently for TII and TIE

```

#sub get_pubissue
# {
# my $year=$_[0];
# $year=$_[0]-1953; #TII starts from year 2004
# my $file="e:/website_manage/TIEpub/".$year.".s.htm";
# open(H,$file) || die "couldn't open the file";
# my @lines=<H>;
# my $total_line=@lines;
# my $title=$_[1];
# $title=~s/(\W+)/ /g;#remove some strange characters such as "-"
# $title=~s/(\W+)$//;
# for(my $i=1;$i<$total_line;$i++){
# if($lines[$i]=~m/<td valign="*top"*/i){ #in 58s.htm "td vAlign=top"

```

```

#     #print($lines[$i]);
#     my @array1=split(/\&nbsp;/,$lines[$i]);
#     my $title_match=$array1[1];
#
#     if ($year==58){
#         my @array2=split(/<\A>/,$title_match);
#         $title_match=$array2[0];
#     }
#     else{
#         my @array2=split(/<i>/,$title_match);
#         $title_match=$array2[0];
#     }
#     $title_match=~s/(\W+)/ /g;#remove some strange characters such as "-"
#     $title_match=~s/(\W+)$/;/;
#
#     if($title=~m/$title_match/i){
#         my ($volume,$issue,$order)=(($lines[$i]=~m/(\d+)\.(\d+)\.(\d+)/);
#         #print("$volume,$issue");
#         return $issue;
#     }
# }
#
# }
# return 0;
# }

```

```

sub get_pubissue
{
    my $year=$_[0];
    $year=$_[0]-2004; #TII starts from year 2004
    my $file="e:/website_manage/TIIpub/" . $year . ".s.htm";
    open(H,$file) || die "couldn't open the file";
    my @lines=<H>;
    my $total_line=@lines;
    my $title=$_[1];
    $title=~s/(\W+)/ /g;#remove some strange characters such as "-"
    $title=~s/(\W+)$/;/;
    for(my $i=1;$i<$total_line;$i++){
        if($lines[$i]=~m/<td valign="*top"*>/i){ #in 58s.htm "td vAlign=top"
            #print($lines[$i]);
            my @array1=split(/\&nbsp;/,$lines[$i]);
            my $title_match=$array1[1];

            if ($year==7){
                my @array2=split(/<\a>/,$title_match);
                $title_match=$array2[0];
            }
        }
    }
}

```

```

}
else{
my @array2=split(/<i>/,$title_match);
$title_match=$array2[0];
}
$title_match=~s/(\W+)/g;#remove some strange characters such as "-"
$title_match=~s/(\W+)$/;

if($title=~m/$title_match/i){
my ($volume,$issue,$order)=(($lines[$i]=~m/(\d+)\.(\d+)\.(\d+)/));
#print("$volume,$issue");
return $issue;
}
}

}
return 0;
}

```

APPENDIX B: analyze.pl

```
use Switch;
use POSIX;
use Spreadsheet::ParseExcel;
use Spreadsheet::ParseExcel::SaveParser;
use Spreadsheet::WriteExcel;

my $parser= Spreadsheet::ParseExcel::SaveParser->new();
my $TII_citation=$parser->Parse('TII_citation.xls');
if ( !defined $TII_citation ) {
    die $parser->error(), "\n";
}

for my $sheetnum(0..7){
    $sheet2=$TII_citation->worksheet($sheetnum);
    ( $row_min1, $row_max1 ) = $sheet2->row_range();
    open(SS,">>SScitation.txt")||die "couldnt open SSCitation.txt!";
    my ($r1,$r2,$r3,$r4,$r5)=&getSScitation();

    my @time=@$r1;
    my @SSname=@$r2;
    my @papernum=@$r3;
    my @SScitation=@$r4;
    my @to_now=@$r5;
    my $SSnum=@SSname;

    for my $i (0..($SSnum-1)){
        print SS "$time[$i],$SSname[$i],$papernum[$i],$SScitation[$i],$to_now[$i]\n";
    }

    close(SS);
}

for my $row(1..$row_max1)
{
    # add Dec-Sub
```

```

my $cell_SubmissionDate=$sheet2->get_cell($row,9);
if(!defined $cell_SubmissionDate)
{
    next;
}
my $SubmissionDate=$cell_SubmissionDate->value();

my $cell_DecisionDate=$sheet2->get_cell($row,11);
if(!defined $cell_DecisionDate)
{
    next;
}
my $DecisionDate=$cell_DecisionDate->value();
my $day_num=&get_days($SubmissionDate,$DecisionDate);
$sheet2->AddCell($row,16,$day_num);
# add pub-Dec and average citation over time for every paper
my $cell_Issue=$sheet2->get_cell($row,15);
if(!defined $cell_Issue)
{
    next;
}
my $Issue=$cell_Issue->value();
if ($Issue!=0){
    my $pubdate=&get_pubdate($year,$Issue);
    my $cell_cites=$sheet2->get_cell($row,0);
    my $cites=$cell_cites->value();
    my $ave_cites_time=cite_ave($pubdate, $cites);
    $sheet2->AddCell($row, 19, $ave_cites_time);
    my $pub_Dec=&get_days($DecisionDate,$pubdate);
    $sheet2->AddCell($row,17,$pub_Dec);
}
#add FirstDec-Sub
my $cell_firstDec=$sheet2->get_cell($row,10);
if(!defined $cell_firstDec)
{
    next;
}
my $firstDec=$cell_firstDec->value();
my $firstDec_Sub=get_days($SubmissionDate, $firstDec);
$sheet2->AddCell($row,18, $firstDec_Sub);

}
}
$TII_citation->SaveAs('TII_Citation.xls');

# *****all subroutines*****

```

```

****This sub takes two arguments, submission date and decision date,
#and return the time difference in days.
sub get_days
{my @month_length=(31,28,31,30,31,30,31,31,30,31,30,31);
  my %month_order=(Jan=>0,
    Feb=>1,
    Mar=>2,
    Apr=>3,
    May=>4,
    Jun=>5,
    Jul=>6,
    Aug=>7,
    Sep=>8,
    Oct=>9,
    Nov=>10,
    Dec=>11);
  my $start_date=$_[0];
  my $end_date=$_[1];
  $start_date=~m/,\s(\d+)/;
  my $start_year=$1;
  $start_date=~m/(\w+)\s(\d+)/;
  my ($start_month,$start_day)=(($1,$2));
  $end_date=~m/,\s(\d+)/;
  my $end_year=$1;
  if($end_year lt $start_year){ return 0};
  $end_date=~m/(\w+)\s(\d+)/;
  my ($end_month,$end_day)=(($1,$2));
  my $total_month=($end_year-$start_year)*12+$month_order{$end_month}-
$month_order{$start_month};
  my $days=0;
  my $i=$month_order{$start_month};
  for my $j (0..($total_month-1))
  {
    $days=$days+$month_length[$i];
    $i++;
    $i=$i%12;
  }
  $days=$days+$end_day;
  $days=$days-$start_day;
  return ($days);
}
*****This sub takes two argument, year and issue number, and translate it to "Month
date, year"
***** for TII*****
sub get_pubdate
{my $year=$_[0];

```

```

my $issue=$_[1];
my $month=0;
switch ($issue) {
  case (1){$month="Feb";}
  case (2){$month="May";}
  case (3){$month="Aug";}
  case (4){$month="Nov";}
}
my $pubdate="$month 10, $year";
return ($pubdate);
}
#####for TIE#####
#sub get_pubdate
#{my $year=$_[0];
# my $issue=$_[1];
# my $month=0
# my
@month_name=("Jan","Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec");
# $month=$month_name[$issue-1];
# my $pubdate="$month 10, $year";
# return ($pubdate);
#}
#####This sub takes two argument, pub_date&citation, returns citation averaged by quarter
#year#####
sub cite_ave()
{my $pub_date=$_[0];
my $cites=$_[1];
my $current="Oct 03, 2011";
my $past_time=get_days($pub_date,$current);
my $past_quarter=ceil($past_time/120);
my $cite_ave=$cites/$past_quarter;
return $cite_ave;
}

```

APPENDIX C: aveCitations_AE

```
#####this program computes the average citations for EICs or AEs, and writes the result to
#####"data.txt". Note, need to first sort "journal_citation.xls" according to EICs or AEs
use Spreadsheet::ParseExcel;
use Spreadsheet::ParseExcel::SaveParser;
use Spreadsheet::WriteExcel;

my $parser= Spreadsheet::ParseExcel::SaveParser->new();
my $TII_citation=$parser->Parse('TII_citation.xls');
if ( !defined $TII_citation ) {
    die $parser->error(), ".\n";
}

for my $sheetnum(0..7){
    $sheet2=$TII_citation->worksheet($sheetnum);
    ( $row_min1, $row_max1 ) = $sheet2->row_range();
    ave_editor(19, 14); # the second input: 13 for EICs, 14 for AEs
}

#####subroutines#####
#####This sub takes two arguments, the column number to be averaged, by EIC or
AE ,and generates
#####a text file of data from "Citation.xls" by EIC(13) or AE(14);
#####Must first sort xls file by EIC or AE accordingly!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

sub ave_editor
{my $col_data=$_[0];
my $col_editor=$_[1];
my $cell_editor=$sheet2->get_cell(1,$col_editor);
my $editor=$cell_editor->unformatted();
my $cell_data=$sheet2->get_cell(1,$col_data);
my $data=$cell_data->unformatted();
my $paperNumber=1;
my $save=0;
open (F,">>data.txt")|| die "couldn't open data.txt!\n";
```



```

for my $row (2..$row_max1){
    my $cell_editorNext=$sheet2->get_cell($row,$col_editor);
    if (!defined $cell_editorNext){last;}
    my $editor_next=$cell_editorNext->unformatted();
    my $cell_dataNext=$sheet2->get_cell($row,$col_data);
    if (!defined $cell_dataNext){next;}
    my $data_next=$cell_dataNext->unformatted();
    if ($editor eq $editor_next){
        $data+=$data_next;
        $paperNumber++;}
    else {
        if($paperNumber!=0){$save=$data/$paperNumber;}
        print F "$editor; $data; $paperNumber; $save;\n";
        $data=$data_next;
        $paperNumber=1;
        $editor=$editor_next;
    }
}
print F "$editor; $data; $paperNumber; $save;\n";
close F;
}

```