feature

Citation searches in on-line databases: possibilities and pitfalls

Damon D. Ridley*
School of Chemistry, University of Sydney, Sydney, N.S.W. 2006, Australia

Until recently the main source of scientific journal citations was the science citation index, which is now available in print and electronic formats. Searches in the electronic format were performed mainly by information professionals, but with the introduction of citations in the chemical abstracts service bibliographic database and the acceptance of the Web of Science as a means to access citations, many scientists and managers are searching directly for citation data. For comprehensive results the most reliable method is to use search terms based on first-listed authors of each publication, but this may be very time consuming. Electronic sources provide search algorithms which fast-track the process, but it is important that users understand the issues and how these algorithms work if reasonably reliable data are to be obtained.

Keywords: Science citation index; Chemical abstracts service; Citations; Citation analysis; On-line searching

1. Introduction

Citations allow researchers to retrieve related documents [1–5], and allow researchers and managers to assess the relative significance of individual publications [6–9]. Accordingly it is important that end-users are familiar with the limitations and idiosyncrasies of the various citation searching methods. On the basis of a practical example, this article examines key issues when searching leading citation databases.

Citation analyses have been questioned both with respect to validity of data and the accuracy with which these data are retrieved [10–13]. Authors cite, or do not cite, papers for a whole host of reasons. For example, the article 'Electrochemically induced nuclear fusion of deuterium' by Fleischmann, Pons and Hawkins in the Journal of Electroanalytical Chemistry (1989) has been cited over 600 times. It could be argued this is an important article that has stimulated much research and the high citation rate is a reflection of its importance, but some would suggest its level of citations is due to other factors. However, the various reasons why papers are cited in the long run tend to balance out, so that most analysts agree that the cautious interpretation of citation data over a period of time provides a valuable indicator. Still there are many traps for the unwary and this article works through them, using as an example the tracing of citations to articles by Paul R. Haddad, University of Tasmania, a key researcher in the area of analytical chemistry.

Most commonly, scientists turn to the ISI science citation index for data on citations. The laborious search through the print version has now mostly been replaced by electronic searching through a variety of on-line networks or through internet-based products like Web of Science. The search algorithms that have been built to help overcome some search difficulties can in turn cause new problems and care must be exercised to ensure that they have provided the most comprehensive data.

On the STN network, the SCISEARCH database covers the literature from almost 6000 journals since 1974. It currently has almost 19 million records, most of which list the citations in the original articles. Parts of a sample record are given in Fig. 1. Recently the Chemical Abstracts Service (CAS) introduced citations into its bibliographic

*E-mail: d.ridley@chem.usyd.edu.au

0165-9936/01/$ – see front matter
PII: S0165-9936(00)00057-1

1

```
SCISEARCH            COPYRIGHT 2000 ISI (R)
ACCESSION NUMBER:    2000:359306  SCISEARCH
TITLE:               On-capillary ion-exchange preconcentration of inorganic
                     anions using open-tubular capillaries followed by elution with a
                     transient isotachophoretic gradient
AUTHOR:              Breadmore M C; Boyce M C; Macka M; Avdalovic N; Haddad P R
                    (Reprint)
REFERENCE(S):
   Referenced Author   |Year | VOL | PG   | Referenced Work
         (RAU)         |(RPY)|(RVL)|(RPG) |      (RWK)
=======================+=====+=====+======+==================
ARCE L                 |1999 |390  |39    |ANAL CHIM ACTA
ARCE L                 |1997 |791  |279   |J CHROMATOGR A
AVDALOVIC N            |     |     |      |UNPUB ANALYST
BURGI D S              |1997 |     |479   |HDB CAPILLARY ELECTR…
```

2

```
HCAPLUS              COPYRIGHT  2000 ACS
ACCESSION NUMBER:    2000:293546  CAPLUS
TITLE:               On-capillary ion-exchange preconcentration of
                     inorganic anions using open-tubular capillaries
                     followed by elution with a transient isotachophoretic
                     gradient
AUTHOR(S):           Breadmore, Michael C.; Macka, Miroslav; Haddad,
                     Paul R.; Boyce, Mary C.; Avdalovic, Nebojsa
REFERENCE COUNT:     16
REFERENCE(S):       (1) Arce, L; Anal Chim Acta 1999, V390, P39 CAPLUS
                    (2) Arce, L; J Chromatogr A 1997, V791, P279 CAPLUS
                    (3) Avdalovic, N; Analyst, submitted
                    (4) Burgi, D; Handbook of Capillary Electrophoresis, 2nd edn
                        1997, P479 CAPLUS…
```

3

```
=> E HADDAD P R, 1994/RE 8
E1           1     HADDAD P R, 1993, V250, P21, ANAL CHIM ACTA/RE
E2          19     HADDAD P R, 1993, V640, P135, J CHROMATOGR/RE
E3           0 --> HADDAD P R, 1994/RE
E4           1     HADDAD P R, 1994, 6TH INT S HIGH PERF/RE
E5           1     HADDAD P R, 1994, P932, IEEE ANT PROP S/RE
E6           1     HADDAD P R, 1994, P932, IEEE AP S INT S/RE
E7           1     HADDAD P R, 1994, V2, P932, IEEE APS/RE
E8           1     HADDAD P R, 1994, V30, ELECTRONICS LETT/RE
```

4

```
=> E HADDAD P/AU 10
E1           4     HADDAD O M/AU
E2           1     HADDAD O S/AU
E3          86 --> HADDAD P/AU
E4           1     HADDAD P A/AU
E5           1     HADDAD P C/AU
E6           5     HADDAD P F/AU
E7           2     HADDAD P J/AU
E8          10     HADDAD P M/AU
E9         211     HADDAD P R/AU
E10          6     HADDAD P S/AU
```

Fig. 1. Part of a record from the SCISEARCH file.

Fig. 2. Part of a record from the HCAPLUS file. The representations of authors and references should be compared with those in Fig. 1.

Fig. 3. Entries for citations in the database may be displayed through the EXPAND command in the reference field, and here some of the listings for first author HADDAD P R (1994) are shown.

Fig. 4. EXPAND in the author field lists entries for authors. Actual records should be checked to ensure only (and all) relevant records are retrieved.

←

database, but the implementation is only on the STN network[1]. Citations are listed from the beginning of 1999 for articles in almost 1300 core journals, about 5300 partly indexed journals, and examiner citations from 4 patent issuing authorities. So the coverage in the sciences compares reasonably with the journals covered in the science citation index since 1999, though patent coverage is exclusive to CAS. Parts of a sample record in the HCAPLUS file are given in Fig. 2. The main points to note (Figs. 1 and 2) in relation to citation searching are the different representations of author names in the author fields, and of the initials for authors in the reference fields.

The search results reported here were obtained from the respective databases on 12 July 2000.

2. Science citation index

In the reference field, the science citation index lists the first named author only, plus the original article name, publication year, volume and page number given in the citation. When citing articles authors may use a variety of journal names and abbreviations, so it is necessary to EXPAND (that is display a list) of possibilities and to check which ones are relevant. For example it is to be noted (Fig. 3) that there is one record for each of the separate entries E5, E6, and E7 although they are references to the same article, and the implications of this will be explained shortly.

This process is very similar to the process of searching citations through the hardcopy. While it may be a little more convenient than the print version, it can also be laborious – particularly when it is necessary to EXPAND on all the first named authors where many publications are involved. In appreciation of this, producers of electronic sources provide options where citations may be retrieved for any authors irrespective of whether or not they are the first-mentioned author in the original publication. The technique is first to find records for the author and then to ask the computer to search automatically for citations to the identified records. In order to use this technique to find citations for Paul R. Haddad, it is necessary to identify search terms for the author by EXPANDing in the author field (Fig. 4). It is noted that the database lists the surname and initials only, and while the 211 entries (E9) are most likely relevant, it is possible that some of the entries under E3 (HADDAD P) may be relevant also. To be sure one needs to check with the author or look through the entries indicated under E3 or E9.

Next the command TRA CIT L1 1- is employed, where the 211 records in set E9 (which have been obtained as an answer set called L1) are now searched for citations, producing an answer set, L3, of 1605 records (Fig. 5). The command TRA CIT L1 1- is literally a request to 'transfer citations from answer number 1 onwards in L1' but in effect it is finding possible terms for citations of all the answers in L1 and then searching them automatically. To do this the system re-looks at the 211 answers in L1, then re-formats them so the first author, the publication year, volume number and page number are used as search terms in the field which contains the list of cited references. That is, the system first creates an intermediate answer set (L2) and Fig. 6 shows some of the 211 terms which have been searched to create answer set L3. Part of one of the 1605 answers is shown in Fig. 7, in which the article by ReynekeBarnard et al. gives a refer-

[1] It is more economical to perform citation searches in the HCAPLUS File since many search terms are required, though citations are implemented on the CA/CAPLUS family of files.

```
5


=> TRA CIT L1 1-
L2          TRANSFER L1 1- CIT :     211 TERMS
L3          1605 L2


6

=> D L2 1-
L2              TRA L1 1- CIT :      211 TERMS

TERM # TERMS
------ -------------------------

         1 ALEXANDER P W, 1981, V209, P29,?/RE              …
         2 ALEXANDER P W, 1984, V17, P309,?/RE              44 HADDAD P R, 1974, V21, P123,?/RE
         3 ALEXANDER P W, 1984, V56, P2417,?/RE             45 HADDAD P R, 1974, V21, P859,?/RE
         4 ALEXANDER P W, 1985, V171, P151,?/RE             46 HADDAD P R, 1975, V22, P61,?/RE
         5 ALEXANDER P W, 1985, V177, P183,?/RE             47 HADDAD P R, 1976, V23, P275,?/RE
         6 ALEXANDER P W, 1985, V18, P1953,?/RE             48 HADDAD P R, 1977, V24, P1,?/RE
         7 ALEXANDER P W, 1985, V20, P179,?/RE              49 HADDAD P R, 1982, V252, P177,?/RE
         8 ALEXANDER P W, 1992, V589, P201,?/RE             50 HADDAD P R, 1982, V5, P853,?/RE…
         9 BAGCHI R, 1986, V351, P541,?/RE…


7

=> D L3 TI AU HIT

L3   ANSWER 1 OF 1639  SCISEARCH  COPYRIGHT 2000 ISI (R)
TI   N,N'-pentamethylenethiuram disulfide- and N,N'-pentamethylenethiuram
     hexasulfide-accelerated sulfur vulcanization. I. Interaction of curatives
     in the absence of rubber
AU   ReynekeBarnard C P; Gradwell M H S; McGill W J (Reprint)
   Referenced Author   |Year | VOL | PG   | Referenced Work
         (RAU)         |(RPY)|(RVL)|(RPG) |      (RWK)
======================+=====+=====+======+====================
HUTCHINS S R           |1982 |252  |185   |J CHROMATOGR          <--


8

=> E HADDAD P R, 1990/RE 14
…
E4          27     HADDAD P R, 1990, ION CHROMATOGRAPHY/RE
E5           1     HADDAD P R, 1990, ION CHROMATOGRAPHY A/RE
E6          89     HADDAD P R, 1990, ION CHROMATOGRAPHY P/RE
E7           1     HADDAD P R, 1990, ION CHROMATOGRAPHYS/RE
E8           1     HADDAD P R, 1990, ION CHROMATOGRAPY PR/RE
E9           2     HADDAD P R, 1990, ION CHROMATOGRPAHY P/RE
E10          1     HADDAD P R, 1990, J CHROMATOGRAPHY LIB/RE
E11          2     HADDAD P R, 1990, P103, ION CHROMATOGRAPHY P/RE
E12          2     HADDAD P R, 1990, P124, ION CHROMATOGRAPHY P/RE
E13          1     HADDAD P R, 1990, P133, ION CHROMATOGRAPHY/RE
E14          5     HADDAD P R, 1990, P133, ION CHROMATOGRAPHY P/RE
```

Fig. 5. A search on E9 (Fig. 4) produces an answer set (L1) of 211 records for the author. The TRAnsfer command selects out the first author (and some journal source information), and then automatically searches the terms (L2) to produce an answer set (here for 1605 citations).

Fig. 6. Some of the 211 terms selected through the TRAnsfer command (Fig. 5). It is critical to check this list against the author's publication list (if known).

Fig. 7. Parts of the most recent citation from the answer set L3 (Fig. 5).

Fig. 8. EXPAND in the reference field shows citations listings for the reference work by Haddad and Jackson. Because the database does not have a record for this work, the search in Fig. 5 did not retrieve these citations.

ence to the article by Hutchins, Haddad and Dilli in the Journal of Chromatography (1982).

There are many points to note:

- the answer set L1 contains records for the required author in the database irrespective of name order;
- the intermediate set L2 looks at the records in L1 and selects out the first author since this should be the author listed in the citation field;
- the intermediate set L2 contains the publication year, volume, and page number for the record in the database; and
- the intermediate set L2 does not contain the journal title because of the many ways it could have been abbreviated in the original publication and because generally name, volume and page numbers are sufficient to give precise answers.

This raises the following questions. First, did answer set L1 contain all the relevant records for the author (i.e. have relevant citations been missed)? Second, did the answer set L1 contain records for different authors with the same surname and initials (i.e. does the initial answer set of citations contain incorrect records)? Third, did the citing author use the correct publication year, volume number, and first page number (i.e. did the search algorithm use the relevant terms)? To check the first two questions, it is necessary to compare the full list of selected entries L2 with the author's publication list. In this case the entry HADDAD P R did contain a different author with the same last name and initials. It is surprising how many different authors have the same surname and initials, and one should always check to see that different authors are not represented. The entries E5–E8 in Fig. 3 are for a different P.R. Haddad. It was noticed that L2 did not contain any listing for a number of publications, including the reference work, 'Ion Chromatography: Principles and Applications' by P.R. Haddad and P.E. Jackson (Elsevier, 1990). The reason for this is that there is no record for this book in the database, so the algorithm did not produce relevant search terms. Nevertheless, the database contains over 300 citations to this work! As a further complication, as various page numbers from reference works of this type are cited, there often are a large number of different listings (in this case there are over 100). An analysis of the information in Fig. 8, which lists some of the citations to the same reference text, gives an indication of the problem. It clearly is very important to check the terms that have been searched in L2 very carefully, and subsequently to add or delete entries. To check the third point it is necessary to return to the tedious process of EXPANDing on the individual entries. For example, each of the 211 terms in Fig. 6 may have to be EXPANDed as shown in Fig. 3.

In some instances, the problem of retrieval of citations to publications that are not present in the database (and hence which are not retrieved through the TRAnsfer command, Fig. 6) may be overcome by searching for authors in the referenced author field (RAU). For example as shown in Fig. 9, use of the TRAnsfer command gives 1605 citations, while searching in the RAU field for HADDAD P R gives 1004 citations. When these answer sets are combined and replicates eliminated, the result is a set of 1829 citations. It must be remembered that entries in the RAU field are for first authors only, and this can cause considerable difficulties when comprehensive results are required. In this case the book had P.R. Haddad as the first author and citations to the book were therefore retrieved. If the search had been for P.E. Jackson, this would not have been the case.

This answer set is a reasonable compromise. The first author problem for citations in the database has largely been overcome and quite a number of additional records have been obtained through a search in the RAU field. There are of course still issues of inclusion of false records, or exclusion of correct records, and of different ways in which citations are entered in the original article. To address these rigorously would require extensive EXPANDs in the RE or RAU fields. Perhaps another 10% of citations would be obtained, but then this probably lies within the 'error' relating to referencing in articles. Assuming that the 1829 records is a reasonable starting point, how does one find which publications are being cited most? The process is easy since the searcher simply asks the system to ANAlyze all the 1829 records and Fig. 10 illustrates the process and displays some of the most listed terms.

There are now two further points. First, the list of terms searched (L2) did not include variations in the author, publication year, volume and page number entries so some of those entries in Fig. 10 (especially that for the reference text for which the figure of 90 citations refers only to that particular way of citing the text) will be less than the actual number of citations. Second, the search in the RAU

9

```
=> S HADDAD P R/AU
L1          211 HADDAD P R/AU

=> TRA L1 1- CIT
L2        TRANSFER L1 1- CIT :     211 TERMS
L3        1605 L2

=> S HADDAD P R/RAU
L4        1004 HADDAD P R/RAU

=> S L3 OR L4
L5        1829 L3 OR L4
```

10

```
=> ANA L5 1- HIT RE
L6          ANALYZE L5 1- RE HIT :     403 TERMS

=> D OGT50
L6          ANALYZE L5 1- RE HIT :     403 TERMS

TERM #   # OCC  # DOC  % DOC RE
------ ------- ------ ------ ---------------
     1      90     89   4.87 HADDAD P R, 1990, ION CHROMATOGRAPHY P
     2      73     73   3.99 BUCHBERGER W, 1992, V608, P59, J CHROMATOGR
     3      59     59   3.23 HUTCHINS S R, 1982, V252, P185, J CHROMATOGR
     4      58     57   3.12 HADDAD P R, 1982, V252, P177, J CHROMATOGR
     5      57     57   3.12 JACKSON P E, 1985, V346, P125, J CHROMATOGR
     6      54     54   2.95 JACKSON P E, 1993, V640, P481, J CHROMATOGR
```

11

```
=> S HADDAD P R/AU; TRA L1 1- CIT
L1          211 HADDAD P R/AU
L2        TRANSFER L1 1- CIT :     211 TERMS
L3        1605 L2

=> S HADDAD P R/RAU; S L3 OR L4
L4        1004 HADDAD P R/RAU
L5        1829 L3 OR L4

=> S L5 NOT HADDAD ?/AU
L6        1663 L5 NOT HADDAD ?/AU
```

12

```
=> E HADDAD P/AU
E1            2     HADDAD O/AU
E2            1     HADDAD O M/AU
E3            4 --> HADDAD P/AU
E4            1     HADDAD P O/AU
E5           95     HADDAD P R/AU
E6            2     HADDAD P S/AU
E7            5     HADDAD PATRICK/AU
E8            7     HADDAD PAUL/AU
E9          143     HADDAD PAUL R/AU
E10           1     HADDAD PAUL RAYMOND/AU
E11           1     HADDAD PAULO DE TARSO/AU
E12           4     HADDAD PETER/AU

=> S E5 OR E8-E10
L4          246 "HADDAD P R"/AU OR ("HADDAD PAUL"/AU OR "HADDAD PAUL R"/AU OR …
```

Fig. 9. Inclusion of citations from the reference author field now produces an answer set of 1829 citations.

Fig. 10. A list of the most cited publications (L5, Fig. 9) can easily be obtained and displayed.

Fig. 11. L6 is an answer set after self-citations have been removed.

Fig. 12. Entries for authors in the CAS database are exactly as in the original article. First names, as well as initials, may thus be listed and it is important to EXPAND to identify all relevant terms.

←

field completely ignored all bibliographic terms other than the referenced author. That is, all the variations (just the first 10 of over 100 for the major reference work are shown in Fig. 8) will have separate entries and so these must be added to the first term in Fig. 10 in order to obtain all the citations to this work. This can be done manually, or through various EDIT or DISPLAY commands. Sometimes it is necessary to remove 'self-citations', in which authors cite their own publications, but this can easily be done. Accordingly, after removal of records which have HADDAD as the citing author from the answer set L5 (1829 records), a new answer set of 1663 records is obtained (Fig. 11).

3. Chemical abstracts database

Although citations have only recently been added to the CAS bibliographic database, it is of interest to compare the citation listings with those in the science citation index. Initially the same process outlined above for the science citation index is followed. An EXPAND on the author reveals one important difference, namely that the authors' first names are listed when they were included in the original article (Fig. 12) and this helps with the identification of the author. The database lists 246 publications for Paul Haddad and these have been cited 266 times. Selection and display of the cited references again is possible and the process and some of the cited publications are given in Fig. 13[2]. It now is of interest to compare records between the ISI and CAS databases and this is again performed by an algorithm (remember that the results must be checked carefully to ensure accuracy), which automatically detects and removes citations which are duplicated in both databases (Fig. 14). It is seen that 210 duplicates are removed and that a new answer set, L7, with 1661 answers is retrieved[3]. Thus, the original number of 1605 citations listed in Fig. 5 (which we know to have omitted the citations to the reference text) has grown by 56 citations that were listed in the CAS database since 1999, but not in the ISI database. The 56 new citations, or just over 20% of the total publications retrieved in the CAS database since 1999, are unique to this database. Again the accuracy of this should be checked and, for example, a display of the titles (Fig. 14) shows which records appear to be unique to the CAS database. Where duplicate answers have been detected the word DUPLICATE appears in the first line of the display and of the first five records, 1 and 2 appear to be unique to the CAS database. To be thorough, further analyses need to be conducted and when this is done it is noted that the answer 1 in Fig. 14 is a patent (which are not listed in the ISI database), while answer 2 actually is in both databases but the full indexing (including citations) has yet to be added.

We can also take records from the CAS database and search for citations in the ISI database. That is, once the records (L4, Fig. 12) in the CAS database have been identified they can be TRAnsferred to the ISI database. This retrieves about 30 more citations over those previously identified (L3, Fig. 5).

---

[2] Even though search terms have more than one initial, only the first initial is searched in the RE field.

[3] The duplicate removal was conducted on L3 and L6 (answer sets which did not contain hits from the RAU field). However it could have been done on answer sets which included RAU field hits. It is of interest to note that there were 327 citations to 'Ion Chromatography: Principles and Applications' in SCISEARCH and 77 in HCAPLUS.

```
13


=> TRA L4 1- CIT
L5         TRANSFER L4 1- CIT :     246 TERMS
L6         266 L5

=> D L5 1-10
L5              TRA L4 1- CIT :      246 TERMS

TERM # TERMS
------ ------------------------
     1 ALEXANDER P W, 1981, V209, P29,?/RE
     2 ALEXANDER P W, 1984, V17, P309,?/RE
     3 ALEXANDER P W, 1984, V56, P2417,?/RE
     4 ALEXANDER P W, 1985, V171, P151,?/RE
     5 ALEXANDER P W, 1985, V177, P183,?/RE
     6 ALEXANDER P W, 1985, V18, P1953,?/RE
     7 ALEXANDER P W, 1985, V20, P179,?/RE
     8 ALEXANDER P W, 1992, V589, P201,?/RE
     9 BAGCHI R, 1986, V351, P541,?/RE
    10 BILLIET H A H, 1996, V17, P1367,?/RE


14


=> DUP REM L3 L6

L7         1661 DUP REM L3 L6 (210 DUPLICATES REMOVED)

=> D 1-10 TI

L7   ANSWER 1 OF 1661  HCAPLUS  COPYRIGHT 2000 ACS
TI   Automated capillary electrophoresis method and apparatus

L7   ANSWER 2 OF 1661  HCAPLUS  COPYRIGHT 2000 ACS
TI   Indirect Fluorescence Detection of Amino Acids on Electrophoretic
     Microchips

L7   ANSWER 3 OF 1661  SCISEARCH  COPYRIGHT 2000 ISI (R)DUPLICATE 1
TI   Simultaneous separation of cationic and anionic proteins using
     zwitterionic surfactants in capillary electrophoresis

L7   ANSWER 4 OF 1661  SCISEARCH  COPYRIGHT 2000 ISI (R)DUPLICATE 2
TI   Potentiometric detection of carboxylic acids, phosphate esters, and
     nucleotides in liquid chromatography using anion-selective coated-wire
     electrodes

L7   ANSWER 5 OF 1661  SCISEARCH  COPYRIGHT 2000 ISI (R)DUPLICATE 3
TI   Separation and detection of explosives on a microchip using micellar
     electrokinetic chromatography and indirect laser-induced fluorescence
```

Fig. 13. The TRAnsfer command is used to identify search terms in the citation field, and then to search for citations. L6 contains 266 citations.

Fig. 14. The DUPlicate command is used to remove duplicate answers between different databases.

4. Web of Science

ISI provides a web interface, the Web of Science, to its science citation index. It offers much more than the science citation index, since it also includes other citation indexes and databases, includes more records, and takes full advantage of hyperlink possibilities open to web-based products. Web of Science is very easy to use – when Person Search is chosen under the Easy Search option it is a simple matter to enter a person's name and then click the Search button. Within a few seconds the next

screen appears. It lists the authors, titles and source information to answer the query 'Show me all of the articles in the database that cite this person's work'. Many people are simply happy to get the answer, but of course the key question is what has occurred between clicking the Search button and the appearance of the answers. This is critical because we want to know the limitations of the search and, in particular, whether there are false positive and false negative answers as well as the correct answers.

Web of Science found 1861 citations, about the same amount as was found in the SCISEARCH file through command line searching, taking into account the issue of first author in the reference field (for papers listed with author P R HADDAD) and the referenced author name (P R HADDAD) in the RAU field[4]. In short, it quickly gave what previously, after extensive command line searches, had been concluded to be a reasonable answer set. In particular it picked up citations to the reference work 'Ion Chromatography: Principles and Applications' since Haddad was the first author. If the same search for citations to works for JACKSON PE is performed, none of the citations to this work is retrieved. This highlights the importance of understanding how the search algorithm behind Web of Science operates, and being alert to issues which may have to be addressed. It may be better to find citations in Web of Science through Full Search and to work slowly through screens, for example by years. Alternatively, when individual articles are displayed through Web of Science, a hyperlink Times Cited appears and citations numbers can be seen. This process may have to be repeated many times in order to see all citations to journal articles by a single author. At this stage the types of data shown in Fig. 10 cannot be generated through Web of Science automatically, and instead these analyses must be obtained manually.

5. Conclusion

The major issues relating to searching for citations in electronic databases are the identification of first authors, and the accuracy and consistency with which the references are entered in the original articles. Further, it is usually necessary to search for citations both in the RE and RAU fields, but terms which give comprehensive results in the latter may be particularly hard to obtain. The full power of on-line commands cannot be employed in Web of Science, and whether reasonably comprehensive results are obtained through this interface depends very much on the author (or articles) being traced.

At the moment the number of citations in the full science citation index greatly exceeds the number of citations in the CAS bibliographic database, but searches in the latter may retrieve a number of unique citations and, as the database is extended, it will become an important source. The on-line commands SEARCH, TRANSFER and DUPLICATE REMOVE help to overcome some of the difficulties inherent in citation searching, but ultimately the most comprehensive results will be obtained only through extensive use of the EXPAND command. The Web of Science has a number of important citation linking features, and both Easy Search and Full Search for citations are conducted automatically through an intelligent algorithm. However caution should always be exercised and if comprehensive results are essential then very lengthy processes based on the names of first authors are required.

6. Glossary

This paper refers to several on-line commands that are needed to find information relating to citations on the STN network. Details of commands may be found at http://www.cas.org/training/stncommands/stncommands.html.

ANA	ANALYZE command is used to extract terms from an L-number and to provide statistical analysis of the terms
CAS	Chemical abstracts service http://www.cas.org/
CIT	CIT is used to extract bibliographic information from a record and re-format it for searches in the RE field
DIS	DISPLAY command is used to display information
DUP	DUPLICATE command is used to detect duplicate answers in multi-file searches
EDIT	EDIT command is used to change field codes in an E# list
EXPAND	EXPAND command is used to show an alphabetical list of entries
HCAPLUS	CAS bibliographic database on STN http://www.cas.org/ONLINE/DBSS/hcaplusss.html
ISI	Institute for Scientific Information http://www.isinet.com

[4] This answer set was obtained in the full Web of Science dating back to 1975. Institutions which have access to fewer years may get fewer citations.

L#	Line numbers on STN used to identify answer sets
RAU	Referenced author field on STN
RE	Field containing citation references on STN
SCISEARCH	Science citation index database on the STN network http://www.cas.org/ONLINE/DBSS/scisearchss.html
STN	International on-line network in the sciences http://www.cas.org/stn.html
TRA	TRANSFER command is used to extract terms and search them in one step

Acknowledgements

The author thanks Professor Paul Haddad who prompted this article, through discussions on citation issues and on the need for greater awareness of the issues among end-users. Permission to reproduce data by the Institute for Scientific Information and the American Chemical Society is gratefully acknowledged.

References

[1] K. Reavie, Methods Mol. Biol. 132 (2000) 471.
[2] A.L. Smith, J. Chem. Educ. 76 (1999) 1153.
[3] S. Komatsu, Joho Kanri 40 (1997) 299.
[4] R.L.M. Synge, J. Chem. Inf. Comput. Sci. 30 (1990) 33.
[5] M.L. Pao, Comput. Biomed. Res. 26 (1993) 143.
[6] E. Zass, Chimia 53 (1999) 253.
[7] S. Redner, Eur. Phys. J. B4 (1998) 131.
[8] J. Reedijk, New J. Chem. 22 (1998) 767.
[9] O. Exner, Chem. Listy 87 (1993) 719.
[10] J.S. Kotiaho, Nature 398 (1999) 19.
[11] P.O. Seglen, Allergy 52 (1997) 1050.
[12] D. Schoonbaert, G. Roelants, Trop. Med. Int. Health 1 (1996) 739.
[13] J. Reedijk, Chem. Mag. 11 (1996) 443.