ELSEVIER

Journal of Power Sources 110 (2002) 163–176

JOURNAL OF POWER SOURCES

www.elsevier.com/locate/jpowsour

# Electrochemical power text mining using bibliometrics and database tomography

Ronald N. Kostoff[a,*], Rene Tshiteya[b], Kirstin M. Pfeil[a], James A. Humenik[c]

[a]Office of Naval Research, 800 North Quincy Street, 22217 Arlington, VA 22217, USA
[b]DDL Omni Engineering, LLC 8260 Greensboro Drive, Suite 600, Mclean, VA 22102, USA
[c]Noesis Inc., Manassas, VA, USA

Received 12 April 2002; accepted 17 April 2002

**Abstract**

Database tomography (DT) is a textual database analysis system consisting of two major components: (1) algorithms for extracting multi-word phrase frequencies and phrase proximities (physical closeness of the multi-word technical phrases) from any type of large textual database, to augment (2) interpretative capabilities of the expert human analyst. DT was used to derive technical intelligence from an electrochemical power database derived from the science citation index (SCI). Phrase frequency analysis by the technical domain experts provided the pervasive technical themes of the electrochemical power database, and the phrase proximity analysis provided the relationships among the pervasive technical themes. Bibliometric analysis of the electrochemical power literature supplemented the DT results with author/journal/institution publication and citation data. 

*Keywords:* Electrochemical power; Database tomography; Bibliometric analysis; Text mining; Information retrieval; Technical intelligence

## 1. Introduction

Science and technology are assuming an increasingly important role in the conduct and structure of domestic and foreign business and government. In the highly competitive civilian and military worlds, there has been a commensurate increase in the need for scientific and technical intelligence to insure that one's perceived adversaries do not gain an overwhelming advantage in the use of science and technology. While there is no substitute for direct human intelligence gathering, there have become available many techniques that can support and complement it. In particular, techniques that identify, select, gather, cull, and interpret large amounts of technological information semi-automatically can expand greatly the capabilities of human beings in performing technical intelligence.

One such technique is DT [2–4], a system for analyzing large amounts of textual computerized material. It includes algorithms for extracting multi-word phrase frequencies and phrase proximities from the textual databases, coupled with the topical expert human analyst to interpret the results and convert large volumes of disorganized data to ordered information. Phrase frequency analysis (occurrence frequency of multi-word technical phrases) provides the pervasive technical themes of a database, and the phrase proximity (physical closeness of the multi-word technical phrases) analysis provides the relationships among pervasive technical themes, as well as among technical themes and authors/journals/institutions/countries, etc. The present paper describes use of the DT process, supplemented by literature bibliometric analyses, to derive technical intelligence from the published literature of electrochemical power science and technology.

Electrochemical power, as defined by the authors for this study, is the generation and conversion of power, and the storage of energy, using electrochemical processes. Since one of the key outputs of the present study is a query that can be used by the community to access relevant electrochemical power documents, a recommended query based on this study is presented in total. This query serves as the operational definition of electrochemical power, and its development is discussed in the database generation section.

### 1.1. Electrochemical power query

(Fuel cell* or sofc* or pemfc* or dmfc* or ultracapacitor* or supercapacitor* or pseudocapacitor* or (capacitor* same (electrochemical or electrolyte* or double-layer)) or ((battery or batteries) same (lithium or Li or electrode* or anode*

* Corresponding author. Tel.: +1-703-696-4198; fax: +1-703-696-4274.
*E-mail address:* kostofr@onr.navy.mil (R.N. Kostoff).

0378-7753/02/$ – see front matter 
PII: S0378-7753(02)00233-1

or cathode* or capacity or material* or electrochemical or charge or charging or discharge* or discharging or rechargeable or electrolyte* or lithium or Li or lithium-ion or nickel or metal hydride* or lead-acid or alloy*)) or ((lithium or Li) same (electrochemical or discharge* or discharging or electrode* or liclo4 or rechargeable or cycling or reversible or insertion or mah or intercalation)) or (electrochemical same (discharge* or discharging or hydrogen storage or mah)) or (hydrogen storage same (alloy* or electrode*)) or (limn2o4 same electrode*) or (lipf6 same electrolyte*) or (charge-discharge same electrode*) or ((discharge capacity or metal hydride*) same electrode*) or (electrolyte* same lsgm) or (hydrogen same storage alloy*) or (nafion same polymer*) or (ptru same co) or (ruo2 same electrode*)) NOT( ((electrode* or hydrogen or discharge*) same plasma*) or (discharge* same gas) or DNA or assay* or biosensor* or rats or blood or capillary or protein* or mercury or clinical or amino or hydrogen peroxide or paste or corona or tissue* or helium or ascorbic acid or receptor* or chromium or radiation or bacteria* or plant* or extracellular or antenna* or magnetron or drug* or vivo or hydrolysis or ml or amperometric or care or cd or buffer or silicon or stress or sensor* or rf or filter* or switching or detection limit* or inhibition* or ar or ms or electrostatic or phi or monolayer* or gate* or sheath* or gc or depletion or combustion or serum* or toxicity or converter* or chromatography or radical* or oil* or generator* or target* or gap* or excitation* or environmental or glow* or ring or rings or diet* or pretreatment* or space charge* or amine* or ultrasound or lamp* or scan rate* or health* or solar or fe2 or reflection* or electromagnetic or carboxylic or deep or diode* or synthetic* or acetic acid or collision* or moiety or dimeric or titanate* or carbon steel* or curvature* or lithium chloride or coercive field or network* or hydrodynamic* or tris or mutant* or backbone* or decay* or monomer* or outcome* or driving or contamination or spatial or cmos or mediator* or excited or led or self-assembled or nitric oxide or *I–V* or array* or mmol or dt or waste* or aromatic or epitaxial or atomic force microscopy or differential pulse or viscosity or sorption or pk or native or shifts or recording* or adhesion* or dye* or surfactants).

To execute the study reported in this paper, a database of relevant electrochemical power articles is generated using the iterative search approach of simulated nucleation [5,6]. Then, the database is analyzed to produce the following characteristics and key features of the electrochemical power field: recent prolific electrochemical power authors; journals that contain numerous electrochemical power papers; institutions that produce numerous electrochemical power papers; keywords most frequently specified by the electrochemical power authors; authors, papers and journals cited most frequently; pervasive technical themes of electrochemical power; and relationships among the pervasive themes and sub-themes.

What is the importance of applying DT and bibliometrics to a topical field such as electrochemical power? The roadmap, or guide, of this field produced by DT and bibliometrics provides the demographics and a macroscopic view of the total field in the global context of allied fields. This allows specific starting points to be chosen rationally for more detailed investigations into a specific topic of interest. DT and bibliometrics do not obviate the need for detailed investigation of the literature or interactions with the main performers of a given topical area in order to make a substantial contribution to the understanding or the advancement of this topical area, but allow these detailed efforts to be executed more efficiently. DT and bibliometrics are quantity-based measures (number of papers published, frequency of technical phrases, etc.), and correlations with intrinsic quality are less direct. The direct quality components of detailed literature investigation and interaction with performers, combined with the DT and bibliometrics analysis, can result in a product highly relevant to the user community.

## 2. Background

### 2.1. Overview

The information sciences background for the approach used in this paper is presented in [7]. This reference shows the unique features of the computer and co-word-based DT process relative to other roadmap techniques. It describes the two main roadmap categories (expert-based and computer-based), summarizes the different approaches to computer-based roadmaps (citation and co-occurrence techniques), presents the key features of classical co-word analysis, and shows the evolution of DT from its co-word roots to its present form.

The DT method in its entirety requires generically three distinct steps. The first step is identification of the main themes of the text being analyzed. The second step is determination of the quantitative and qualitative relationships among the main themes and their secondary themes. The final step is tracking the evolution of these themes and their relationships through time. The first two steps are summarized in 2.1.1 and 2.1.2. Time evolution of themes has not yet been studied.

At this point, a variety of different analyses can be performed. For databases of non-journal technical articles [2], the final results have been identification of the pervasive technical themes of the database, the relationship among these themes, and the relationship of supporting sub-thrust areas (both high and low frequency) to the high-frequency themes. For the more recent studies in which the databases are journal article abstracts and associated bibliometric information (authors, journals, addresses, etc.), the final results have also included relationships among the technical themes and authors, journals, institutions, etc. [7–11].

These more recent DT/bibliometrics studies were conducted of the technical fields of: (1) near-earth space (NES) [8]; (2) hypersonic and supersonic flow over aerodynamic bodies (HSF) [7]; (3) chemistry (JACS) [9] as represented by

Table 1
DT studies of topical fields

| Topical area | Number of SCI articles | Years covered |
|---|---|---|
| (1) Near-earth space (NES) | 5480 | 1993 to mid 1996 |
| (2) Hypersonics (HSF) | 1284 | 1993 to mid 1996 |
| (3) Chemistry (JACS) | 2150 | 1994 |
| (4) Fullerenes (FUL) | 10515 | 1991 to mid 1998 |
| (5) Aircraft (AIR) | 4346 | 1991 to mid 1998 |
| (6) Hydrodynamics (HYD) | 4608 | 1991 to mid 1998 |
| (7) Electric power sources (EPS) | 20835 | 1991 to beginning of 2000 |
| (8) Research assessment (RIA) | 2300 | 1991 to beginning of 1995 |
| (9) Electrochemical power sources (ECHEM) | 6985 | 1991 to mid 2001 |

the Journal of the American Chemical Society; (4) fullerenes (FUL) [10]; (5) aircraft (AIR) [11]; (6) hydrodynamic flow over surfaces (HYD); (7) electric power sources (EPS); (8) the non-technical field of research impact assessment (RIA). Overall parameters of these studies from the SCI database results and the current electrochemical study are shown in Table 1.

### 2.1.1. First step

The frequencies of appearance in the total text of all single word phrases (e.g. matrix), adjacent double word phrases (e.g. metal matrix), and adjacent triple word phrases (e.g. metal matrix composites) are computed. The highest frequency significant technical content phrases are selected by topical experts as the pervasive themes of the full database.

### 2.1.2. Second step

*2.1.2.1. Numerical boundaries*. For each theme phrase, the frequencies of phrases within $+/-M$ (nominally 50) words of the theme phrase are computed for every occurrence of the theme phrase in the full text, and a phrase frequency dictionary is constructed. This dictionary contains the phrases closely related to the theme phrase. Numerical indices are employed to quantify the strength of this relationship. Both quantitative and qualitative analyses are performed by the topical expert for each dictionary (hereafter called cluster) yielding, among many results, those sub-themes closely related to and supportive of the main cluster theme.

Threshold values are assigned to the numerical indices, and these indices are used to filter out the phrases most closely related to the cluster theme. However, because numbers are limited in their ability to portray the conceptual relationships among themes and sub-themes, the qualitative analyses of the extracted data by the topical experts have been at least as important as the quantitative analyses. The richness and detail of the extracted data in the full text analysis allow an understanding of the theme inter-relationships not heretofore possible with previous text abstraction techniques (using index words, key words, etc.).

*2.1.2.2. Semantic boundaries*. The approach is conceptually similar to 2.1.2.1, with the difference being that semantic boundaries are used to define the co-occurrence domain rather than numerical boundaries. The only semantic boundaries used for the present studies were paper abstract boundaries. Software is being developed that will allow paragraphs or sentences to be used as semantic boundaries.

It is an open question as to whether semantic boundaries or numerical boundaries provide more accurate results. The elemental messages of text are contained in concepts or thoughts. Sentences or paragraphs are the vehicles by which the concepts or thoughts are expressed. The goal of text mining is to usually quantify relationships occurring in the concepts or thoughts, not in the fragments of their vehicles of expression. In particular, while intra-sentence relationships will be very strong, they may be overly restrictive for text mining purposes, and many cross-discipline relationships can be lost by adhering to intra-sentence relationships only. Intra-paragraph relationships are more inclusive and reasonable. For journal paper abstracts of the type found in SCI, many abstracts constitute a single paragraph.

### 2.2. Unique study features

The study reported in the present paper is in the journal article abstract category. It differs from the previous published papers in this category [7–11] in three respects. First, the topical domain (electrochemical power) is completely different. Second, a much more rigorous statistically-based technical theme clustering approach is used. Third, bibliometric clustering is presented for two database fields: authors and countries.

## 3. Database generation

The key step in the electrochemical power literature analysis is the generation of the database to be used for processing. There are three key elements to database generation: the overall objectives, the approach selected, and the database used. Each of these elements is described.

### 3.1. Overall study objectives

The main objective was to identify global S&T that had both direct and indirect relations to electrochemical power. A sub-objective was to estimate the overall level of global effort in electrochemical power S&T, as reflected by the emphases in the published literature.

### 3.2. Databases and approach

For the present study, the SCI database was used. The approach used for query development was the DT-based iterative relevance feedback concept [5].

*3.2.1. Science citation index [12]*

The database consists of selected journal records (including authors, titles, journals, author addresses, author keywords, abstract narratives, and references cited for each paper) obtained by searching the web version of the SCI for electrochemical power articles. At the time the data was extracted for the present paper (mid 2001), the version of the SCI used accessed about 5600 journals (mainly in physical, engineering, and life sciences basic research).

The SCI database selected represents a fraction of the available electrochemical power (mainly research) literature, that in turn represents a fraction of the electrochemical power S&T actually performed globally [13]. It does not include the large body of classified literature, or company proprietary technology literature. It does not include technical reports or books or patents on electrochemical power. It covers a finite slice of time (1991 to mid 2001). The database used represents the bulk of the peer-reviewed high quality electrochemical power research, and is a representative sample of all electrochemical power research in recent times.

To extract the relevant articles from the SCI, the title, keyword, and abstract fields were searched using keywords relevant to electrochemical power, although different procedures were used to search the title and abstract fields [5]. The resultant abstracts were culled to those relevant to electrochemical power. The search was performed with the aid of two powerful DT tools (multi-word phrase frequency analysis and phrase proximity analysis) using the process of simulated nucleation [5].

An initial query of electrochemical power-related terms produced two groups of papers: one group was judged by domain experts to be relevant to the subject matter, the other was judged to be non-relevant. Gradations of relevancy or non-relevancy were not considered. An initial database of titles, keywords, and abstracts was created for each of the two groups of papers. Phrase frequency and proximity analyses were performed on this textual database for each group. The high frequency single, double, and triple word phrases characteristic of the relevant group, and their boolean combinations, were then added to the query to expand the papers retrieved. Similar phrases characteristic of the non-relevant group were effectively subtracted from the query to contract the papers retrieved. The process was repeated on the new database of titles, keywords, and abstracts obtained from the search. A few more iterations were performed until the number records retrieved stabilized (convergence). The final phrase-based query used for the electrochemical power study was shown in the Introduction.

The authors believe that queries of these magnitudes and complexities are required when necessary to provide a tailored database of relevant records that encompasses the broader aspects of target disciplines. In particular, if it is desired to enhance the transfer of ideas across disparate disciplines, and thereby stimulate the potential for innovation and discovery from complementary literatures [1], then even more complex queries using simulated nucleation may be required.

## 4. Results

The results from the publications bibliometric analyses are presented in Section 4.1, followed by the results from the citations bibliometrics analysis in Section 4.2. Results from the DT analyses are shown in Section 4.3. The SCI bibliometric fields incorporated into the database included, for each paper, the author, journal, institution, and keywords. In addition, the SCI included references for each paper.

### 4.1. Publication statistics on authors, journals, organizations, countries

The first group of metrics presented is counts of papers published by different entities. These metrics can be viewed as output and productivity measures. They are not direct measures of research quality, although there is some threshold quality level inferred, since these papers are published in the (typically) high caliber journals accessed by the SCI.

*4.1.1. Author frequency results*

There were 6985 papers retrieved, 11,051 different authors, and 25,465 author listings. The occurrence of each author's name on a paper is defined as an author listing. While the average number of listings per author is about 2.3, the 20 most prolific authors (see Table 2) have listings more than an order of magnitude greater than the average. The number of papers listed for each author are those in the database of records extracted from the SCI using the query, not the total number of author papers listed in the source SCI database.

Of the 20 most prolific authors listed in Table 2, seven are from Japan. In fact, 13 are from the far East, four are from Europe (western), two are from North America, and one is from the Middle East. Eighteen are from universities, and two are from research institutes. Total publications listed in the SCI for each of these 20 authors were scanned visually, and, on average, these authors were rarely listed as first authors. For example, in their 100 most recent papers, J.R. Dahn was listed as first author five times, and J.M. Tarascon was listed as first author six times.

*4.1.2. Journals containing most electrochemical power papers*

There were 587 different journals represented, with an average of 11.90 papers per journal. The journals containing the most power-related electrochemistry papers (see Table 3) had more than an order of magnitude more papers than the average.

The majority of the journals are electrochemistry, with the remainder divided between chemistry and materials. There appear to be three primary groups at the top layer.

Table 2
Most Prolific Authors (present institution listed)

| Author name | Institution | Country | Number of papers |
|---|---|---|---|
| J.R. Dahn | Dalhousie University | Canada | 67 |
| J.M. Tarascon | University of Picardie | France | 53 |
| Q.D. Wang | Zhejiang University | China | 51 |
| Y.Q. LEI | Zhejiang University | China | 46 |
| H.K. Liu | University of Wollongong | Australia | 44 |
| S.X. Dou | University of Wollongong | Australia | 44 |
| B. Scrosati | University of Roma LA Sapienza | Italy | 43 |
| J.Y. Lee | National University of Singapore | Singapore | 42 |
| N. Kumagai | Iwate University | Japan | 41 |
| O. Yamamoto | Aichi Institute of Technology | Japan | 40 |
| M. Yoshio | Saga University | Japan | 40 |
| D. Aurbach | Bar Ilan University | Israel | 38 |
| I. Uchida | Tohoku University | Japan | 37 |
| M. Watanabe | University of Yamanashi | Japan | 37 |
| L.Q. Chen | Chinese Academy Science | China | 36 |
| Y. Takeda | MIE University | Japan | 36 |
| S. Passerini | Enea | Italy | 35 |
| J.L. Tirado | University of Cordoba | Spain | 33 |
| C. Iwakura | University fo Osaka Prefecture | Japan | 32 |
| R.E. White | University of South Carolina | USA | 32 |

The Journal of Power Sources, an international journal devoted to the science and technology of electrochemical energy systems, contains the most articles by far. This is not surprising, since its stated mission is fully aligned with the main objective of the present study. While many of its articles were retrieved by the query, essentially all of its articles are relevant to the topic of the present study.

The next group consists of the Journal of the Electrochemical Society (JES) and Solid State Ionics (SSI). The JES focuses on solid-state and electrochemical science and technology, while SSI is devoted to the physics, chemistry and materials science of diffusion, mass transport, and reactivity of solids. While these journals include aspects of electrochemistry/electrochemical power sources in their charters, they include other aspects of chemistry (and physics) as well. The next five journals listed constitute the third group.

Table 3
Journals containing most papers

| Journal names | Number of papers |
|---|---|
| J. Power Sources | 1240 |
| J. Electrochem. Soc. | 771 |
| Solid State Ionics | 546 |
| Electrochim. Acta | 403 |
| J. Alloy Compd. | 290 |
| Denki Kagaku | 198 |
| J. Appl. Electrochem. | 167 |
| J. Electroanal. Chem. | 138 |
| Electrochem. Solid State Lett. | 119 |
| Int. J. Hydrog. Energ. | 112 |
| Russ. J. Electrochem. | 100 |
| Electrochemistry | 86 |
| J. Mater. Chem. | 81 |
| J. Solid State Chem. | 72 |
| Chem. Mat. | 70 |
| J. New Mat. Electrochem. Syst. | 60 |
| Electrochem. Commun. | 56 |
| Synth. Met. | 55 |
| Bull. Electrochem. | 54 |
| J. Phys. Chem. B | 50 |

#### 4.1.3. Institutions producing most electrochemical power papers

A similar process was used to develop a frequency count of institutional address appearances. It should be noted that many different organizational components may be included under the single organizational heading (e.g., Harvard University could include the Chemistry Department, Biology Department, Physics Department, etc.). Identifying the higher level institutions is instrumental for these DT studies. Once they have been identified through bibliometric analysis, subsequent measures may be taken (if desired) to identify particular departments within an institution.

The most prolific institutions are listed in Table 4. Of the 20 most prolific institutions, 10 are from Asia, five are from western Europe, four from the USA, and one from eastern Europe. Twelve are universities, and the remaining institutions are research institutes.

#### 4.1.4. Countries producing most electrochemical power papers

There are 78 different countries listed in the results. The country bibliometric results are summarized in Table 5. The dominance of a handful of countries is clearly evident.

Table 4
Prolific institutions

| Institution names | Country | Number of papers |
|---|---|---|
| Chinese Academy of Sciences | China | 118 |
| Kyoto University | Japan | 108 |
| CNRS | France | 104 |
| Korea Advanced Institute of Science & Technology | Korea | 90 |
| Russian Academy of Sciences | Russia | 89 |
| Zhejiang University | China | 85 |
| Argonne National Laboratory | USA | 79 |
| University of Calif Berkeley | USA | 78 |
| Tohoku University | Japan | 73 |
| MIT | USA | 66 |
| CNR | Italy | 63 |
| Central Electrochemical Research Institute | India | 60 |
| Seoul National University | Korea | 60 |
| Tokyo Institute of Technology | Japan | 55 |
| CSIC | Spain | 55 |
| KFA Julich GMBH | Germany | 54 |
| University of South Carolina | USA | 54 |
| Osaka National Research Institute | Japan | 52 |
| University of Tokyo | Japan | 51 |
| Delft University of Technology | The Netherlands | 51 |

Table 5
Prolific countries

| Country name | Number of papers |
|---|---|
| Japan | 1552 |
| USA | 1318 |
| France | 558 |
| Peoples Republic of China | 499 |
| South Korea | 380 |
| Germany | 341 |
| Canada | 318 |
| England | 285 |
| Italy | 250 |
| India | 249 |
| Russia | 206 |
| Spain | 151 |
| Sweden | 126 |
| Australia | 121 |
| Switzerland | 113 |
| The Netherlands | 97 |
| Taiwan | 90 |
| Brazil | 83 |
| Israel | 78 |
| Poland | 73 |

There appear to be three dominant groups in the 20 most prolific countries. The US and Japan constitute the most dominant group, and were the only two countries to have published more than 1000 papers on power-related electrochemistry during the past 8 years. France and China constitute the next group, but had less papers combined than either member of the first group. The next seven countries constitute the third group.

Interestingly, unlike all previous DT studies, the United States (US) was not the most prolific country. Japan had more published papers (nearly 18% more) than the US. Overall, eastern Asia (Japan, China, South Korea, Taiwan), northern North America (US, Canada), and western Europe (France, Germany, UK) accounted for most of the electrochemistry research activity.

Fig. 1 contains a co-occurrence matrix of the top 15 countries. In terms of absolute numbers of co-authored papers, the USA major partners are Japan, France, Italy, Canada, and South Korea. Overall, countries in similar geographical regions tend to co-publish substantially, the US being a moderate exception.

### 4.2. Citation statistics on authors, papers, and journals

The second group of metrics presented is counts of citations to papers published by different entities. While citations are ordinarily used as impact or quality metrics [14], much caution needs to be exercised in their frequency count interpretation, since there are numerous reasons why authors cite or do not cite particular papers [15,16].

The citations in all the retrieved SCI papers were aggregated, the authors, specific papers, years, journals, and countries cited most frequently were identified, and were presented in order of decreasing frequency. A small percentage of any of these categories received large numbers of

| | # Records | 1552 | 1318 | 558 | 499 | 380 | 344 | 341 | 318 | 250 | 249 | 206 | 151 | 126 | 121 | 113 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Records | COUNTRY | JAPAN | USA | FRANCE | CHINA | SOUTH KOREA | UK | GERMANY | CANADA | ITALY | INDIA | RUSSIA | SPAIN | SWEDEN | AUSTRALIA | SWITZERLAND |
| 1552 | JAPAN | 1552 | 52 | 15 | 17 | 16 | 14 | 8 | 3 | 5 | 4 | 0 | 0 | 3 | 3 | 5 |
| 1318 | USA | 52 | 1318 | 36 | 6 | 17 | 10 | 9 | 24 | 27 | 9 | 6 | 5 | 5 | 2 | 4 |
| 558 | FRANCE | 15 | 36 | 558 | 3 | 4 | 10 | 13 | 9 | 9 | 5 | 4 | 31 | 3 | 0 | 4 |
| 499 | CHINA | 17 | 6 | 3 | 499 | 1 | 2 | 5 | 5 | 1 | 0 | 0 | 1 | 8 | 0 | 0 |
| 380 | SOUTH KOREA | 16 | 17 | 4 | 1 | 380 | 0 | 4 | 1 | 5 | 2 | 1 | 0 | 0 | 2 | 1 |
| 344 | UK | 14 | 10 | 10 | 2 | 0 | 344 | 9 | 7 | 2 | 5 | 2 | 0 | 4 | 2 | 3 |
| 341 | GERMANY | 8 | 9 | 13 | 5 | 4 | 9 | 341 | 3 | 7 | 5 | 6 | 0 | 2 | 0 | 13 |
| 318 | CANADA | 3 | 24 | 9 | 5 | 1 | 7 | 3 | 318 | 0 | 0 | 1 | 0 | 2 | 1 | 1 |
| 250 | ITALY | 5 | 27 | 9 | 1 | 5 | 2 | 7 | 0 | 250 | 5 | 2 | 1 | 2 | 0 | 2 |
| 249 | INDIA | 4 | 9 | 5 | 0 | 2 | 5 | 5 | 0 | 5 | 249 | 0 | 1 | 0 | 4 | 2 |
| 206 | RUSSIA | 0 | 6 | 4 | 0 | 1 | 2 | 6 | 1 | 2 | 0 | 206 | 2 | 2 | 0 | 1 |
| 151 | SPAIN | 0 | 5 | 31 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 151 | 0 | 0 | 0 |
| 126 | SWEDEN | 3 | 5 | 3 | 8 | 0 | 4 | 2 | 2 | 2 | 0 | 2 | 0 | 126 | 0 | 1 |
| 121 | AUSTRALIA | 3 | 2 | 0 | 0 | 2 | 2 | 0 | 1 | 0 | 4 | 0 | 0 | 0 | 121 | 0 |
| 113 | SWITZERLAND | 5 | 4 | 4 | 0 | 1 | 3 | 13 | 1 | 2 | 2 | 1 | 0 | 1 | 0 | 113 |

Fig. 1. Country co-occurrence matrix.

Table 6
Most cited authors (cited by other papers in this database only)

| Author names | Institutions | Countries | Times cited |
|---|---|---|---|
| T. Ohzuku | Osaka City University | Japan | 1066 |
| M.M. Thackeray | Argonne National Laboratory | USA | 845 |
| D. Aurbach | Bar Ilan University | Israel | 808 |
| J.M. Tarascon | University of Picardie | France | 755 |
| J.R. Dahn | Dalhousie University | Canada | 698 |
| M. Watanabe | University of Yamanashi | Japan | 601 |
| K.M. Abraham | Covalent Associates | USA | 461 |
| R.J. Gummow | CSIR | South Africa | 455 |
| C. Delmas | CNRS | France | 429 |
| T. Sakai | Osaka National Research Institute | Japan | 412 |
| G. Pistoia | CNR | Italy | 391 |
| N.Q. Minh | Allied Signal Aero | USA | 381 |
| J.B. Goodenough | University of Texas | USA | 379 |
| T. Ishihara | Oita University | Japan | 370 |
| B.C.H. Steele | University of London Imperial | England | 351 |
| J.N. Reimers | Moli Energy | Canada | 345 |
| E. Peled | Tel Aviv University | Israel | 335 |
| D. Guyomard | University of Nantes | France | 332 |
| J. Mizusaki | Tohoku University | Japan | 324 |
| A.J. Appleby | Texas A&M | USA | 300 |

citations. From the citation year results, the most recent papers tended to be the most highly cited. This reflected rapidly evolving fields of research.

#### 4.2.1. Most cited authors

The most highly cited authors are listed in Table 6.

Of the 20 most cited authors, five are from Japan, five from the USA, five from Europe (western), two from Canada, two from Israel, and one from Africa. This is a far different distribution from the most prolific authors, where 13 were from the far East. There are a number of potential reasons for this difference, including difference in quality and late entry into the research discipline. In another 3 or 4 years, when the papers from present-day authors have accumulated sufficient citations, firmer conclusions about quality can be drawn.

The lists of 20 most prolific authors and 20 most highly cited authors only had four names in common (Aurbach, Tarascon, Dahn, Watanabe). This phenomenon of minimal intersection between most prolific and most cited authors has been observed in all other text mining studies performed by the first author.

Thirteen of the authors' institutions are universities, four are government-sponsored research laboratories, and three are private companies. The appearance of the companies on this list is another differentiator from the list of most prolific authors.

The citation data for authors and journals represents citations generated only by the specific records extracted from the SCI database for this study. It does not represent all the citations received by the references in those records; these references in the database records could have been cited additionally by papers in other technical disciplines.

#### 4.2.2. Most cited papers

The most highly cited papers are listed in Table 7.

The theme of each paper is shown in italics on the line after the paper listing. The order of paper listings is by number of citations by other papers in the extracted database analyzed. The total number of citations from the SCI paper listing, a more accurate measure of total impact, is shown in the last column on the right.

Table 7
Most cited papers (total citations listed in SCI)

| Author name | Year | Journal | Volume | SCI cites |
|---|---|---|---|---|
| J.M. Tarascon | 1991 | J. Electrochem. Soc. ($LiMn_2O_4$ spinel phase as secondary lithium cell cathode) | 138 | 272 |
| N.Q. Minh | 1993 | J. Am. Ceram. Soc. (Ceramic fuel cells—review) | 76 | 476 |
| T. Ohzuku | 1993 | J. Electrochem. Soc. (Synthesis of $LiNiO_2$ for secondary lithium cell) | 140 | 217 |
| R.J. Gummow | 1994 | Solid State Ionics (Improved rechargeable capacity of $LiMn_2O_4$ cathodes) | 69 | 281 |
| T. Ohzuku | 1990 | J. Electrochem. Soc. (Electrochemistry of $MnO_2$ in lithium cells) | 137 | 314 |
| K. Mizushima | 1980 | Mater. Res. Bull. ($LiXCOO_2$ for high-energy density battery cathodes) | 15 | 392 |
| D. Guyomard | 1992 | J. Electrochem. Soc. (Li metal-free rechargeable $LiMn_2O_4$/carbon cells) | 139 | 300 |
| M.M. Thackeray | 1983 | Mater. Res. Bull. (Lithium insertion into manganese spinels) | 18 | 358 |
| J.M. Tarascon | 1994 | J. Electrochem. Soc. (Lithium insertion into the spinel $LiMn_2O_4$) | 141 | 247 |
| R. Fong | 1990 | J. Electrochem. Soc. (Lithium intercalation into carbon using non-aqueous cells) | 137 | 334 |
| J.N. Reimers | 1992 | J. Electrochem. Soc. (Lithium intercalation in lixcoo2) | 139 | 227 |
| I.A. Courtney | 1997 | J. Electrochem. Soc. (Lithium reaction with tin oxide composites in lithium ion cell) | 144 | 147 |
| K. Sato | 1994 | Science (Lithium storage in disordered carbons) | 254 | 221 |
| M.M. Thackeray | 1992 | J. Electrochem. Soc. (Spinel electrodes from limno system for secondary batteries) | 139 | 202 |
| M.M. Thackeray | 1984 | Mater. Res. Bull. (Electrochemical extraction of lithium from $LiMn_2O_4$) | 19 | 235 |
| T. Ishihara | 1994 | J. Am. Chem. Soc. (Doped lago3 oerovskite oxide ionic conductor) | 116 | 201 |
| R.D. Shannon | 1976 | Acta Crystallogr. A (Ionic-radii and interatomic distances in halides and chalcogenides) | 32 | 10254 |
| J.J.G. Willems | 1984 | Phillips J. Res. (Metal hydride electrodes for rechargeable battery) | 39 | 285 |
| K.M. Abraham | 1990 | J. Electrochem. Soc. ($Li^+$-conductive solid polymer electrolytes with liquid-like conduct) | 137 | 202 |
| T. Ohzuku | 1993 | Electrochim. Acta (Li–N–Co oxides for secondary lithium cells | 38 | 139 |

The JES contains the most papers, 12 out of the 20 listed. Most of the journals are fundamental science journals, and most of the topics have a fundamental science theme. Most of the papers are from the 1990s, with four being from the 1980s, and one extremely highly cited paper being from 1976. This reflects a dynamic research field, with seminal works being performed in the recent past.

Sixteen of the papers address issues related to lithium secondary batteries, with the dominant issue theme being lithium insertion/intercalation to avoid free-metal formation. Two of the papers address issues related to ceramic fuel cells, with the dominant issue theme being solid oxides for high ionic conductivity. One paper addresses issues related to nickel metal hydride rechargeable batteries.

Thus, the major intellectual emphasis of cutting edge electrochemical power sources research, as evidenced by the most cited papers, is well aligned with the intellectual heritage and performance emphasis, as will be evidenced by the clustering approaches.

#### *4.2.3. Most cited journals*

The most highly cited journals are listed in Table 8. The JES received as many citations as the next three journals combined. Most of the highly cited journals are electrochemistry, some are materials, some chemistry, with one physics journal represented. Based on all the citation results, there is little evidence that disciplines outside the tightly knit electrochemistry-materials groups relevant to the specific applications are being accessed.

The authors end this bibliometrics section by recommending that the reader interested in researching the topical field of interest would be well-advised to, first, obtain the highly-cited papers listed and, second, peruse those sources that are highly cited and/or contain large numbers of recently published papers.

Table 8
Most cited journals (cited by other papers in this database only)

| Journal names | Times cited |
|---|---|
| J. Electrochem. Soc. | 22363 |
| Solid State Ionics | 9782 |
| J. Power Sources | 8265 |
| Electrochim. Acta | 5994 |
| J. Electroanal. Chem. | 4607 |
| J. Solid State Chem. | 2364 |
| J. Alloy Compd. | 2269 |
| J. Appl. Electrochem. | 2008 |
| Mater. Res. Bull. | 1811 |
| Phys. Rev. B | 1672 |
| J. Am. Chem. Soc. | 1491 |
| J. Phys. Chem. (US) | 1470 |
| J. Am. Ceram. Soc. | 1417 |
| J. Less-Common Met. | 1399 |
| Denki Kagaku | 1157 |
| Synth. Met. | 1041 |
| Chem. Mater. | 969 |
| Electrochem. Soc. | 851 |
| Science | 841 |

### *4.3. Database tomography results*

There are two major analytic methods used in this section to generate taxonomies of the SCI databases: non-statistical clustering, based on ‘phrase frequency analysis’, and statistical clustering, based on ‘phrase proximity analysis’.

#### *4.3.1. Non-statistical clustering*

*4.3.1.1. Keyword taxonomy.* All the keywords from the extracted SCI records, and their associated frequencies of occurrence, were tabulated, and then grouped into categories by visual inspection. The phrases were of two types: ‘system-related’ and ‘tech base-related’. While the system sub-categories were relatively independent, there was substantial overlap between some of the tech base categories. In particular, the generic system components (electrolytes, electrodes) had overlap with the generic tech base components. These results are summarized now.

The main system categories were sources-converters, storage, and other applications. Sources-converters included fuel cells, storage included batteries and capacitors (to a much smaller extent), and other applications covered many uses of electrochemical cells.

The main tech-base categories were electrolytes, electrodes, materials, processes, properties, phenomena, parameters, experiments, theory, macrostructure, microstructure, and region.

Fuel cell concepts emphasized were solid oxide, molten carbonate, and polymer electrolyte. While most efforts addressed pure hydrogen input, or reformed hydrocarbon input to supply hydrogen (e.g. the CO-tolerant internal reforming molten carbonate fuel cell), some fuel cell research efforts focused on direct hydrocarbon conversion. These concepts included e.g. solid metal electrolyte and polymer electrolyte fuel cells.

Batteries focused strongly on lithium-ion rechargeable, with substantial effort on nickel-metal hydride and lead-acid, and somewhat less emphasis on rechargeable air-electrode batteries. The lithium-ion battery emphasis was two-fold: increasing electrode intercalation to minimize lithium free-metal formation with non-aqueous electrolytes, and modifying the electrolyte (and electrodes) to increase conductivity and allow higher specific energies. Research emphasis to improve specific energy appeared to be on solid-polymer electrolytes, backed by highly conducting current collectors. The nickel-metal hydride battery is a potential replacement for the balanced performance (specific energy, specific power, cycle-life, reliability), but environmentally contaminating Ni–Cd battery. NiMH emphasized developing metal hydrides for negative electrodes that can be decomposed and reformed reversibly, and have high hydrogen retention to reduce battery self-discharge rate. The lead-acid battery emphasis appears to be materials, packaging, and fabrication. Much of the associated modeling and simulation reflected in the keywords was related to lead-acid battery system optimization, focusing on maximizing energy at optimum power density, minimizing internal resistance, retaining maximum charge, and maximizing mechanical strength and cycle life. Environmental impact of lead-acid batteries, mainly disposal and material recyclability, was a research area as well. The rechargeable air-electrode batteries, such as zinc-air, focused on methods to increase charging rates and reduce electrolyte carbonation from air-based carbon compounds.

There was very little effort reflected in capacitors. Relatively high specific power/specific energy supercapacitors appear to be a key research area, with potential for filling the gap between the high power density low energy density conventional capacitors and the high energy density low power batteries or very high energy density low power density fuel cells. Research focus appears aimed at electrode geometries (larger surface areas) and materials that will allow higher capacitance, and electrolyte materials (and geometries) that will allow the relatively low voltage ceilings (due to potential breakdown across the small separator gap) to be raised.

*4.3.1.2. Abstract taxonomy.* A taxonomy of electrochemical energy-related technologies was developed through visual inspection of the abstract phrase frequencies. The developed taxonomy was subsequently used to approximate global levels of emphasis (GLE). This type of analysis would help identify adequately and inadequately supported system and subsystem tech base areas. It could also differentiate the developed and developing technology components of a particular system.

In this section, a three level taxonomy was required to provide sufficient detail on the various electrochemical energy-related technologies. The first two levels of the taxonomy were developed using a phrase frequency-only analysis. Phrases generated with the phrase frequency analysis could be classified into two generic types of categories: system specific (e.g. solid oxide fuel cell, lithium-ion batteries, steam reforming, electric double-layer capacitors) and generic (electrolytes, capacity, electrodes, discharge, cathodes, anodes). Since one feature of the manually generated taxonomy was allocation of abstract phrases and associated frequencies to specific categories in order to estimate GLE of specific systems, a method was required to relate the generic phrases to their associated specific systems (e.g. what fraction of the electrodes frequencies should be allocated to the batteries or fuel cells categories?).

The method selected was to perform a proximity analysis using the second level taxonomy categories as themes. The second level of the taxonomy consisted of high technical content system-specific phrases from the phrase frequency analysis data. Phrases in close physical (and thematic) proximity to the system-specific phrases were generated, and the more generic tech base phrases were assigned to the related system-specific categories weighted by their occurrence frequencies.

*4.3.1.2.1. Taxonomy level 1.* The highest taxonomy level consisted of two categories: electrochemical converters (17 227) that were comprised of fuel cell technologies, and electrochemical source and storage devices (24 804) consisted of battery and electrochemical capacitor technologies.

An evaluation of the current status of these technologies showed that electrochemical source/storage devices provided power to a wider variety of applications (especially for small/portable electronic devices) compared to fuel cells. As the numbers of applications requiring portable power continue to increase rapidly, near term solutions are required, and batteries currently seem to offer the most feasible solution. Hence, the higher level of emphasis in that category.

The literature indicated that fuel cell research was substantial. The potential of fuel cell technology is promising, but many technical and economic issues such as the need for expensive catalysts (in low temperature fuel cells), the large size/weight/miniaturization, and the corrosion/breakdown of components (for high temperature fuel cells) would need to be resolved for it to become more competitive. Another issue is the lack of infrastructure for some of the fuels (methanol, ethanol, and hydrogen) required to supply these fuel cells.

Supercapacitors remain ideal for delivering high power over a short duration, but are most useful when combined with other electric power sources (i.e. to provide extra power boost in an electric application). Technical advances were still required before higher energy density systems could be mass-produced.

*4.3.1.2.2. Taxonomy level 2. Fuel cells*: Fuel cell research addressed the following tech base areas: system components and component configurations (4038); properties and characteristics (2683); sources/fuels (2385); materials (2358); conversion processes (1363); conversion byproducts (1011); operating conditions (885); and potential applications (267).

In general, fuel cell research was aimed at improving the performance of fuel cells (for a wider variety of applications) while lowering the cost (manufacturing, operation, and maintenance) of the systems, and reducing the size (for portable power applications). Advances in the individual fuel cell components could help achieve many of those objectives. For example:

- use of porous electrodes to increase electrode surface area, and subsequently, improve electrode current density and overall fuel cell performance;
- use of polymer electrolytes to reduce costly corrosion and electrolyte loss problems, and to improve hydrogen ion conductivity (which improves fuel cell performance);
- development of electrocatalysts to improve the rate of electrode reactions/exchange current density.

The importance of advances in component technologies was reflected by the high GLE for that category.

The fuel cell properties/characteristics (such as ionic conductivity, current density, corrosion rate), sources/fuels (such as hydrogen, methane, natural gas, hydrocarbons, methanol, ethanol), and materials (such as nafion, zironia, platinum, ceramics) seemed to have equal amounts of research according to the GLE. All three of these categories provided critical supporting/enabling technologies for improving fuel cells from a performance or cost perspective.

Other important fuel cell categories were conversion processes, byproducts, and operating conditions. Lastly, there was little emphasis on applications. Current fuel cell technologies did not address a large variety of applications (based on range of required energy or rate of power output).

Most of the recent research on fuel cell technologies focused on solid oxide and molten carbonate fuel cells, due to their high operating temperatures and consequently higher efficiencies and relaxed reforming requirements. Polymer electrolyte/proton exchange membrane (PEM) fuel cells were also being pursued, but at a lower level. The fuel cells' main technical areas of interest were the same as described in the keyword taxonomy. For solid oxide fuel cells (SOFC), electrolyte doping, synthesis and characterization of various anode/cathode materials, and the development of low temperature SOFC seemed to be the focal points of research. Molten carbonate fuel cell (MCFC) research was focused on the development of novel cathode materials (often a nickel based alloy for corrosion resistance between the cathode and current collector plate), and the selection of anode catalyst for the reforming reaction. Polymer electrolyte and proton exchange membrane (PEM) fuel cell research was developing new techniques for preparing the catalyst layer in the electrolyte, developing new fabrication methods for composite membranes, and was assessing the physical and morphological characteristics and electrochemical behavior of various catalysts (i.e. PtRu/C).

*Batteries*: Batteries research addressed the following tech base areas: materials (7850), properties and characteristics (4643), component technologies (4531), processes and phenomena (2658), types (2195), and applications (1121).

Materials seemed to be an integral part of advancing battery technologies. The enhanced properties and characteristics of novel materials, and their application in the electrodes, electrolytes, or at the interface were projected to result in lighter, higher energy capacity, less expensive batteries. For example, research was being conducted on:

- porous metals/high surface area materials for NiCd/NiMH electrode;
- hydrogen absorbing metals for NiMH anodes;
- alloy powders for higher energy capacity and improved cycle life;
- carbon coated silicon for anodes to improved cycleability;
- analyses of reactions at electrode–electrolyte (non-aqueous) interface;
- polymer electrolytes;
- carbon anodes/composite anodes;
- microwave synthesis of electrode materials.

Materials were expected to improve properties and characteristics of batteries and battery components (electrodes and electrolytes) with similar impacts. This was reflected by the equal amounts of emphasis in each category, according to the GLE numbers.

Little was discussed in the literature about the basic theory behind the electrochemical processes on which batteries are based, other than material synthesis and characterization. There were some efforts in the areas of modeling and simulation, mainly focused on lead-acid batteries. There was a modest amount of the literature that focused on applications of batteries. The reputation, variety, and current availability of batteries have made them the top option for many applications requiring portable electric power. However, because of the large unit energy content of hydrogen, if fuel cells can be made to operate efficiently and cheaply at small sizes, the portable electric power market share between batteries and fuel cells could shift substantially over the next decade.

Technical area research for battery technologies had a strong focus on secondary lithium batteries (lithium ion, lithium polymer, etc.). Research included the investigation of carbon–carbon (C–C) composite as an anode material for lithium-ion batteries. Lithium-ion cells made with the C–C composite anode showed many advantages, such as excellent performance and enhanced safety. Other research focused on improving the charge–discharge characteristics of polyaniline films used as cathodes in lithium batteries and improving the ionic conductivity of crystalline polymer electrolytes. Lead-acid batteries also continued to be a major thrust area. Research there focused on battery additives and their influence on the separator behavior. There was research on the use of gas-recombining noble metal catalysts in valve regulated lead acid batteries. Also, researchers were developing high rate discharge (HRD) lead-acid batteries, geared for automobile applications. Nickel metal hydride (NiMH) batteries constituted another major thrust area. Research was conducted to develop new/improved methods for synthesizing and characterizing materials. For example, surface modification of hydrogen storage alloy electrodes by the hot-charging treatment was being investigated. New hydrogen storage materials were being developed. Gelatin-pretreated graphite anodes were being tested to see if they could improve the irreversible loss of charge, reversible capacity and efficiency of NiMH batteries.

*Electrochemical capacitors*: Electrochemical capacitors research addressed the following tech base areas: properties and characteristics (604); component technologies (568); materials (435); and types (199). The major concerns were associated with improving the energy density, and power density, hence the higher GLE for that category. Potential solutions for achieving this included improving components (dielectric/gel electrolyte, solid electrolyte; polarizable electrodes/composite electrodes) and improvements based on component materials (glassy carbon, carbon fibers, aerogels, thin films).

A focal point of electrochemical capacitor research was the chemical synthesis and characterization of various materials (i.e. MoxSy(CO)(*n*) and Mo-x(CO)(*n*)) and their possible application as catalysts for oxygen reduction reaction. Other electrochemical capacitor thrust areas included: the use of industrial carbon blacks (CBs) materials in electrochemical supercapacitors because of their high specific areas; the development of thin film supercapacitors using a sputtered $RuO_2$ electrode; the development of carbon nanotubes/$RuO_2$ electrodes for electrochemical capacitors; the synthesis, structure-property characterization, and performance of carbon aerogels; and the fabrication and application of Cu-carbon composite (prepared from sawdust) to electrochemical capacitor electrodes.

The absence of any categories/sub-categories in this taxonomy should not be interpreted that S&T efforts are not being pursued in those areas. The correct interpretation is that within the frequency threshold constraints of the electrochemical database, mid-high frequency phrases related to these categories do not appear.

*4.3.2. Statistical clustering*

Two statistically-based clustering methods were used to develop taxonomies, factor matrix clustering and multi-link clustering. Both offer different perspectives on taxonomy category structure from the non-statistical manual clustering approach described above. None of the three approaches are inherently superior.

*4.3.2.1. Factor matrix clustering*. A correlation matrix of the 218 highest frequency high technical content phrases was generated, and a factor analysis was performed using the WINSTAT statistical package. The eigenvalue floor was set equal to unity, and a factor matrix consisting of 20 factors resulted. A description of these factors, and their aggregation into a taxonomy, follows. The capitalized phrases in parentheses represent typical high factor loading phrases for that factor.

Factor 1 (Cr, Mn, Fe, Ni, Ti, V, Zr, Cu, Mg, Mo, hydride, discharge capacity, electrochemical properties): materials used in electrochemical systems to improve discharge capacity, especially for cathodes; some emphasis on hydrogen storage alloys for metal hydride batteries.

Factor 2 (electrical conductivity, ionic conductivity, Sr, oxygen partial pressure, thermal expansion, SOFC): focuses on rare earth oxides for higher electrical and ionic conductivity and lower thermal expansion electrodes in solid oxide fuel cells.

Factor 3 (methanol oxidation, platinum, ruthenium, catalysts): focuses on the use of platinum/ruthenium catalysts for enhancing methanol oxidation at electrodes of direct methanol oxidation fuel cells.

Factor 4 (La, Nd, Pr, YSZ, SOFCs, thermal expansion): focuses on rare earth oxide composite electrodes, especially with Ysz added to either/both electrolytes and electrodes, to increase conductivity and reduce thermal expansion in solid oxide fuel cells.

Factor 5 (lithium, intercalation, electrochemical intercalation, chemical diffusion coefficient, graphite, secondary lithium batteries): focuses on intercalation of lithium into graphite-based electrodes, as evidenced by chemical diffusion coefficient and other metrics, for lithium-ion secondary batteries.

Factor 6 (anode, cathode, current density, maximum power density, fuel cell, SOFC): focuses on fuel cell performance characteristics (especially SOFC), such as current density and maximum power density, as functions of different electrode materials.

Factor 7 (propylene carbonate, ethylene carbonate, electrolytes, Li, ionic conductivity, graphite): focuses on interactions between lithium-based electrodes and non-aqueous carbonate-based electrolytes for lithium-ion secondary batteries.

Factor 8 (Na, K, Li, electrochemical intercalation, S): focuses on electrochemical intercalation of sodium family-based compounds, especially lithium, especially in graphite, especially for re-chargeable lithium-ion batteries, and the role of sulfur compound additives in enhancing the intercalation–deintercalation process.

Factor 9 (catalysts, oxygen reduction, CB, platinum, nitrogen, methanol): focuses on platinum-based catalysts on carbon-based electrodes for enhancing oxygen reduction in fuel cells.

Factor 10 (hydrogen, methane, fuel cell, storage, ethanol, methanol, catalyst, molten carbonate fuel cells): focuses on conversion of methane (or methanol/ethanol) to hydrogen for use in fuel cells, especially molten carbonate fuel cells, and addresses storage of hydrogen in solid material in parallel.

Factor 11 (thin films, cyclic voltammetry, chronoamperometry, impedance spectroscopy, lithium ions): focuses on electrochemical characterization of thin film electrodes, including intercalation behaviour, mainly for lithium secondary batteries.

Factor 12 (spinel structure, capacity fading, sol–gel method, cathode material, cyclability, rechargeable lithium batteries): focuses on use of sol–gel method to fabricate spinel lithium compounds, and subsequent capacity fading and cyclability as a function of sol–gel materials and environmental parameters, for eventual use in lithium secondary batteries.

Factor 13 (nickel, iron, cobalt, titanium, copper, magnesium, manganese, aluminum, reversibility, electrochemical behaviour, specific capacity, batteries): materials for electrochemical systems, especially for electrodes, especially for batteries.

Factor 14 (X-ray diffraction, transmission electron microscopy, X-ray photoelectron spectroscopy, electrochemical measurements, thermogravimetric analysis, reversible capacity, films, crystal structure): experimental measurements of structures and properties and reactive phenomena to characterize electrochemical systems.

Factor 15 (discharge, charge, lead-acid batteries, electrochemical reactions, power, simulations, mathematical model, corrosion): focuses on lead-acid battery phenomena, such as charging and discharging, electrochemical reactions and subsequent corrosion, including modeling and simulation of battery processes.

Factor 16 (propylene carbonate, ethelyne carbonate, lithium salts, polymer electrolyte, ionic conductivity): focuses on increasing electrolyte electrical and ionic conductivities by varying composition and concentrations of polymer, lithium salt, and plasticizer in polymer electrolytes, emphasizing propylene carbonate and ethylene carbonate plasticizers, for eventual use in secondary lithium-ion batteries.

Factor 17 (electrode surface, hydrogen diffusion, exchange current density, cyclic voltammetry, electrode performance, electrochemical impedance spectroscopy, cathodic polarization, discharge capacity): focuses on measurement of hydrogen diffusion and charge transfer (exchange current density) across electrodes, mainly for their effect on discharge capacity, mainly for metal hydrides as a precursor to NiMH batteries.

Factor 18 (negative electrode, positive electrode, lithium metal, manganese, cobalt, rechargeable lithium batteries, specific capacity, lead-acid battery): focuses on use of layered lithium oxides, especially cobalt and manganese, as cathode materials that can reversibly intercalate lithium ions for lithium-ion rechargeable batteries.

Factor 19 (energy density, capacitor, power density, specific energy, capacitance, battery, electric vehicles, cycle life): primary focus on advanced capacitors with reasonable energy density for high power density high discharge long cycle life applications, with secondary focus on hybrid battery/capacitor systems for high energy density high power density applications.

Factor 20 (reversible capacity, graphite, carbon, tin, anode material, lithium, lithium-ion batteries, first discharge): focuses on use of graphite-tin composites as anode materials for high reversible capacity lithium-ion batteries.

Thus, the 20 factors can be viewed as thrust areas constituting the lowest level taxonomy. Each factor contains one or more of the following elements: (1) system-specific (e.g. lithium-ion battery); (2) system-generic (e.g. re-chargeable battery); (3) class (e.g. battery); (4) phenomenon (e.g. electrolyte conductivity). Thus, there are myriad ways to combine the factors, depending on which dominant characteristics are chosen. In practice, the aggregation methodology will depend on the application for the taxonomy. For example, if the taxonomy is used to identify participants for a comprehensive workshop on electrochemical power sources, categorizing by phenomena would identify, e.g. intercalation experts, while categorizing by system would identify, e.g. lithium-ion battery experts. Selection of aggregation attributes would depend on the workshop objectives. Conversely, assume the taxonomy is used by a program manager to estimate global levels of effort in specific technologies, in order to identify technology areas of adequacy and deficiency. Then, categorizing by phenomena would identify, e.g. intercalation deficiencies, whereas categorizing by system would identify, e.g. lithium-ion battery deficiencies.

The factors above were aggregated by generic system at the highest level, then by specific system (when identifiable) at the next level, or by phenomenon when specific system is not identifiable, then the specific systems were de-aggregated to phenomena at the following level. The following hierarchical level taxonomy resulted (numbers in parenthesis are factor numbers from above).

The highest level taxonomy (level 1) consisted of fuel cells (2, 3, 4, 6, 9, 10), batteries (5, 7, 8, 11, 12, 15, 16, 17, 18, 20), and capacitors (19). The next highest level taxonomy (level 2) is:

1. Fuel cells
   - 1.1. SOFC (2, 4, 6)—improve electric conductivity and thin film properties, and reduce thermal expansion, to increase current density and maximum power density;
   - 1.2. DMFC (3, 9, 10)—catalysts and hydrogen storage alloys to improve hydrocarbon oxidation and hydrogen storage;
   - 1.3. materials (1, 13)—materials for electrochemical systems, including fuel cells;
   - 1.4. diagnostics (14)—diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including fuel cells.
2. Batteries
   - 2.1. lithium-ion (5, 7, 8, 11, 12, 16, 18, 20)—rechargeable;
     - 2.1.1. intercalation (5, 8, 11, 18, 20);
     - 2.1.2. thin films (11);
     - 2.1.3. conductivity (7, 16);
     - 2.1.4. reversibility (12, 18, 20);
     - 2.1.5. fabrication (12);
   - 2.2. lead-acid (15)—modeling, simulation, and performance characteristics measurement;
   - 2.3. NiMH (1, 17)—hydrogen storage alloys structure and reactions;
   - 2.4. materials (1, 13)—materials for electrochemical systems, including batteries;
   - 2.5. diagnostics (14)—diagnostic techniques used to characterize phenomena and properties of electrochemical systems, including batteries.
3. Capacitors
   - 3.1. thin films (19)—capacitance and energy density of thin-film capacitors.

*4.3.2.2. Multi-link clustering*. A symmetrical co-occurrence matrix of the 218 highest frequency high technical content phrases was generated. The matrix elements were normalized using the equivalence index ($E_{ij} = C_{ij}{}^{\wedge}2/C_iC_j$, where $C_i$ is the total occurrence frequency of the ith phrase, and $C_j$ the total occurrence frequency of the *j*th phrase, for the matrix element *ij*), and a multi-link clustering analysis was performed using the WINSTAT statistical package. The average linkage method was used. The main output, a dendogram or tree-like structure, showed the relationships among the phrases, and allowed clusters of related phrases to be identified. A description of the final dendogram, and the aggregation of its branches into a taxonomy of categories, follows.

The 218 phrases in the dendogram are grouped into 14 clusters. These clusters form the lowest level of the taxonomy hierarchy. Each cluster is assigned a letter, ranging from A to N. The cluster hierarchies are determined by the branch structure. Overall, there are two main branches (clusters). The clusters differ in size and coherence. The larger cluster (A–M) covers myriad electrochemical systems and phenomena, while the smaller cluster (N) has a strong focus about materials to improve electrical conductivity and reduce thermal expansion in SOFCs.

The smaller cluster (N) will not be sub-divided for the more detailed taxonomy, while the larger cluster (A–M) can be divided further into two clusters. Again, the clusters differ in size and coherence. The larger cluster (A–M) covers myriad electrochemical systems and phenomena, while the smaller cluster (A–C) has a strong focus about electrolyte and electrode materials to improve electrical conductivity and reversibility of lithium-ion secondary batteries.

The smaller cluster (A–C) can be divided further into its elemental clusters. Cluster A focuses on the use of polymer electrolytes and carbonate plasticizers to increase ionic conductivity within lithium-ion batteries. Cluster B focuses on the lithium intercalation into carbon-based electrodes to increase reversible capacity of secondary lithium-ion batteries, with a secondary emphasis on examination of lithium family member intercalation processes. Cluster C focuses on electronic and diffusion phenomena related to the intercalation process.

The larger cluster (D–M) can be subdivided further into two clusters. The larger of these two clusters (D–J) covers myriad electrochemical system types and phenomena, while the smaller of these two clusters (K–M) is focused on improving discharge capacity, reversibility, and cyclability of lithium-ion secondary batteries.

The smaller cluster (K–M) can be sub-divided into its elemental clusters. Cluster K focuses on materials to improve discharge capacity and cycle life of lithium-ion secondary batteries. Cluster L focuses primarily on materials to improve reversibility of lithium cells, and secondarily on materials to reduce corrosion in molten carbonate fuel cells. Cluster M focuses on material fabrication to improve capacity and cyclability of lithium-ion batteries.

The larger cluster (D–J) can be divided further into two clusters. The larger of these two clusters (D–H) covers mainly battery, capacitor, and fuel cell system performance issues, while the smaller of these two clusters (I–J) covers direct methanol oxidation fuel cells and associated thin film characterization issues.

The larger cluster (D–H) can be sub-divided into its elemental clusters. Cluster D focuses on characterization of electrode performance in electrochemical cells. Cluster E focuses on electrode-driven charge-discharge phenomena in rechargeable batteries, including modeling and simulation of battery performance. Cluster F focuses on preparation of electrode materials by the sol-gel process for use in electrochemical cells in general, and capacitors in particular. Cluster G focuses on electrode-related electrochemical cell performance, with emphasis on solid oxide fuel cells. Cluster H focuses on the direct electrochemical oxidation of methane for solid oxide fuel cells, bypassing the need for hydrogen storage, and addresses carbon deposition reduction at the electrodes by current density variation.

The smaller cluster (I–J) can be sub-divided into its elemental clusters. Cluster I focuses on measurement and characterization of thin film electrochemical properties. Cluster J focuses on methanol oxidation in platinum–ruthenium catalyst-based thin film polymer–electrolyte membrane direct methanol fuel cells.

From these multi-link clustering results, the main research focal points appear to be solid oxide fuel cells (increasing electric/ionic conductivity, direct hydrocarbon oxidation) and lithium secondary batteries (improving electrode intercalation, increasing electrolyte conductivity, increasing discharge capacity/reversibility). Research also emphasizes other direct hydrocarbon fuel cells, and other secondary batteries.

#### *4.3.3. Recommended taxonomy*

The different statistical and non-statistical taxonomies generated above used different methodologies and some different phrases. Therefore, the results are not directly comparable. A taxonomy that reflects the levels of effort and specific research thrusts would have the structure of the non-statistical abstract field taxonomy. It would reflect high emphasis on the solid oxide fuel cells and rechargeable lithium batteries, especially the sub-thrusts identified in the statistical clustering approaches. It would also reflect emphasis on other rechargeable battery approaches, such as NiMH, and on direct hydrocarbon fuel cells. Its capacitor component would be low, with emphasis on thin-film voltage-breakdown-resistant supercapacitors.

## 5. Summary and discussion

Two final observations on the technical results. Based on reading a large number of the retrieved SCI Abstracts, and examining the phrase listings and taxonomy results, there appears to be a large imbalance between theory and experiment. The discipline appears almost Edisonian in nature. In addition, there is little evidence of extrapolation of concepts and insights from other technical disciplines. The research appears very parochial. Some of the hybrid literature-based discovery/multi-discipline workshop techniques [1], performed at the initiation of research projects, could systematically access this extra-disciplinary information.

This paper has presented a number of advantages of using DT and bibliometrics for deriving technical intelligence from the published literature. Large amounts of data can be accessed and analyzed, well beyond what a finite group of expert panels could analyze in a reasonable time period. Preconceived biases tend to be minimized in generating roadmaps. Compared to standard co-word analysis, DT uses full text, not index words, and can make maximum use of the rich semantic relationships among the words. It also has the potential of identifying low occurrence frequency but highly theme related phrases that are 'needles-in-a-haystack', a capability unavailable to any of the other co-occurrence methods.

Combined with bibliometric analyses, DT identifies not only the technical themes and their relationships, but relationships among technical themes and authors, journals, institutions, and countries. Unlike other roadmap development processes, DT generates the roadmap in a 'bottom-up' approach. Unlike other taxonomy development processes, DT can generate many different types of taxonomies (because it uses full text, not keywords) in a 'bottom-up' process, not the typical arbitrary 'top-down' taxonomy specification process. Compared to co-citation analysis, DT can use any type of text, not only published literature, and it is a more direct approach to identifying themes and their relationships.

The maximum potential of the DT and bibliometrics combination can be achieved when these two approaches are combined with expert analysis of selected portions of the database. If a manager, for example, wants to identify high quality research thrusts as well as science and technology gaps in specific technical areas, then an initial DT and bibliometrics analysis will provide a contextual view of work in the larger technical area, i.e. a strategic roadmap. With this strategic map in hand, the manager can then commission detailed analysis of selected abstracts to assess the quality of work done as well as identify work that needs to be done (promising opportunities).

## Acknowledgements

The views in this paper are solely those of the authors, and do not represent the views of the department of the Navy or any of its components, DDL Omni Inc., or Noesis Inc. In addition, the authors acknowledge the contributions of Dr. Karen Swidder Lyons, Naval Research Laboratory, for categorizing the retrieved records for query development purposes, and Dr. Richard Carlin, Office of Naval Research, for sponsoring this effort.

## References

[1] R.N. Kostoff, Science and technology innovation, Technovation 19 (1999) 593–604.
[2] R.N. Kostoff, Database tomography for technical intelligence, Competitive Intell. Rev. 4 (1993) 1.
[3] R.N. Kostoff, Database tomography: origins and application, Competitive Intell. Rev. (Special Issue on Technology) 5 (1994) 1.
[4] R.N. Kostoff, et al. System and Method for Database Tomography, US Patent number 5440481, 1995.
[5] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Database tomography for information retrieval, J. Inf. Sci. 23 (1997) 4.
[6] R.N. Kostoff, The Handbook of Research Impact Assessment, 7th Edition, DTIC Report Number ADA296021, 1997, http://www.dtic.mil/dtic/kostoff/index.html
[7] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography, JASIS, 15 April 1999.
[8] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Database tomography for technical intelligence: a roadmap of the near-earth space science and technology literature, Inf. Process. Manage. 34 (1998) 1.
[9] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, R. Pellenbarg, Database tomography for technical intelligence: comparative roadmaps of the research impact assessment literature and the Journal of the American Chemical Society, Scientometrics 40 (1997) 1.
[10] R.N. Kostoff, T. Braun, A. Schubert, D.R. Toothman, J. Humenik, Fullerene roadmaps using bibliometrics and database tomography, J. Chem. Inf. Comput. Sci. (2000).
[11] R.N. Kostoff, K.A. Green, D.R. Toothman, J. Humenik, Database tomography applied to an aircraft science and technology investment strategy, J. Aircraft 37 (2000) 4.
[12] Science citation index (SCI), Institute for Scientific Information, Philadelphia, PA, 1999.
[13] R.N. Kostoff, The underpublishing of science and technology results, The Scientist 14 (2000) 6.
[14] E. Garfield, History of citation indexes for chemistry—a brief review, JCICS 25 (3) (1985) 170–174.
[15] R.N. Kostoff, The use and misuse of citation analysis in research evaluation, Scientometrics 43 (1998) 1.
[16] M. MacRoberts, B. MacRoberts, Problems of citation analysis, Scientometrics 36 (1996) 3.