

# **SITKIS: Software for Bibliometric Data Management and Analysis**

Version 1.0

## **MANUAL**

**Henri Schildt**

Helsinki University of Technology  
Institute of Strategy and International Business,  
Department of Industrial and Engineering Management

E-mail: [henri.schildt@hut.fi](mailto:henri.schildt@hut.fi)

Latest update: March 10<sup>th</sup>, 2004

## Table of Contents

Table of Contents .....	2
1 Introduction .....	3
2 Sitkis Package Contents.....	3
3 How to Install SITKIS .....	4
4 Exporting Data from ISI.....	7
5 Using SITKIS to read the citations.....	8
6 The Access Database .....	9
7 Using Article View Form.....	10
8 Analyzing Data .....	11
9 Using Multiple Databases .....	11
10 Robustness of the Results.....	11
11 Known Issues .....	13
12 How to Cite SITKIS .....	13
13 Acknowledgements & Licensing information .....	14

# 1 Introduction

This document contains information required to install and use citation data processing software “Sitkis”. It is a little Java software for most versions of Windows. The software imports ISI Web of Science files into a Microsoft Access database that can be modified easily. Sitkis also exports data from the database into UCINET compatible network graphs and Excel-compatible reports. The purpose of the program is to enable researchers to easily and quickly download and analyze bibliometric records.

I got interested in quantitative bibliometric analysis after reading several related articles, and most importantly because of the enthusiasm towards the methodology within my faculty. I had worked a little with social networks before (UCINET). However, what struck me was the amount of effort and time needed to gather and encode the citation data, which was effectively in a database, and thus in electronic format already. I decided to find out whether the data collection and manipulation could be made easier by automating it. This software is the result, and I have been adding new features one after another without a grand plan. Sitkis allows the researcher to quickly add and remove articles to and from the database, and most importantly easily running analyses for datasets with different sets of keywords and/or journals.

This file contains the installation instructions and instructions for importing data to Sitkis, as well as general related issues. Another document, “Sitkis User’s Guide 1.0.pdf” contains information regarding how to make various kinds of analyses. This division was originally made because most analyses combine the functionality of Sitkis with Ucinet and/or Excel. The guide also contains a list of further useful references.

# 2 Sitkis Package Contents

The package you have downloaded contains documentation, database, and program files:

citations_db.mdb	– Access database where your data will reside in.
Sitkis Manual 1.0.pdf	– This file. Installation instructions, etc.
Sitkis User’s Guide 1.0.pdf	– Instructions for doing analysis with Sitkis.
sitkis.bat	– The file you launch Sitkis with.
sitkisgui\...	– The program files
jung.jar	– Java networking package (see Chapter 13)
Return Me.doc	– A form I would like you to fill & return.

### 3 How to Install SITKIS

Installing SITKIS is easy. It consists of three things:

- (1) Unpack Sitkis zip file anywhere on your hard disk
- (2) Install Java Run Time Environment (1.3 or newer) and Microsoft Access
- (3) Set up the database in windows Administrative Tools

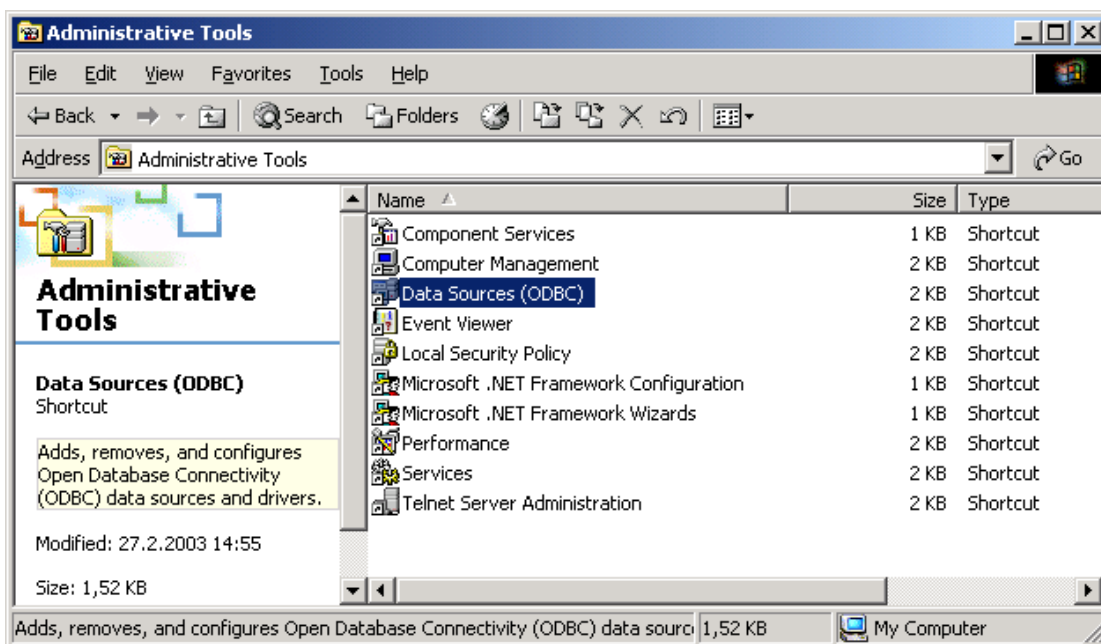
First, copy the contents of Sitkis zip-file into any place you like. It is easier if you copy them to some directory near the root (like C:\sitkis\).

Then you will need to install Microsoft Access (preferably 2000 or XP, Access 97 might work too) and Java Runtime Environment (JRE) version 1.3 or newer (available at [www.javasoft.com](http://www.javasoft.com)). It is highly possible that you have the both installed already.

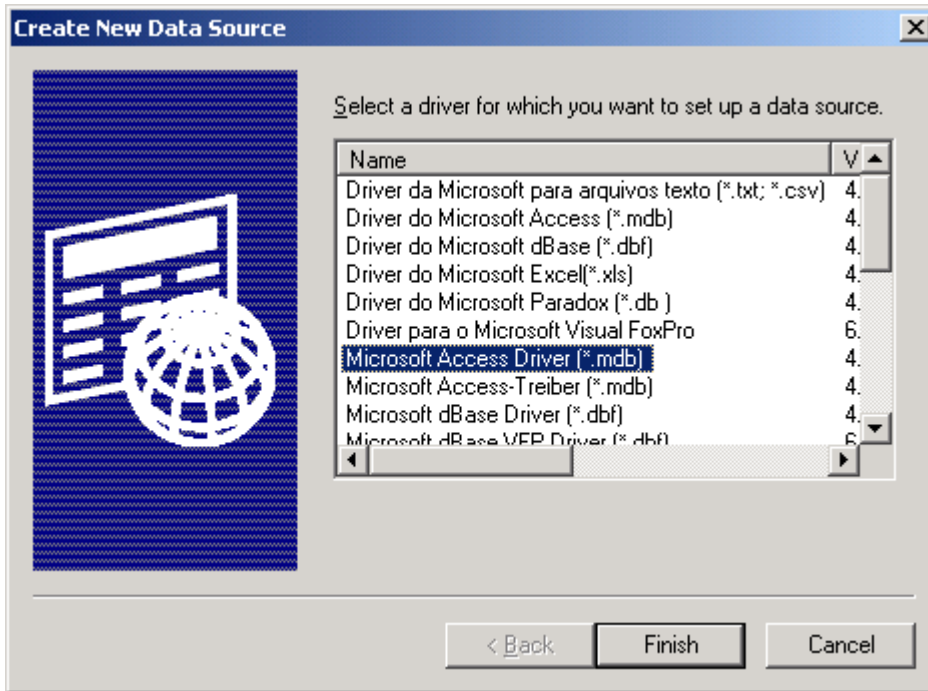
Then you need to set up the access database to your Windows ODBC registry. I advice you to first copy citations\_db.mdb file to a new name like "MyProject.mdb". This way you can keep an empty database and create copies for your various projects.

All you need to do is set up the database to Windows ODBC service and you are ready. It goes like this:

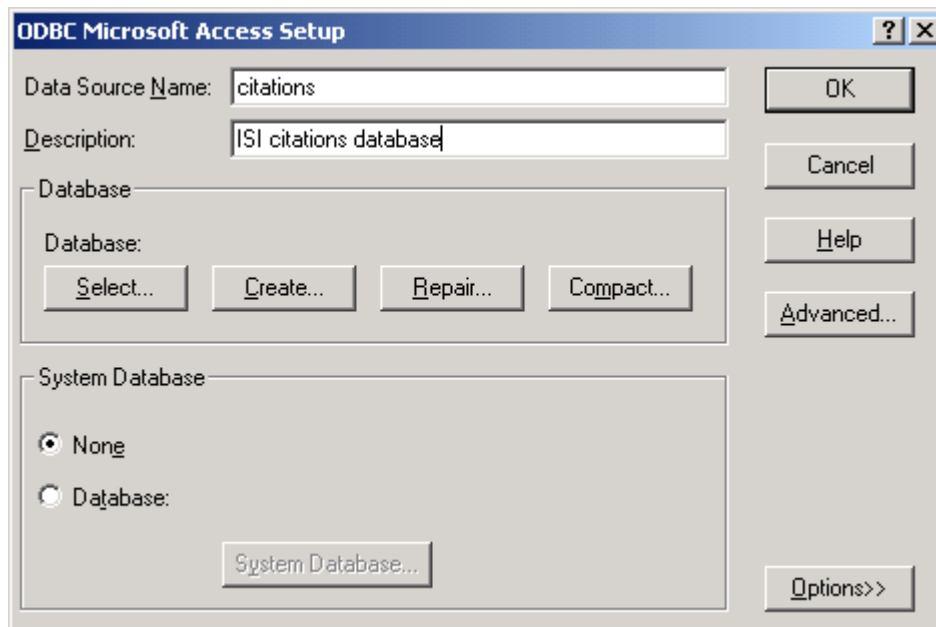
1. Go to control panel and open administrative tools



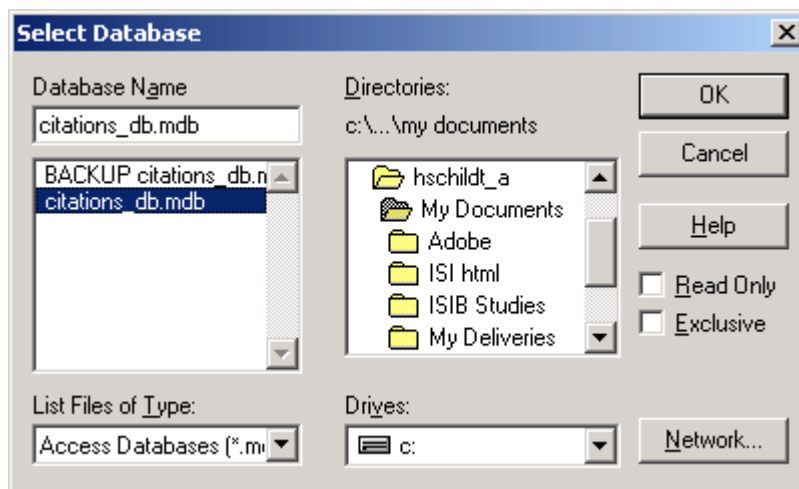
2. Select Data sources (ODBC)
3. Click on the button 'Add...' this brings the following window



4. Select Microsoft Access Driver (\*.mdb)
5. Next, enter the name of your database. It is easiest to use the name '*citations*', since it is the default database name in Sitkis. You can set up multiple names (by repeating the process) for different databases (.mdb files).



6. Click select... and choose the .mdb file included



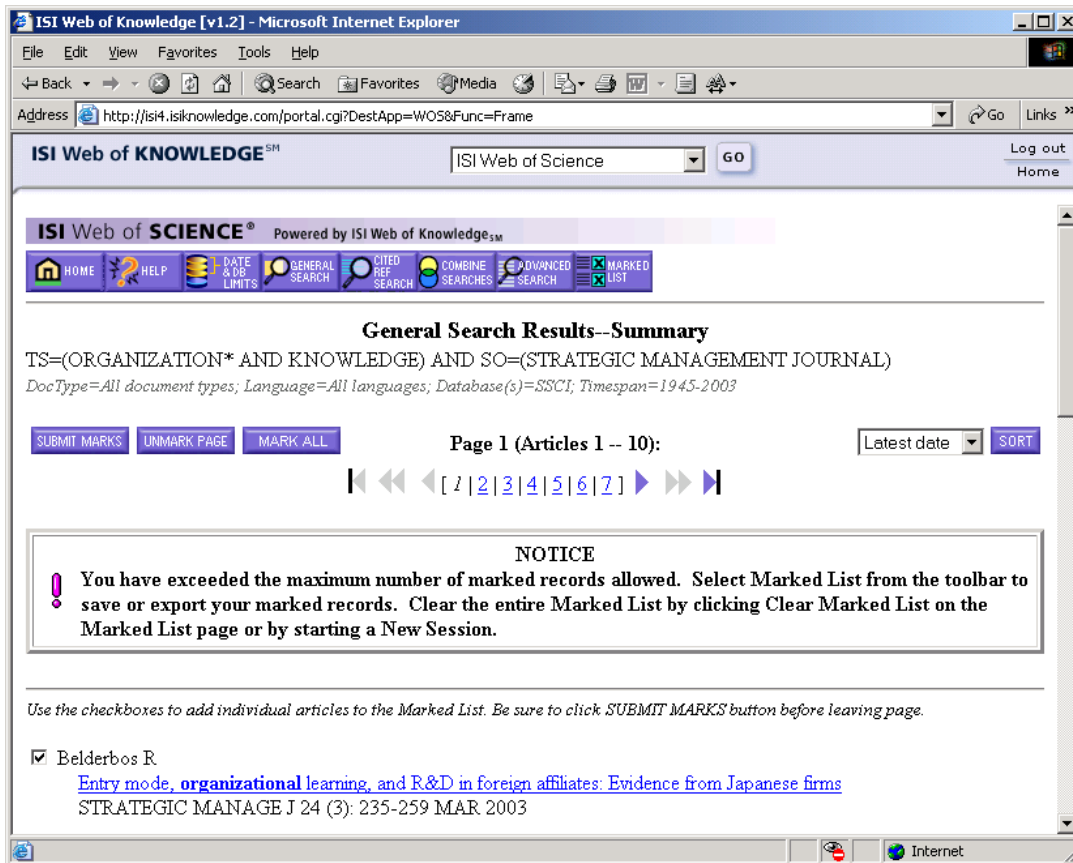
7. Now just click OK a couple of times and the database should be available via ODBC as required by Sitkis.

You have successfully installed the software! In some cases you may need to boot the computer after installing JRE (Java) even although the installer does not say so.

## 4 Exporting Data from ISI

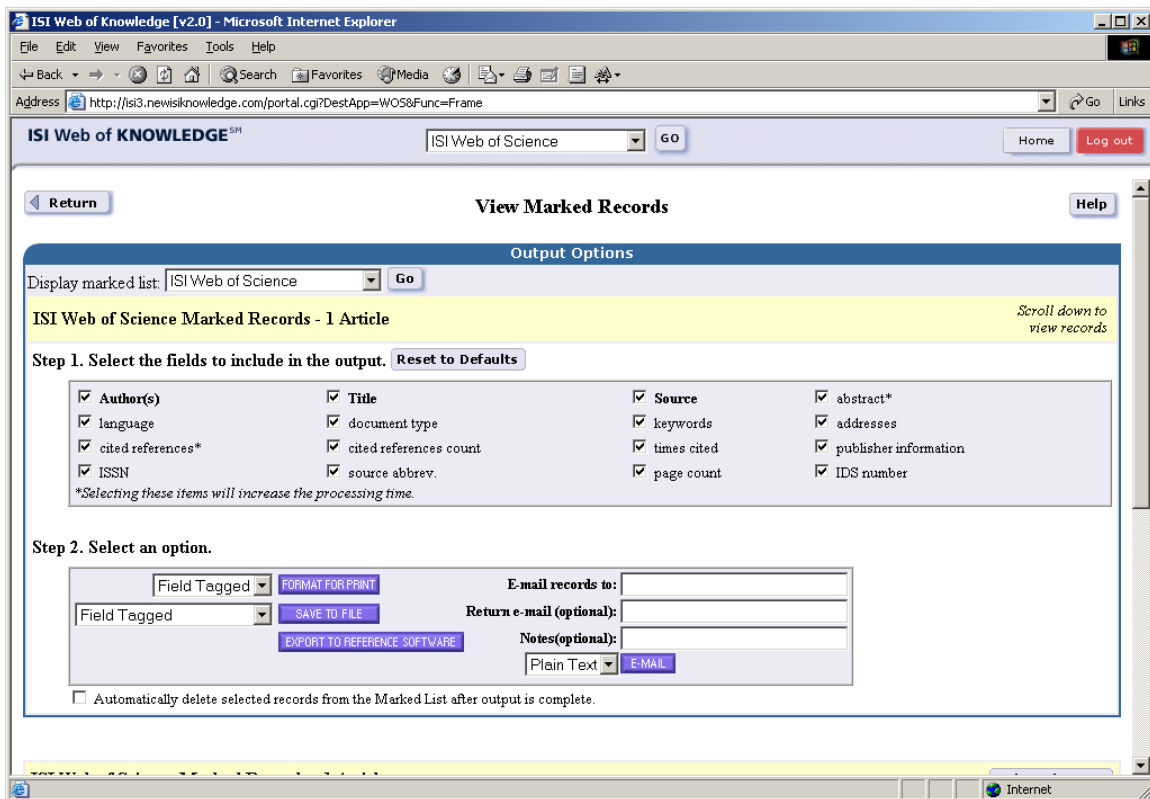
Exporting data from ISI web of knowledge / science is easy. You just conduct the search for articles as always, and then click 'Mark all' button to select articles for export (you can also manually check the articles you wish to use in your analysis).

After marking the articles click on "MARKED" button on the center top of the page.



Note that if you mark too many articles (more than 500) then you get a message like the one in window above! You need to clear previous marked articles in the mark articles-window.

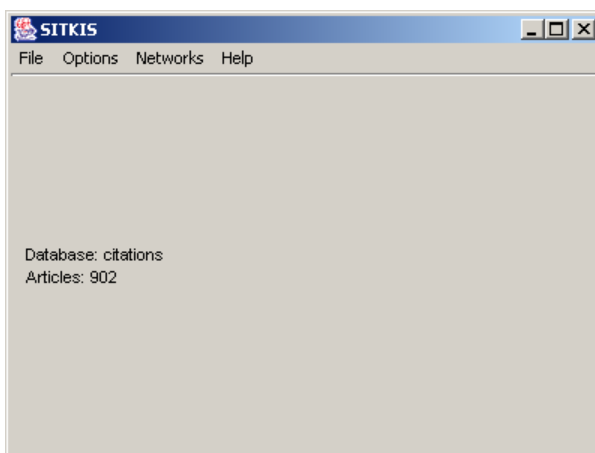
Next, you get a window like one below with all the articles you have marked so far. Check all the fields available and click on the 'Save to File' button.



## 5 Using SITKIS to read the citations

To run Sitkis, you need to locate the sitkis.bat (in the directory where you extracted the zip file). Then you just run it by double clicking it.

Now you should see the Sitkis program appear.





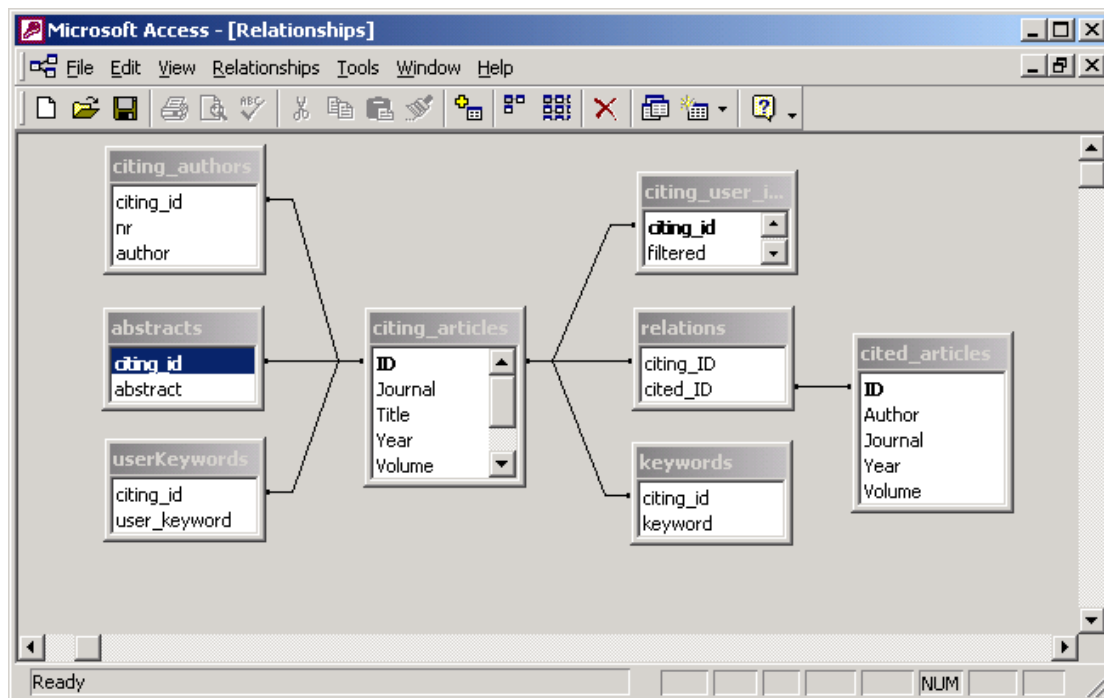
If you set up the name of the database as “citations” you can immediately start importing ISI records. If you used another name, set the name of the database from Options / Set Database (just enter the name you entered in Administrative Tools ODBC setup).

Then select File / Import. You will be asked to browse the file you saved from ISI Web of Science. Now the program appears to halt for a long period of time. You can see the progress from DOS command prompt. Once the file has been imported fully you will see the basic Sitkis window again. Now choose Options / Fix database. You should see now see a proper number of articles (not 0) in the window.

## 6 The Access Database

From this point forward you have multiple options. You can either do various operations with the data in Access, or just use Sitkis to create networks / analyses of the data.

The following table shows the relationships in the Access database.



You can import as many .txt files as you want to your database using Sitkis. You can also import the same articles multiple times without getting duplicate entries. All articles you have imported so far act as the material for analyses.

## 7 Using Article View Form

Once you have imported the data you can in Access and start a form labeled 'Citing Articles Information Editing' (click the Forms tab in the Access database view). The form is shown below.

**Microsoft Access - [citing\_articles1]**

File Edit View Insert Format Records Tools Window Help

**ID** **Journal**

1234 ACAD MANAGE J

**Title**

INTERNATIONAL EXPANSION STRATEGY OF JAPANESE FIRMS - CAPABILITY BUILDING THROUGH SEQUENTIAL ENTRY

**Year** **Volume** **Issue** **TC**

1995 38 2 59

**abstract**

This study examined the sequential entry process of Japanese electronic manufacturing firms into the United States during the period 1976-89. The firms first entered their core businesses and those in which they had a strong competitive advantage over local firms to reduce the hazard of failure. Learning from early entries enabled them to launch further entries into noncore businesses and into areas of weaker competitive advantage. The overall evidence suggests that Japanese firms are building capabilities to operate overseas through sequential entry.

**filtered**

**User Comments:**

**citing\_authors subform1**

nr	author
▶	0 CHANG SJ
*	0

Record: 1 of 1

**keywords subform1**

keyword
▶ FOREIGN DIRECT-INVESTM
UNITED-STATES
JOINT VENTURES
DIVERSIFICATION
OPTION
VIEW
*

Record: 1 of 6

Record: 20 of 512

Form View NUM

You can edit any information in the form. The check box ‘filtered’ is intended to allow removal of some (non-related) articles from the analyses. User comments field simply saves comments related to the particular article in the database.

The form records are sorted by publication and year, but it is possible to change sorting method (this is for advanced users – see Access documentation).

## 8 Analyzing Data

Please refer to Sitkis\_analyses.pdf for all information on how to analyze the data you have gathered.

## 9 Using Multiple Databases

It is advisable to use multiple databases if you want to analyze different kinds of topics. It is easy to create a new database. You just need to:

- (1) Copy the citations\_db.mdb (make sure it is the empty database that came in the .zip!) to a new name of your choice
- (2) Record the new .mdb file to ODBC (in Control Panel / Administrative Tools), using a new name of your choice for the database
- (3) When you use Sitkis, make sure to select Options / Set database each time before you import / export any data. When you have set the name of the database, all operations will use the database you have defined. If you misspell the name of the database, nothing will happen (see console).

## 10 Robustness of the Results

ISI Web of Science entries have lots of errors regarding to citations. If you want to aim at perfection, you need to do two things: look at citation entries that could not be processed by SITKIS, and match similar citation entries with different spelling. The latter is much more important and should probably be done to at least the most commonly cited authors if you are aiming to publish your bibliometric research.

### 10.1 Discarded citations entries

First thing you can do is to check table “*discarded\_citations*”. These are entries that could not be properly interpreted by Sitkis. You can manually enter those citations into cited\_articles table, with automatically created keyword and then link them to articles by inserting an entry in the table “*relations*”.

An example helps here. Let's assume you have the following line in `discarded_citations` table:

```
citing_id   discarded
58          CHEUNG SNS, J LAW ECON, V26, P1
```

You can easily notice that the year for the article is missing from ISI database, but the reference should probably be included in your study. You then add the entry to `cited_articles`. Let's say you have found out the year is 1992. Notice that the ID is automatically created. You don't enter it!

```
ID    Author    Journal    Year Volume
319    CHEUNG SNS  J LAW ECON  1992   26
```

Next, you want to add the link from the citing article (ID 58) to the article it cited (now ID 319) by entering the following line in the table *relations*.

```
citing_ID  cited_ID
58          319
```

It was not exactly easy, but not that difficult either. You may usually want to ignore this kind of errors if you have a large number of articles, and you are not aiming to publish your bibliometric results. It takes time to correct the mistakes and they do not usually change your findings substantially. In case you wonder, it does not matter if you have the same cited article stored with multiple different ID numbers, the ID numbers are used in *relations* table, but all analysis queries utilize author name, journal, year, and volume.

## 10.2 Connecting Similar Entries

There are various spelling mistakes in ISI Web of Science citation information, and these can cause a lot of distortion in your results. The mistakes are related to spelling of the authors' names, volume numbers and years of publications. These can be checked easily in table "cited\_articles" and corrected by hand. I suggest that if you have a very large number of cited articles, identify your top 20-50 most cited authors and check them only.

Go to table `cited_articles` in Access and sort them by author. For example the entries for author Levinthal are:

<b>ID</b>	<b>Author</b>	<b>Journal</b>	<b>Year</b>	<b>Volume</b>
2225	LEVINTHAL D	J ECON BEHAV ORGAN	1981	2
780	LEVINTHAL D	DECISIONS ORG	1988	0
2226	LEVINTHAL D	STRATEGIC MANAGEMENT	1993	14
673	LEVINTHAL DA	STRATEGIC MANAGE J	1993	14
1753	LEVINTHAL DA	EVOLUTIONARY DYNAMIC	1994	0
1056	LEVINTHAL DA	ORGAN SCI	1999	10

961 LEVINTHAL DA

MANAGE SCI

1997 43

It is easy to notice that entries with ID number 2226 and 673 are the same article. It is not uncommon that there are substantial numbers of articles in ISI with multiple entries, effectively distorting the perceived article influence in your analyses. It is necessary to make entries 2226 and 673 identical. By changing all these author names to "LEVINTHAL DA", you will also get correct author-specific citation counts.

## 11 Known Issues

- Business press articles (e.g. Los Angeles Times, News Week) without author field in ISI are ignored. Also citations that have author beginning with "\*" are ignored at this stage (these are mostly conference papers, etc.). You can find these in *discarded\_citations* table.
- IN PRESS articles are problematic. Currently the system discards articles that are marked as being "in press" or "forthcoming".
- The same problems appear with doctoral theses and those citations where year or comma before the year is missing.
- If you want to make *sure* that none of the in print / thesis (without specified year) citations are not missing, you have to check the discarded table with key-words "THESIS" and "IN PRINT" and insert those into access manually (you will also need to find out the year from somewhere!). Easier way could be to actually modify the ISI export file in notepad and rerun the software. However, if you open & save the export file with MS Word, it changes line feeds to linefeed & carriage return (Unix txt file to windows txt file), and SITKIS won't be able to process it. I suggest you use some notepad-like freeware editor.

It is advisable NOT to delete the export files because, (i) there may be updates to SITKIS that can make more out of those files, (ii) there may be updates to SITKIS database like additional queries, (iii) you may want to send them to your friend, (iv) it's better be safe than sorry.

## 12 How to Cite SITKIS

If you use SITKIS and find it useful in conducting your literature review, I will greatly appreciate if you cite the software in a footnote. The following is an example citation that can be adjusted to suit publication guidelines, the hyperlink being optional:

Schildt, H.A. 2002. SITKIS: Software for Bibliometric Data Management and Analysis v0.6.1. Helsinki: Institute of Strategy and International Business. [Available at: [www.hut.fi/~hschildt/sitkis](http://www.hut.fi/~hschildt/sitkis)]

## 13 Acknowledgements & Licensing information

I would like to acknowledge the financial support from the VCR project I worked at while made this program. The project was funded by the Research Programme for Advanced Technology Policy (ProACT) of the Ministry of Trade and Industry and the National Technology Agency, Tekes.

I would also like to thank my fellow researchers at the Institute of Strategy and International Business who have been involved in testing the software and giving valuable suggestions related to both the software and this documentation.

Sitkis utilizes JUNG (Java Universal Network/Graph Framework) open source package, which is licensed under Berkeley Software Distribution (BSD) license. For more information on JUNG, please see <http://jung.sourceforge.net/>. Future versions of Sitkis will hopefully take the full advantage of graphical features provided by Jung.