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 WebOfKnowledge. 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
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Associate Editor</td>
</tr>
<tr>
<td>EIC</td>
<td>Editor in Chief</td>
</tr>
<tr>
<td>IF</td>
<td>Impact Factor</td>
</tr>
<tr>
<td>IES</td>
<td>Industrial Electronics Society</td>
</tr>
<tr>
<td>TIE</td>
<td><em>IEEE Trans. on Industrial Electronics</em></td>
</tr>
<tr>
<td>TII</td>
<td><em>IEEE Trans. on Industrial Informatics</em></td>
</tr>
<tr>
<td>IEM</td>
<td><em>IEEE Industrial Electronics Magazine</em></td>
</tr>
</tbody>
</table>
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:

\[
A = \text{number of citations of articles published in 2009 and 2010 during 2011.}
\]

\[
B = \text{the total number of articles published in 2009 and 2010 by that journal.}
\]

\[
2011 \text{ impact factor} = \frac{A}{B}.
\]

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

<table>
<thead>
<tr>
<th></th>
<th>TIE</th>
<th>TII</th>
<th>IEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of citations to 2008 papers</td>
<td>2121</td>
<td>62</td>
<td>31</td>
</tr>
<tr>
<td>Number of citations to 2009 papers</td>
<td>1220</td>
<td>48</td>
<td>29</td>
</tr>
<tr>
<td>Number of citations to 2008 &amp; 2009 papers</td>
<td><strong>3341</strong></td>
<td><strong>110</strong></td>
<td><strong>60</strong></td>
</tr>
<tr>
<td>Number of papers published in 2008</td>
<td>454</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>Number of papers published in 2009</td>
<td>505</td>
<td>39</td>
<td>17</td>
</tr>
<tr>
<td>Number of papers published in 2008 &amp; 2009</td>
<td><strong>959</strong></td>
<td><strong>67</strong></td>
<td><strong>32</strong></td>
</tr>
<tr>
<td>IF Impact Factor</td>
<td><strong>3.48</strong></td>
<td><strong>1.64</strong></td>
<td><strong>1.87</strong></td>
</tr>
</tbody>
</table>

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.

![One year IF trends for IES journals](image)

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

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). 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 cp
d 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:

```perl
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);
```
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”.

11
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cities</td>
<td>Authors</td>
<td>Title</td>
<td>Year</td>
<td>ArticleURL</td>
<td>CiteseURL</td>
<td>Author Institution</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>FL Lian, J</td>
<td>Network architecture and cor 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>National Taiwan Uni</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>AP Kaloge</td>
<td>Vertical integration of enterpr 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Porto</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>M Barranc</td>
<td>An active star topology for im 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Porto</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>G Cena...</td>
<td>On the properties of the flexi 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Porto</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>85</td>
<td>JLM Lastra</td>
<td>Semantic web services in fac 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>Tampere University</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>48</td>
<td>T Skeie, S</td>
<td>Timeliness of real-time IP coi 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Porto</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>Q Morante</td>
<td>Pervasive grid for large-scale 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Samnic</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>C Ryan, D</td>
<td>Interactive consistency on a 1206</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Limerick</td>
<td>Richard, Sauter, Thilo</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>Manuscript Title</td>
<td>Manuscript Type</td>
<td>Author Names</td>
<td>Submission Date</td>
<td>Decision Date</td>
<td>Decision</td>
<td>Editor Full Name</td>
</tr>
</tbody>
</table>
| 125 | On the Use of Wireless Networks at Low Level of Factory Automation Systems | Regular Paper | Villegas, Darío, De Felice, Francesco | Jan 9, 800, 5:36:34 AM | Feb 19, 2006, 1:45:26 PM | Accept | Zarawski, Richard
| 126 | An Internet-Based Distributed Manufacturing System Utilizing a Recommanded Network Rate Controller for Applying and Adapting the USC 5334 Standard to Industrial Automation | Regular Paper | Liu, Wilfred | Jan 12, 2006, 3:00:32 PM | Mar 25, 2006, 9:28:45 AM | Major Revision | Zarawski, Richard

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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cities</td>
<td>Authors</td>
<td>Title</td>
<td>Year</td>
<td>ArticleURL</td>
<td>CiteseURL</td>
<td>Author Institution</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>FL Lian, J</td>
<td>Network architecture and cor 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>National Taiwan Uni</td>
<td>Zarawski, Richard</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>G Cena...</td>
<td>On the properties of the flexi 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Porto</td>
<td>Zarawski, Richard Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>85</td>
<td>JLM Lastra</td>
<td>Semantic web services in fac 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>Tampere University</td>
<td>Zarawski, Richard Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>Q Morante</td>
<td>Pervasive grid for large-scale 2006</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Samnic</td>
<td>Zarawski, Richard Sauter, Thilo</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>C Ryan, D</td>
<td>Interactive consistency on a 1206</td>
<td><a href="http://ieeex">http://ieeex</a> <a href="http://scho">http://scho</a>...</td>
<td>University of Limerick</td>
<td>Zarawski, Richard Sauter, Thilo</td>
<td></td>
</tr>
</tbody>
</table>

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

12
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.

```perl
sub match
    {$_[0]=$_[0];
     $_[1]=$_[1];
     $_[0]=~s/\W+/ /;
     $_[0]=~s/\W+$/;
     $_[1]=~s/\W+/ /;
     $_[1]=~s/\W+$/;
     if (lc($_[0]) eq lc($_[1])){
        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

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

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.

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 10\textsuperscript{th} 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 10\textsuperscript{th}, 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.
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.
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.

![Table of Content](image)

**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 inconsistence 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.
Open "Journal_citation.xls" and "Journal_ManuscriptReceived.xls"

For every entry in "Journal_citation.xls", assign the paper title to variable $title1

For every entry in "Journal_ManuscriptReceived.xls" check whether the title matches $title1

Match?

Yes

Record the row number of the matching entry to variable $LastEntry; Check the decision state.

No

Accepted?

Yes

Break the inner for loop;

No

Extract the Acceptance Date Info from the row $LastEntry

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:

```perl
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

<table>
<thead>
<tr>
<th>AE num</th>
<th># of Papers</th>
<th># of cit.</th>
<th>Citations/quarter</th>
<th>Citations/pap/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE# 050</td>
<td>1</td>
<td>12</td>
<td>6.00</td>
<td>24.00</td>
</tr>
<tr>
<td>AE# 009</td>
<td>24</td>
<td>1709</td>
<td>104.21</td>
<td>17.37</td>
</tr>
<tr>
<td>AE# 029</td>
<td>8</td>
<td>378</td>
<td>31.78</td>
<td>15.89</td>
</tr>
<tr>
<td>AE# 054</td>
<td>34</td>
<td>1701</td>
<td>121.33</td>
<td>14.27</td>
</tr>
<tr>
<td>AE# 024</td>
<td>8</td>
<td>309</td>
<td>28.05</td>
<td>14.02</td>
</tr>
<tr>
<td>AE# 001</td>
<td>11</td>
<td>312</td>
<td>36.77</td>
<td>13.37</td>
</tr>
<tr>
<td>AE# 037</td>
<td>11</td>
<td>433</td>
<td>34.12</td>
<td>12.41</td>
</tr>
<tr>
<td>AE# 041</td>
<td>5</td>
<td>205</td>
<td>15.32</td>
<td>12.26</td>
</tr>
<tr>
<td>AE# 088</td>
<td>28</td>
<td>762</td>
<td>85.38</td>
<td>12.20</td>
</tr>
<tr>
<td>AE# 031</td>
<td>1</td>
<td>44</td>
<td>2.93</td>
<td>11.73</td>
</tr>
<tr>
<td>AE# 043</td>
<td>5</td>
<td>152</td>
<td>14.51</td>
<td>11.61</td>
</tr>
<tr>
<td>AE# 076</td>
<td>16</td>
<td>470</td>
<td>45.22</td>
<td>11.30</td>
</tr>
<tr>
<td>AE# 008</td>
<td>7</td>
<td>224</td>
<td>19.35</td>
<td>11.06</td>
</tr>
<tr>
<td>AE# 063</td>
<td>27</td>
<td>541</td>
<td>72.60</td>
<td>10.76</td>
</tr>
<tr>
<td>AE# 086</td>
<td>7</td>
<td>121</td>
<td>18.63</td>
<td>10.65</td>
</tr>
<tr>
<td>AE# 094</td>
<td>2</td>
<td>89</td>
<td>5.24</td>
<td>10.47</td>
</tr>
<tr>
<td>AE# 061</td>
<td>21</td>
<td>423</td>
<td>53.77</td>
<td>10.24</td>
</tr>
<tr>
<td>AE# 010</td>
<td>15</td>
<td>292</td>
<td>37.63</td>
<td>10.04</td>
</tr>
<tr>
<td>AE# 057</td>
<td>7</td>
<td>129</td>
<td>17.46</td>
<td>9.98</td>
</tr>
<tr>
<td>AE# 051</td>
<td>25</td>
<td>567</td>
<td>61.85</td>
<td>9.90</td>
</tr>
<tr>
<td>AE# 044</td>
<td>7</td>
<td>97</td>
<td>17.25</td>
<td>9.85</td>
</tr>
<tr>
<td>AE#</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>AE# 052</td>
<td>4</td>
<td>95</td>
<td>9.73</td>
<td>9.73</td>
</tr>
<tr>
<td>AE# 002</td>
<td>16</td>
<td>361</td>
<td>38.30</td>
<td>9.58</td>
</tr>
<tr>
<td>AE# 012</td>
<td>20</td>
<td>680</td>
<td>45.33</td>
<td>9.07</td>
</tr>
<tr>
<td>AE# 102</td>
<td>2</td>
<td>17</td>
<td>4.25</td>
<td>8.50</td>
</tr>
<tr>
<td>AE# 084</td>
<td>19</td>
<td>314</td>
<td>39.93</td>
<td>8.41</td>
</tr>
<tr>
<td>AE# 046</td>
<td>8</td>
<td>137</td>
<td>16.71</td>
<td>8.36</td>
</tr>
<tr>
<td>AE# 073</td>
<td>11</td>
<td>285</td>
<td>22.79</td>
<td>8.29</td>
</tr>
<tr>
<td>AE# 064</td>
<td>9</td>
<td>139</td>
<td>18.50</td>
<td>8.22</td>
</tr>
<tr>
<td>AE# 069</td>
<td>2</td>
<td>21</td>
<td>4.00</td>
<td>8.00</td>
</tr>
<tr>
<td>AE# 042</td>
<td>2</td>
<td>68</td>
<td>3.87</td>
<td>7.73</td>
</tr>
<tr>
<td>AE# 027</td>
<td>31</td>
<td>627</td>
<td>59.67</td>
<td>7.70</td>
</tr>
<tr>
<td>AE# 090</td>
<td>9</td>
<td>204</td>
<td>17.18</td>
<td>7.64</td>
</tr>
<tr>
<td>AE# 096</td>
<td>11</td>
<td>118</td>
<td>20.40</td>
<td>7.42</td>
</tr>
<tr>
<td>AE# 055</td>
<td>14</td>
<td>228</td>
<td>25.94</td>
<td>7.41</td>
</tr>
<tr>
<td>AE# 058</td>
<td>5</td>
<td>45</td>
<td>9.20</td>
<td>7.36</td>
</tr>
<tr>
<td>AE# 070</td>
<td>13</td>
<td>254</td>
<td>23.68</td>
<td>7.29</td>
</tr>
<tr>
<td>AE# 095</td>
<td>16</td>
<td>279</td>
<td>28.71</td>
<td>7.18</td>
</tr>
<tr>
<td>AE# 062</td>
<td>23</td>
<td>318</td>
<td>40.89</td>
<td>7.11</td>
</tr>
<tr>
<td>AE# 033</td>
<td>20</td>
<td>318</td>
<td>35.10</td>
<td>7.02</td>
</tr>
<tr>
<td>AE# 066</td>
<td>17</td>
<td>284</td>
<td>29.67</td>
<td>6.98</td>
</tr>
<tr>
<td>AE# 038</td>
<td>31</td>
<td>445</td>
<td>53.31</td>
<td>6.88</td>
</tr>
<tr>
<td>AE# 087</td>
<td>13</td>
<td>283</td>
<td>22.13</td>
<td>6.81</td>
</tr>
<tr>
<td>AE# 018</td>
<td>3</td>
<td>65</td>
<td>5.07</td>
<td>6.76</td>
</tr>
<tr>
<td>AE# 078</td>
<td>15</td>
<td>303</td>
<td>25.24</td>
<td>6.73</td>
</tr>
<tr>
<td>AE# 007</td>
<td>13</td>
<td>288</td>
<td>21.85</td>
<td>6.72</td>
</tr>
<tr>
<td>AE# 019</td>
<td>13</td>
<td>280</td>
<td>21.76</td>
<td>6.70</td>
</tr>
<tr>
<td>AE# 003</td>
<td>7</td>
<td>145</td>
<td>11.66</td>
<td>6.67</td>
</tr>
<tr>
<td>AE# 098</td>
<td>13</td>
<td>269</td>
<td>21.48</td>
<td>6.61</td>
</tr>
<tr>
<td>AE# 059</td>
<td>12</td>
<td>166</td>
<td>19.81</td>
<td>6.60</td>
</tr>
<tr>
<td>AE# 015</td>
<td>3</td>
<td>61</td>
<td>4.81</td>
<td>6.42</td>
</tr>
<tr>
<td>AE# 083</td>
<td>3</td>
<td>21</td>
<td>4.77</td>
<td>6.36</td>
</tr>
<tr>
<td>AE# 077</td>
<td>7</td>
<td>117</td>
<td>10.95</td>
<td>6.26</td>
</tr>
<tr>
<td>AE# 092</td>
<td>4</td>
<td>88</td>
<td>6.20</td>
<td>6.20</td>
</tr>
<tr>
<td>AE# 099</td>
<td>7</td>
<td>151</td>
<td>10.79</td>
<td>6.16</td>
</tr>
<tr>
<td>AE# 013</td>
<td>14</td>
<td>158</td>
<td>21.37</td>
<td>6.11</td>
</tr>
<tr>
<td>AE# 040</td>
<td>5</td>
<td>74</td>
<td>7.55</td>
<td>6.04</td>
</tr>
<tr>
<td>AE# 045</td>
<td>4</td>
<td>12</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>AE# 049</td>
<td>1</td>
<td>6</td>
<td>1.50</td>
<td>6.00</td>
</tr>
<tr>
<td>AE# 075</td>
<td>4</td>
<td>61</td>
<td>6.00</td>
<td>6.00</td>
</tr>
<tr>
<td>AE# 103</td>
<td>2</td>
<td>8</td>
<td>3.00</td>
<td>6.00</td>
</tr>
<tr>
<td>AE# 060</td>
<td>15</td>
<td>183</td>
<td>22.22</td>
<td>5.92</td>
</tr>
<tr>
<td>AE# 080</td>
<td>5</td>
<td>131</td>
<td>7.36</td>
<td>5.89</td>
</tr>
<tr>
<td>AE# 004</td>
<td>15</td>
<td>221</td>
<td>22.05</td>
<td>5.88</td>
</tr>
<tr>
<td>AE#</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>77</td>
<td>14.62</td>
<td>5.85</td>
</tr>
<tr>
<td>026</td>
<td>5</td>
<td>181</td>
<td>7.22</td>
<td>5.78</td>
</tr>
<tr>
<td>039</td>
<td>18</td>
<td>290</td>
<td>25.63</td>
<td>5.70</td>
</tr>
<tr>
<td>035</td>
<td>12</td>
<td>164</td>
<td>16.99</td>
<td>5.66</td>
</tr>
<tr>
<td>068</td>
<td>7</td>
<td>59</td>
<td>9.69</td>
<td>5.54</td>
</tr>
<tr>
<td>020</td>
<td>8</td>
<td>120</td>
<td>11.00</td>
<td>5.50</td>
</tr>
<tr>
<td>056</td>
<td>11</td>
<td>123</td>
<td>14.77</td>
<td>5.37</td>
</tr>
<tr>
<td>022</td>
<td>1</td>
<td>20</td>
<td>1.33</td>
<td>5.33</td>
</tr>
<tr>
<td>017</td>
<td>4</td>
<td>32</td>
<td>5.26</td>
<td>5.26</td>
</tr>
<tr>
<td>085</td>
<td>3</td>
<td>55</td>
<td>3.88</td>
<td>5.17</td>
</tr>
<tr>
<td>005</td>
<td>14</td>
<td>159</td>
<td>17.50</td>
<td>5.00</td>
</tr>
<tr>
<td>011</td>
<td>3</td>
<td>11</td>
<td>3.58</td>
<td>4.78</td>
</tr>
<tr>
<td>053</td>
<td>22</td>
<td>260</td>
<td>26.24</td>
<td>4.77</td>
</tr>
<tr>
<td>081</td>
<td>1</td>
<td>19</td>
<td>1.19</td>
<td>4.75</td>
</tr>
<tr>
<td>067</td>
<td>1</td>
<td>20</td>
<td>1.18</td>
<td>4.71</td>
</tr>
<tr>
<td>093</td>
<td>2</td>
<td>5</td>
<td>2.33</td>
<td>4.67</td>
</tr>
<tr>
<td>079</td>
<td>2</td>
<td>36</td>
<td>2.30</td>
<td>4.60</td>
</tr>
<tr>
<td>091</td>
<td>4</td>
<td>56</td>
<td>4.42</td>
<td>4.42</td>
</tr>
<tr>
<td>048</td>
<td>3</td>
<td>43</td>
<td>3.00</td>
<td>3.99</td>
</tr>
<tr>
<td>089</td>
<td>3</td>
<td>20</td>
<td>2.93</td>
<td>3.91</td>
</tr>
<tr>
<td>071</td>
<td>3</td>
<td>41</td>
<td>2.93</td>
<td>3.90</td>
</tr>
<tr>
<td>032</td>
<td>5</td>
<td>13</td>
<td>4.83</td>
<td>3.87</td>
</tr>
<tr>
<td>097</td>
<td>10</td>
<td>113</td>
<td>8.80</td>
<td>3.52</td>
</tr>
<tr>
<td>030</td>
<td>4</td>
<td>14</td>
<td>3.33</td>
<td>3.33</td>
</tr>
<tr>
<td>028</td>
<td>5</td>
<td>32</td>
<td>4.00</td>
<td>3.20</td>
</tr>
<tr>
<td>047</td>
<td>6</td>
<td>45</td>
<td>4.52</td>
<td>3.02</td>
</tr>
<tr>
<td>072</td>
<td>1</td>
<td>10</td>
<td>0.75</td>
<td>3.00</td>
</tr>
<tr>
<td>074</td>
<td>3</td>
<td>22</td>
<td>2.21</td>
<td>2.95</td>
</tr>
<tr>
<td>006</td>
<td>1</td>
<td>5</td>
<td>0.67</td>
<td>2.67</td>
</tr>
<tr>
<td>016</td>
<td>3</td>
<td>47</td>
<td>2.00</td>
<td>2.67</td>
</tr>
<tr>
<td>034</td>
<td>1</td>
<td>8</td>
<td>0.67</td>
<td>2.67</td>
</tr>
<tr>
<td>025</td>
<td>4</td>
<td>12</td>
<td>2.03</td>
<td>2.03</td>
</tr>
<tr>
<td>023</td>
<td>1</td>
<td>2</td>
<td>0.50</td>
<td>2.00</td>
</tr>
<tr>
<td>101</td>
<td>2</td>
<td>5</td>
<td>1.00</td>
<td>2.00</td>
</tr>
<tr>
<td>014</td>
<td>11</td>
<td>13</td>
<td>4.83</td>
<td>1.76</td>
</tr>
<tr>
<td>065</td>
<td>1</td>
<td>5</td>
<td>0.42</td>
<td>1.67</td>
</tr>
<tr>
<td>036</td>
<td>3</td>
<td>5</td>
<td>0.82</td>
<td>1.10</td>
</tr>
<tr>
<td>021</td>
<td>4</td>
<td>4</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>082</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 4 Citation Analysis for Papers processed by Different Associate Editors in TII

<table>
<thead>
<tr>
<th>AE num</th>
<th># of Papers</th>
<th># of cit.</th>
<th>Citations /quarter</th>
<th>Citations /pap/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE #07</td>
<td>3</td>
<td>253</td>
<td>14.32</td>
<td>19.09</td>
</tr>
<tr>
<td>AE #50</td>
<td>6</td>
<td>168</td>
<td>16.84</td>
<td>11.23</td>
</tr>
<tr>
<td>AE #15</td>
<td>3</td>
<td>101</td>
<td>6.86</td>
<td>9.14</td>
</tr>
<tr>
<td>AE #55</td>
<td>5</td>
<td>143</td>
<td>10.69</td>
<td>8.55</td>
</tr>
<tr>
<td>AE #32</td>
<td>1</td>
<td>27</td>
<td>1.80</td>
<td>7.20</td>
</tr>
<tr>
<td>AE #58</td>
<td>6</td>
<td>201</td>
<td>10.64</td>
<td>7.04</td>
</tr>
<tr>
<td>AE #31</td>
<td>2</td>
<td>55</td>
<td>3.44</td>
<td>6.88</td>
</tr>
<tr>
<td>AE #43</td>
<td>7</td>
<td>164</td>
<td>10.21</td>
<td>5.83</td>
</tr>
<tr>
<td>AE #49</td>
<td>2</td>
<td>37</td>
<td>2.33</td>
<td>4.65</td>
</tr>
<tr>
<td>AE #53</td>
<td>2</td>
<td>32</td>
<td>2.29</td>
<td>4.57</td>
</tr>
<tr>
<td>AE #37</td>
<td>6</td>
<td>45</td>
<td>6.81</td>
<td>4.54</td>
</tr>
<tr>
<td>AE #35</td>
<td>3</td>
<td>58</td>
<td>3.26</td>
<td>4.34</td>
</tr>
<tr>
<td>AE #10</td>
<td>1</td>
<td>16</td>
<td>1.07</td>
<td>4.27</td>
</tr>
<tr>
<td>AE #01</td>
<td>1</td>
<td>3</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>AE #27</td>
<td>1</td>
<td>4</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>AE #30</td>
<td>1</td>
<td>2</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>AE #48</td>
<td>1</td>
<td>2</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>AE #26</td>
<td>1</td>
<td>14</td>
<td>0.93</td>
<td>3.73</td>
</tr>
<tr>
<td>AE #52</td>
<td>2</td>
<td>28</td>
<td>1.87</td>
<td>3.73</td>
</tr>
<tr>
<td>AE #57</td>
<td>2</td>
<td>5</td>
<td>1.75</td>
<td>3.50</td>
</tr>
<tr>
<td>AE #05</td>
<td>11</td>
<td>75</td>
<td>9.06</td>
<td>3.30</td>
</tr>
<tr>
<td>AE #23</td>
<td>4</td>
<td>23</td>
<td>3.28</td>
<td>3.28</td>
</tr>
<tr>
<td>AE #18</td>
<td>13</td>
<td>92</td>
<td>10.42</td>
<td>3.21</td>
</tr>
<tr>
<td>AE #14</td>
<td>1</td>
<td>4</td>
<td>0.80</td>
<td>3.20</td>
</tr>
<tr>
<td>AE #25</td>
<td>1</td>
<td>7</td>
<td>0.78</td>
<td>3.11</td>
</tr>
<tr>
<td>AE #24</td>
<td>2</td>
<td>8</td>
<td>1.50</td>
<td>3.00</td>
</tr>
<tr>
<td>AE #02</td>
<td>5</td>
<td>24</td>
<td>3.69</td>
<td>2.95</td>
</tr>
<tr>
<td>AE #03</td>
<td>8</td>
<td>41</td>
<td>5.62</td>
<td>2.81</td>
</tr>
<tr>
<td>AE #34</td>
<td>2</td>
<td>7</td>
<td>1.40</td>
<td>2.80</td>
</tr>
<tr>
<td>AE #36</td>
<td>10</td>
<td>40</td>
<td>6.86</td>
<td>2.74</td>
</tr>
<tr>
<td>AE #28</td>
<td>1</td>
<td>10</td>
<td>0.67</td>
<td>2.67</td>
</tr>
<tr>
<td>AE #21</td>
<td>9</td>
<td>42</td>
<td>5.58</td>
<td>2.48</td>
</tr>
<tr>
<td>AE #16</td>
<td>2</td>
<td>6</td>
<td>1.20</td>
<td>2.40</td>
</tr>
<tr>
<td>AE #44</td>
<td>1</td>
<td>5</td>
<td>0.56</td>
<td>2.22</td>
</tr>
<tr>
<td>AE #41</td>
<td>2</td>
<td>6</td>
<td>1.02</td>
<td>2.04</td>
</tr>
<tr>
<td>AE #56</td>
<td>4</td>
<td>31</td>
<td>2.03</td>
<td>2.03</td>
</tr>
<tr>
<td>AE #09</td>
<td>1</td>
<td>5</td>
<td>0.45</td>
<td>1.82</td>
</tr>
<tr>
<td>AE #45</td>
<td>8</td>
<td>22</td>
<td>3.60</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>AE #39</td>
<td>4</td>
<td>7</td>
<td>1.75</td>
<td>1.75</td>
</tr>
<tr>
<td>AE #42</td>
<td>1</td>
<td>6</td>
<td>0.43</td>
<td>1.71</td>
</tr>
<tr>
<td>AE #04</td>
<td>2</td>
<td>2</td>
<td>0.67</td>
<td>1.33</td>
</tr>
<tr>
<td>AE #46</td>
<td>1</td>
<td>3</td>
<td>0.33</td>
<td>1.33</td>
</tr>
<tr>
<td>AE #54</td>
<td>3</td>
<td>2</td>
<td>0.41</td>
<td>1.33</td>
</tr>
<tr>
<td>AE #22</td>
<td>1</td>
<td>3</td>
<td>0.30</td>
<td>1.20</td>
</tr>
<tr>
<td>AE #33</td>
<td>3</td>
<td>2</td>
<td>0.50</td>
<td>0.67</td>
</tr>
<tr>
<td>AE #51</td>
<td>2</td>
<td>2</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>AE #38</td>
<td>2</td>
<td>3</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>AE #19</td>
<td>3</td>
<td>1</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>AE #06</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #08</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #11</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #12</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #13</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #17</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #20</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #29</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #40</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>AE #47</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Table 5 Citation Analysis for Paper processed by Different EICs in TIE (Grouped by years)

<table>
<thead>
<tr>
<th>year</th>
<th>EIC num</th>
<th># of paper</th>
<th>citation</th>
<th>cites/quarter</th>
<th>cites/paper/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>EIC #1</td>
<td>70</td>
<td>3641</td>
<td>219.2</td>
<td>12.52</td>
</tr>
<tr>
<td>2007</td>
<td>EIC #1</td>
<td>200</td>
<td>6109</td>
<td>433.7</td>
<td>8.67</td>
</tr>
<tr>
<td>2008</td>
<td>EIC #1</td>
<td>299</td>
<td>6573</td>
<td>575.04</td>
<td>7.69</td>
</tr>
<tr>
<td>2009</td>
<td>EIC #2</td>
<td>1</td>
<td>9</td>
<td>1,285714</td>
<td>5.14</td>
</tr>
<tr>
<td>2009</td>
<td>EIC #1</td>
<td>379</td>
<td>4742</td>
<td>626.47</td>
<td>6.61</td>
</tr>
<tr>
<td>2010</td>
<td>EIC #3</td>
<td>1</td>
<td>2</td>
<td>0.33</td>
<td>1.32</td>
</tr>
<tr>
<td>2010</td>
<td>EIC #4</td>
<td>93</td>
<td>343</td>
<td>86.5</td>
<td>3.72</td>
</tr>
<tr>
<td>2010</td>
<td>EIC #5</td>
<td>7</td>
<td>34</td>
<td>4</td>
<td>2.28</td>
</tr>
<tr>
<td>2010</td>
<td>EIC #2</td>
<td>3</td>
<td>19</td>
<td>3.06</td>
<td>4.08</td>
</tr>
<tr>
<td>2010</td>
<td>EIC #1</td>
<td>207</td>
<td>1648</td>
<td>366.05</td>
<td>7.07</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #3</td>
<td>23</td>
<td>50</td>
<td>28</td>
<td>4.86</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #4</td>
<td>75</td>
<td>198</td>
<td>81.5</td>
<td>4.34</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #5</td>
<td>22</td>
<td>32</td>
<td>13.5</td>
<td>2.45</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #2</td>
<td>23</td>
<td>68</td>
<td>34</td>
<td>5.91</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #6</td>
<td>7</td>
<td>11</td>
<td>5.5</td>
<td>3.14</td>
</tr>
<tr>
<td>2011</td>
<td>EIC #1</td>
<td>18</td>
<td>96</td>
<td>24</td>
<td>5.33</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>EIC num</th>
<th># of paper</th>
<th>Citation</th>
<th>cites/quarter</th>
<th>cites/paper/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIC #1</td>
<td>1173</td>
<td>22809</td>
<td>2244.46</td>
<td>47.91</td>
</tr>
<tr>
<td>EIC #2</td>
<td>27</td>
<td>96</td>
<td>38.34</td>
<td>15.13</td>
</tr>
<tr>
<td>EIC #3</td>
<td>24</td>
<td>52</td>
<td>28.33</td>
<td>6.18</td>
</tr>
<tr>
<td>EIC #4</td>
<td>168</td>
<td>541</td>
<td>168</td>
<td>8.06</td>
</tr>
<tr>
<td>EIC #5</td>
<td>29</td>
<td>66</td>
<td>17.5</td>
<td>4.74</td>
</tr>
<tr>
<td>EIC #6</td>
<td>7</td>
<td>11</td>
<td>5.5</td>
<td>3.14</td>
</tr>
</tbody>
</table>
Table 7 Citation Analysis for Papers processed by Different EICs in TII
( Grouped by years )

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

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

<table>
<thead>
<tr>
<th>EIC num</th>
<th># of paper</th>
<th>Citation</th>
<th>cites/quarter</th>
<th>cites/paper/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIC #1</td>
<td>162</td>
<td>1590</td>
<td>145.2</td>
<td>22.97</td>
</tr>
<tr>
<td>EIC #2</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>0.88</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>SS Name</th>
<th>printed</th>
<th>citations /pap/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS on Wireless Technologies in Facto</td>
<td>2007/2</td>
<td>7.99</td>
</tr>
<tr>
<td>SS on Wireless Technologies in Facto</td>
<td>2007/3</td>
<td>6.99</td>
</tr>
<tr>
<td>SS on Factory Communication System</td>
<td>2005/3</td>
<td>5.79</td>
</tr>
<tr>
<td>SS on Communication in Automation</td>
<td>2006/2</td>
<td>4.87</td>
</tr>
<tr>
<td>SS on Real-Time Systems part 2</td>
<td>2009/1</td>
<td>3.96</td>
</tr>
<tr>
<td>SS on Real-Time and (Networked) En</td>
<td>2009/3</td>
<td>3.18</td>
</tr>
<tr>
<td>SS on Communication in Automation</td>
<td>2008/2</td>
<td>2.35</td>
</tr>
<tr>
<td>SS on In-Vehicle Embedded Systems</td>
<td>2009/4</td>
<td>2.11</td>
</tr>
<tr>
<td>SS on Industrial Control</td>
<td>2010/1</td>
<td>2.09</td>
</tr>
<tr>
<td>SS on Communication in Automation</td>
<td>2009/2</td>
<td>1.97</td>
</tr>
<tr>
<td>SS on Industrial Communication Syst</td>
<td>2010/3</td>
<td>1.88</td>
</tr>
<tr>
<td>SS on Real-Time Applications and To</td>
<td>2010/4</td>
<td>1.85</td>
</tr>
<tr>
<td>SS on Formal Methods in Manufacture</td>
<td>2010/2</td>
<td>1.70</td>
</tr>
<tr>
<td>SS on Real-Time and (Networked) En</td>
<td>2010/4</td>
<td>1.21</td>
</tr>
<tr>
<td>SS on Real-Time Systems Part 1</td>
<td>2008/4</td>
<td>1.03</td>
</tr>
<tr>
<td>SS on Power-Aware Computing</td>
<td>2010/3</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Table 10 Citation Analysis for Special Section Papers Published in TIE

<table>
<thead>
<tr>
<th>SS Name</th>
<th>printed</th>
<th>citations /pap/year</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS on Automotive</td>
<td>2008/6</td>
<td>23.64</td>
</tr>
<tr>
<td>SS on Digital Ecosystems and Cyber</td>
<td>2011/6</td>
<td>15.82</td>
</tr>
<tr>
<td>SS on Predictive Control of Power Eng</td>
<td>2009/6</td>
<td>11.74</td>
</tr>
<tr>
<td>SS on Photovoltaic Power Processing</td>
<td>2008/7</td>
<td>11.63</td>
</tr>
<tr>
<td>SS on Industrial Wireless Sensor Net</td>
<td>2009/10</td>
<td>10.77</td>
</tr>
<tr>
<td>SS on FPGA</td>
<td>2007/4</td>
<td>10.02</td>
</tr>
<tr>
<td>SS on Advances in Electrical Machine</td>
<td>2006/12</td>
<td>9.70</td>
</tr>
<tr>
<td>SS on FPGA</td>
<td>2008/4</td>
<td>9.60</td>
</tr>
<tr>
<td>SS on Voltage and Current Control</td>
<td>2009/2</td>
<td>9.41</td>
</tr>
<tr>
<td>SS on Application and Control of Dc</td>
<td>2009/10</td>
<td>9.20</td>
</tr>
<tr>
<td>SS on Multilevel Inverters</td>
<td>2010/8</td>
<td>9.03</td>
</tr>
<tr>
<td>SS on Active and Hybrid Filters to En</td>
<td>2009/8</td>
<td>8.27</td>
</tr>
<tr>
<td>SS on Efficient and reliable photovoltaic</td>
<td>2009/11</td>
<td>6.84</td>
</tr>
<tr>
<td>SS on Renewable Energy Systems</td>
<td>2011/1</td>
<td>6.68</td>
</tr>
<tr>
<td>SS on Education</td>
<td>2007/6</td>
<td>6.45</td>
</tr>
<tr>
<td>SS on Energy Harvesting</td>
<td>2010/3</td>
<td>6.37</td>
</tr>
<tr>
<td>SS on Predictive Control of Power Eng</td>
<td>2008/12</td>
<td>6.25</td>
</tr>
<tr>
<td>SS on Hardware-in-the-loop</td>
<td>2010/4</td>
<td>6.17</td>
</tr>
<tr>
<td>SS on Advances in Electrical Machine</td>
<td>2010/1</td>
<td>6.12</td>
</tr>
<tr>
<td>SS on Fuel cells power</td>
<td>2010/6</td>
<td>5.89</td>
</tr>
<tr>
<td>SS on Applications of Embedded Systems</td>
<td>2011/3</td>
<td>5.88</td>
</tr>
<tr>
<td>SS on Plug-in hybrid electric vehicle</td>
<td>2010/2</td>
<td>5.86</td>
</tr>
<tr>
<td>SS on Sliding Mode Control in Indus</td>
<td>2008/11</td>
<td>5.70</td>
</tr>
<tr>
<td>SS on Radio Frequency Identification</td>
<td>2009/7</td>
<td>5.23</td>
</tr>
<tr>
<td>SS on Thermal Issues in Electrical Mach</td>
<td>2008/10</td>
<td>5.15</td>
</tr>
<tr>
<td>SS on Education</td>
<td>2008/6</td>
<td>4.79</td>
</tr>
<tr>
<td>SS on Advances in Motion Control</td>
<td>2009/10</td>
<td>4.63</td>
</tr>
<tr>
<td>SS on Automotive Power and Energy</td>
<td>2010/3</td>
<td>4.59</td>
</tr>
<tr>
<td>SS on Education</td>
<td>2009/12</td>
<td>4.58</td>
</tr>
<tr>
<td>SS on Advances in Electrical Machine</td>
<td>2009/11</td>
<td>4.15</td>
</tr>
<tr>
<td>SS on Sliding Mode Control in Indus</td>
<td>2009/9</td>
<td>4.11</td>
</tr>
<tr>
<td>SS on Diagnostics of EMPED</td>
<td>2011/5</td>
<td>3.66</td>
</tr>
<tr>
<td>SS on Energy Storage System</td>
<td>2010/12</td>
<td>3.44</td>
</tr>
<tr>
<td>SS on Electronic Devices and System</td>
<td>2011/7</td>
<td>3.33</td>
</tr>
<tr>
<td>SS on Building Automation, Control</td>
<td>2010/11</td>
<td>3.17</td>
</tr>
<tr>
<td>SS on Switched Reluctance Machine</td>
<td>2010/9</td>
<td>2.80</td>
</tr>
<tr>
<td>SS on Wireless Technology</td>
<td>2010/5</td>
<td>2.55</td>
</tr>
<tr>
<td>SS on Advances in Microelectromechanics</td>
<td>2009/4</td>
<td>2.54</td>
</tr>
<tr>
<td>SS on Education</td>
<td>2010/10</td>
<td>2.42</td>
</tr>
</tbody>
</table>
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.

![Graph showing average time between submission and the first decision for TIE and TII.](image)

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 $TII_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 $TII_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.

```perl
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);
            $title_match=@array2[0];
        } else{
            my @array2=split(/"$/,$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+)/); 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 align="top">5.1.2\&nbsp;&nbsp;&nbsp;\&nbsp;&nbsp;&nbsp;\&nbsp;&nbsp;&nbsp;</td>\> Junyoung Heo, Jiman Hong, Yookun Cho,\&nbsp;&nbsp;"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 2\textsuperscript{nd} 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.

![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.](image-url)
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.

```perl
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".

```perl
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 $ave=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){$ave=$data/$paperNumber;}
       print F "$editor; $data; $paperNumber; $ave;\n";
       $data=$data_next;
       $paperNumber=1;
       $editor=$editor_next;
     }
   }
   print F "$editor; $data; $paperNumber; $ave;\n";
   close F;
  }
```

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.
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.

```perl
#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();
#push 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;
  }
}

47
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”.
<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>Year</td>
<td>Source</td>
<td>ArticleURL</td>
<td>CheURL</td>
<td>Paper Type</td>
<td>Submission</td>
<td>FirstDec</td>
<td>C</td>
<td>Author Inst</td>
<td>EIC</td>
<td>Full</td>
<td>N</td>
<td>Editor</td>
<td>Full</td>
<td>Issue</td>
</tr>
<tr>
<td>3</td>
<td>Network</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>May 04, 21</td>
<td>Jul 20, 20</td>
<td>Jul 20, 20</td>
<td>National T</td>
<td>Zurich, R</td>
<td>1</td>
<td>77</td>
<td>205</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Combining</td>
<td>2006</td>
<td>EEE</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>SS on Cor</td>
<td>Dec 08, 20</td>
<td>Jan 04, 20</td>
<td>Mar 17, 20</td>
<td>Facultade Zurich,</td>
<td>Sauter, T</td>
<td>2</td>
<td>99</td>
<td>54</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>Vertical im</td>
<td>2006</td>
<td>EE</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>SS on Cor</td>
<td>Dec 08, 20</td>
<td>Jan 04, 20</td>
<td>Mar 29, 20</td>
<td>University Zurich,</td>
<td>Sauter, T</td>
<td>2</td>
<td>111</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>An active</td>
<td>2006</td>
<td>EEE</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>SS on Cor</td>
<td>Dec 08, 20</td>
<td>Jan 04, 20</td>
<td>Mar 29, 20</td>
<td>University Zurich,</td>
<td>Sauter, T</td>
<td>2</td>
<td>111</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>On the pro</td>
<td>2006</td>
<td>Informal</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>SS on Cor</td>
<td>Dec 08, 20</td>
<td>Jan 04, 20</td>
<td>Mar 29, 20</td>
<td>University Zurich,</td>
<td>Sauter, T</td>
<td>2</td>
<td>111</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>Pervasive</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Dec 12, 20</td>
<td>Feb 19, 20</td>
<td>May 01, 20</td>
<td>University Zurich,</td>
<td>Gomes, L</td>
<td>3</td>
<td>140</td>
<td>101</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Interactive</td>
<td>2006</td>
<td>Informal</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Mar 14, 20</td>
<td>May 01, 20</td>
<td>Aug 08, 20</td>
<td>University Zurich,</td>
<td>Naedele, R</td>
<td>4</td>
<td>147</td>
<td>94</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Web-base</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>May 09, 20</td>
<td>Jul 20, 20</td>
<td>Oct 05, 20</td>
<td>NUT</td>
<td>Zurich,</td>
<td>Hs,</td>
<td>1</td>
<td>149</td>
<td>128</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>Energy-eff</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Jan 20, 20</td>
<td>Sep 14, 20</td>
<td>Dec 14, 20</td>
<td>Auburn Un</td>
<td>Zurich,</td>
<td>1</td>
<td>169</td>
<td>50</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Two-stage</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Dec 11, 20</td>
<td>Feb 20, 20</td>
<td>Sep 02, 20</td>
<td>The Hong</td>
<td>Zurich,</td>
<td>1</td>
<td>223</td>
<td>170</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Development</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Dec 15, 20</td>
<td>Feb 19, 20</td>
<td>Sep 02, 20</td>
<td>The Hong</td>
<td>Zurich,</td>
<td>1</td>
<td>261</td>
<td>69</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Network tr</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Sep 23, 20</td>
<td>Nov 30, 20</td>
<td>Jun 14, 20</td>
<td>Auburn Un</td>
<td>Zurich,</td>
<td>1</td>
<td>264</td>
<td>149</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>On the use</td>
<td>2006</td>
<td>Industrial</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>May 20, 20</td>
<td>Dec 14, 20</td>
<td>Feb 19, 20</td>
<td>111</td>
<td>Zurich,</td>
<td>1</td>
<td>275</td>
<td>80</td>
<td>208</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Service con</td>
<td>2006</td>
<td>Informal</td>
<td>kce</td>
<td>explore</td>
<td><a href="http://ieeexplore.ieee.org">http://ieeexplore.ieee.org</a></td>
<td>Sep 02, 20</td>
<td>Mar 07, 20</td>
<td>Jun 08, 20</td>
<td>Tampere U</td>
<td>Zurich,</td>
<td>1</td>
<td>279</td>
<td>155</td>
<td>186</td>
<td></td>
</tr>
</tbody>
</table>

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


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

going use Spreadsheet::ParseExcel;
use Spreadsheet::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-&gt;get_cell($row1,1);
if(!defined $cell_title1)
{next;}
my $title_match=$cell_title1-&gt;unformatted();
if (match($title,$title_match))
{
  if ($FirstEntry==0)
  {
    $FirstEntry=1;
    my $cell_SubDate=$Page2_2-&gt;get_cell($row1,4);
    if (defined $cell_SubDate){
      my $SubDate=$cell_SubDate-&gt;value();
      $worksheet-&gt;AddCell($row,9,$SubDate);
    }
    my $cell_FirstDecisionDate=$Page2_2-&gt;get_cell($row1,5);
    if(defined $cell_FirstDecisionDate){
      my $FirstDecisionDate=$cell_FirstDecisionDate-&gt;value();
      $worksheet-&gt;AddCell($row,10,$FirstDecisionDate);
    }
  }
  $LastMatchRow=$row1;
  my $cell_Decision=$Page2_2-&gt;get_cell($row1,6);
  if(!defined $cell_Decision){next;}
  my $Decision=$cell_Decision-&gt;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-&gt;get_cell($row1,5);
   if(defined $cell_DecisionDate){
     my $DecisionDate=$cell_DecisionDate-&gt;value();
     $worksheet-&gt;AddCell($row,11,$DecisionDate);
   }

   my $cell_Ins=$Page2_2-&gt;get_cell($row1,7);
   if(defined $cell_Ins){
     my $Ins=$cell_Ins-&gt;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 auguments, 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('/',$lines[$i]);
# my $title_match=@array1[1];
#
# if ($year==58){
#    my @array2=split('<VA>',$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;
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);

}
#***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;
  $start_date=~m/\s+(\d+)/;
  my ($start_month,$start_day)=($1,$2);
  $end_date=~m/\s+(\d+)/;
  my ($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 arguments, 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
#   @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 $ave=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;