Indexing the World of Information

Australian Society of Indexers International Conference

12–13 September 2003
Sydney, NSW

Table of contents

Papers

External link

Reports

Changes in *The Health and Ageing Thesaurus* and reindexing in Health *Insite*

Prue Deacon
Health Insite *Editorial Team*
Australian Government Department of Health and Ageing

This paper looks at the practical issues of aligning a database to updates of a thesaurus, with particular reference to the Health Insite *database and the 6th edition of* The Health and Ageing Thesaurus. *With the 6th edition, some new terms were added to the thesaurus to enable better indexing in the areas of disease prevention and chronic disease. This affected over 20 per cent of the records in the Health* Insite *database and reindexing took some months. The costs and benefits of reindexing are discussed.*

To remain useful, a thesaurus must evolve with the terminology of its discipline. However a thesaurus does not usually exist on its own. It is a component of one or more information retrieval systems and so it is likely that development must also take into account the parameters of particular information retrieval systems—software capabilities, search mechanisms and indexing procedures.

Generally thesaurus changes are batched into regular updates (perhaps annual or biennial) to allow incremental but balanced evolution. Occasionally niggling concerns raised by database indexers or searchers will come to a head and require significant changes to the thesaurus.

Thesaurus changes will in turn have implications for the databases. When a thesaurus changes, database managers must decide whether to reindex the affected older records in the database or rely on searchers adjusting their searches to match the terminology of the period they are searching. Generally larger databases would not be reindexed because of the cost. For smaller databases reindexing might be feasible—and it may be more necessary because the inconsistencies between old and new indexing may be more obvious. (Libraries have to make a similar decision —whether or not to reshelve collections—when there are changes to the shelf classification scheme.)

In the case of *The Health and Ageing Thesaurus* [1] , a very significant change occurred with the latest edition. The scope of the term 'prevention and control' was narrowed to public health aspects only. Three new terms were introduced ('living with disease', 'primary prevention' and 'prophylaxis') and the scope of the term 'disease management' was widened to cover some areas where 'prevention and control' might have been used in the past.

Now, while Health *Insite* [2] staff were the main instigators of this change, the effect was that over 20 per cent of the records in our database had to be reviewed. Why didn't we let sleeping dogs lie?!

Relationship between Health *Insite* and *The Health and Ageing Thesaurus*

Health *Insite* is a gateway website managed by the Australian Government Department of Health and Ageing. It links to over 10,000 Australian health resources on the sites of our 65 information partner organisations. Each resource has a metadata record in the Health *Insite* database. The most important metadata element is the subject element which utilises *The Health and Ageing Thesaurus* .

The Health and Ageing Thesaurus is managed by the Library of the Department of Health and Ageing. It is now in its 6th edition. It is based on MeSH [3] , with a similar hierarchical

structure but far fewer preferred terms. The Thesaurus does not include history notes, although lists of new, changed and deleted terms appear in the introduction to each edition.

Some features of how Health *Insite* uses the Thesaurus are:

- Indexing depth is somewhat greater than in cataloguing—most records have 5–10 thesaurus terms.
- Indexing is as specific as possible.
- 'Subheading' terms are used but they are free floating. This leads to a loss of precision—for example, if a resource has the indexing terms 'pneumonia', 'pain' and 'therapy' then you cannot be sure if it is about therapy for pneumonia or therapy for pain or both.
- There is no weighting of terms.
- The thesaurus hierarchy numbers are built into the database so that they can be used in searching.
- The browsing structure in Health *Insite* consists of a range of topics arranged in a hierarchy and also accessible from a topics list. Topics range from broad (for example, 'Conditions/Diseases') to narrow (for example, 'Respiratory conditions in childhood'). Most topic pages have a dynamic search embedded within them, so that users see the latest list of resources relevant to the topic. These searches rely heavily on the thesaurus terms used in the metadata.

The thesaurus term 'prevention and control'

When indexing a resource about a disease, indexers routinely consider a set of possible 'subheading' terms from the thesaurus (such as 'causes', 'symptoms', 'therapy', 'prevention and control'). Most of these terms were adopted from the MeSH list of subheadings.

However, as noted above, they are free floating terms in Health *Insite* . The MeSH definition of 'prevention and control' is:

Used with disease headings for increasing human or animal resistance against disease (e.g., immunization), for control of transmission agents, for prevention and control of environmental hazards, or for prevention and control of social factors leading to disease. It includes preventive measures in individual cases.

In practice, it was a very easy term to use in Health *Insite* —and to overuse. We used it for how an individual could avoid a disease as well as the public health procedures in controlling a disease. We also used it for the individual control of a disease in the sense of preventing complications or preventing flare-ups of a chronic disease.

This broad use had been of concern for some time because of recall/ precision problems with some topic queries. However it was a user feedback query which persuaded us to finally take action. The user wanted to find some information on preventing breast cancer but was overwhelmed by resources about mammograms. Now mass screening with mammograms is an important public health measure to control breast cancer through early diagnosis and prompt treatment. It is a preventive measure at the population level, but at the individual level it is about diagnosis, not about how to prevent the disease.

We realised that our indexing did not enable us to separate resources about primary prevention (likely to be of particular interest to consumers) from those about public health measures. We also had a problem with chronic diseases like asthma, where preventing an asthmatic person from having an asthma attack is quite different from preventing asthma in the first place.

Since consumers are a key target for Health *Insite* , we recommended some thesaurus changes which were adopted. The Health *Insite* database is relatively small and our usual practice has been to reindex the database to match thesaurus changes. One reason for doing this in a web gateway is that the resources themselves are being frequently updated

and metadata upgrading needs to be part of resource maintenance. In addition there are usability issues for both users and content managers. For example, when indexing a particular concept, it is useful for indexers to be able to check the database to see how that concept has been indexed before—it is confusing if the database is inconsistent.

There are also software reasons for reindexing. For example, in the Health *Insite* metadata creation tool, the validation program for subject terms is unable to cater for multiple versions of the thesaurus. The downside of reindexing is the time required. In the case of 'prevention and control' this was likely to be substantial because a large proportion of the *Changes in* The Health and Ageing Thesaurus Health *Insite* database (over 20 per cent) was affected.

Reindexing process

Below are the new terms introduced into the thesaurus (with scope notes) and the revised scope notes for disease management and prevention and control:

Living with disease: Covers social, psychological, physical, and lifestyle issues. Add name of specific disease

Primary prevention: Prevention of disease in susceptible individuals through promotion of health, including mental health, and specific protection, as in immunisation, as distinguished from the prevention of complications or after-effects of existing disease (that is Disease management)

Prophylaxis : The administration of a drug or therapy to prevent disease

Disease management: Use as an additional heading for the management of chronic disease and the prevention of flare-ups and/or complications

Prevention and control: Use for the public health aspects of prevention and control of particular diseases. Covers the activities associated with developing control mechanisms, national responses and procedures to prevent the occurrence of outbreak of diseases

The review was done over a period of ten weeks. Preliminary guidelines for reindexing were defined and a supplement was prepared after the first batch of reindexing. The reindexing was done by two members of the Health *Insite* Editorial Team, consulting with the rest of the team as needed. For example, we sometimes needed to discuss the effects of reindexing on search strategies.

The crucial guideline when considering a resource was: is this about someone who has a disease? (in which case treatment/management terms are needed) or is it about someone who does not have the disease? (in which case the prevention terms are appropriate).

We found that the resource had to be checked again as well as the metadata record in about half the cases. Often it was necessary to check all the indexing terms, not just 'prevention and control'. The most difficult cases were the complex disease processes, such as the lead-up to stroke or heart attacks (with risk factors like high cholesterol, high blood pressure and atherosclerosis). In some cases the therapy was the same as the prevention method—for example, a high fibre diet for constipation —and we had to decide whether the resource focus was prevention or therapy. Of course many resources are about both prevention and therapy and both aspects can be covered in indexing. A third area of difficulty was the prevention of infectious diseases, where the main process of control at the population health level is also primary prevention at the individual level—for example, washing your hands before preparing food. We worked out some guidelines so that only a small proportion of records were indexed as both 'primary prevention' and 'prevention and control'.

We did not keep a record of the time taken for this project. The review and reindexing was often done in small batches between other jobs. Sometimes slow system speeds held us up. We estimate that the task took between 3 and 6 minutes for each record, with a total time of over 100 hours and maybe as much as 150 hours.

Results

The number of records checked was 2277. Of these, 164 were not included in the analysis—reasons for noninclusion were 404 errors, records being flagged for removal, topic pages being deferred till later and miscellaneous queries. See Table 1 .

Discussion

One of the principles of thesaurus development is to provide a balance with respect to the key databases for which it is used. This means not getting too specific in areas where there are likely to be few relevant resources in the databases and also trying to split terms that are retrieving far too many results in a database search. It was good to see high numbers of records for both 'prevention and control' with its new definition and 'primary prevention' —this indicated that it was worthwhile to make the distinction. 'Prophylaxis' was well used. 'Disease management' was less well used—however both 'disease management' and 'living with disease' will get more usage when other areas of the database are reviewed.

Table 1. Reindexing decision made on the use of 'prevention and control' (P&C) in a record

Reindexing decision	Number of Records	%
No change (ie meets new definition of P&C)	433	20.5
Changed to Primary Prevention	643	30.4
Changed to Prophylaxis	160	7.5
Changed to Disease Management, Living with Disease or another therapy term	177	8.4
Changed to another term not listed above	39	1.8
Deleted	661	31.3
Total	2113	

The biggest surprise was the percentage of records where 'prevention and control' was simply deleted. In many cases the resource was about preventing complications of a disease and we had decided to index this as 'complications' without any prevention term. In other cases a therapy term or a more specific prevention term (for example, 'immunisation') was more appropriate and was already in the record. In a lot of cases, the resource was not really about prevention at all and it was a puzzle that it was indexed as such.

The reason lies in the indexing procedure noted earlier: when indexing a resource about a disease, indexers routinely consider a set of possible 'subheading' terms and quickly check them off, for example: causes—yes, symptoms—yes, diagnosis—no, therapy—no, prevention and control —yes. Sometimes you fail to register that there are only a couple of sentences about prevention and that the main focus of the resource is one of the other terms. Furthermore, when a resource is about chronic disease —about controlling symptoms or being in control of the disease—it is easy to bend the definition and think that maybe 'prevention and control' is a suitable term.

Given that we had been aware of problems with 'prevention and control' for years, it is of some regret to us that we did not make changes to the thesaurus much earlier. One of the reasons was a reluctance to keep moving away from MeSH with each edition of the thesaurus. At one stage we thought of the thesaurus as a cut-down version of MeSH with an Australian slant—and hoped that it could be incorporated into the UMLS 4 (a medical meta-thesaurus). Development of the thesaurus is now tending to give precedence to the needs of our own databases.

(It should be noted that, when MeSH is used in Medline, the indexing process and the search software provided in many versions of Medline allow for much higher recall and precision than we can achieve in Health *Insite* . The problems that we have had with 'prevention and control' might not occur.)

Now that the reindexing has been done, it will be possible to review the Health *Insite* topic pages relating to disease prevention. Depending on the number of resources retrieved by topic queries, there will be some diseases where separating the different aspects of prevention will be very useful. Tightening our indexing procedure so that 'subheading ' terms are considered more carefully, rather than ticking off all that seem to apply, will make topic results more precise.

We are hoping to develop a search application which will make the complete thesaurus available to users and guide them in constructing a search, using thesaurus terms and hierarchies. The reindexing will mean that a user with a query like 'How can I avoid getting Disease X' could be given a much better search result.

Conclusion

This project showed that the thesaurus changes in the 'prevention' terms enabled a useful partitioning of the Health *Insite* database where the term 'prevention and control' had previously been overused and sometimes incorrectly used. The real benefits will show as the Health *Insite* topic pages are reviewed and more precise searching is possible in the area of disease prevention. Further benefits will come as we develop the search functionality of Health *Insite* to allow guided thesaurus searches.

We feel that the benefits of reindexing justify the costs (estimated at 100–150 hours work). The situation would have continued to worsen without the thesaurus changes and without reindexing.

What are the implications for other database indexers and managers? Firstly, it is important to keep a whole system perspective. Indexers should never regard themselves as 'merely the indexer'—their alerts that particular concepts are hard to index with the thesaurus, or that thesaurus terms are not well-defined, are vital for both database managers and thesaurus developers. Secondly, database maintenance to match thesaurus changes can be expensive. Database managers need to weigh the costs against the benefits of having consistent indexing.

Acknowledgements

References

1. *The Health and Ageing Thesaurus* , 6th edn, Commonwealth Department of Health and Ageing. 2003 www.health.gov.au/thesaurus.htm
2. Health *Insite* www.healthinsite.gov.au
3. *MeSH: Medical Subject Headings* , US National Library of Medicine www.nlm.nih.gov/Mesh
4. *UMLS: Unified Medical Language System* www.nlm.nih.gov/research/umls *Changes in* The Health and Ageing Thesaurus

Paper from Indexing the World of Information: International Conference of the Australian Society of Indexers, Sydney, 12–13 September, 2003.

A PASSIONATE SEARCHER'S VIEW OF INDEXING AND INDEXERS

Elizabeth Swan

In a career of over 40 years, this passionate searcher migrated from manual searching of voluminous printed indexes to online searching in 1975 and now searches the Web and services such as Dialog, Factiva and Informit for clients. In a Google dominated online world, indexing is still important but as information professionals we need to debate the virtues of codes and controlled versus uncontrolled indexing, as well as the challenges of bibliographic searching versus fulltext searching. The business of indexing also needs to be addressed - how can indexing time lags be minimised, how can costs be contained? What are the innovative indexing initiatives of major players such as Factiva, Derwent, Dialog, and others? How successful are tools such as mapping, weighting algorithms, proximity operators and qualifiers, and clustering search engines? Do metatags solve the problems of and reveal the Invisible Web? What are the keys to success for human indexers in a fulltext search engine world?

A PASSIONATE SEARCHER'S VIEW OF INDEXING AND INDEXERS

On a hot summer's day in Melbourne in the late 1960's I was doing a literature search for one of the scientists using several tools - *Biological Abstracts, Index Veterinarius, Veterinary Bulletin* , and so on. The CSIRO Animal Health Research Laboratory library in the Veterinary Precinct at the University of Melbourne had wooden shelves to the ceiling, wooden ladders, shiny, highly polished vinyl tiles. I climbed to the top of the ladder to "retrieve" a large blue bound volume of *Biological Abstracts* from the top shelf. The bindings had stuck in the heat and I had to use some force to separate the volume from its neighbours. The force not only removed the volume but it also moved the ladder! As I tumbled to the floor on top of the ladder with my skirt over my head, a startled research scientist who had been snoozing in the large leather chair woke up and said "I know the librarian is alive as I can hear muffled swearing emerging from under her skirt"!

Not long after that, the National Library announced a series of MEDLARS workshops - the batch predecessor to MEDLINE - and I was keen to attend and learn how I could use a computer to do literature searches. Betty Doubleday the CSIRO Librarian at the time said "Bah! There will never be a role for computers in libraries" and so my odyssey as "a passionate searcher" began.

Not being convinced the Chief Librarian was right, I left CSIRO, flew to Europe and killed a few birds with one stone by getting a job in continental Europe , with a database producer (Embase) in Amsterdam - close enough to the ski slopes of Europe and so on.

I thought I needed to learn how to program but a wonderful mentor Dr. Pierre Vinken, at the time Director of Excerpta Medica, advised me there were many other things other than programming I needed to learn and in a shortish stint at Excerpta Medica (10 months in all) I was introduced to many things that I doubt I could have learned in a library.

I learned about:

- The politics of database production and the competition between MEDLINE and EMBASE and the results of different indexing philosophies at the commercial Excerpta Medica and government National Library of Medicine. At EM the indexing terms were chosen by subject specialists and supported by a massive thesaurus called MALIMET. At NLM the indexing terms were chosen by librarians supported by MESH headings.
- The importance of secondary terms and codes.

- The difference between abstracting and index services, and between manual sources and electronic sources. In fact, I was instructed to write a letter to the editor of the *British Journal of Industrial Medicine* explaining the differences as the difference between a laboratory animal and a human!
- The value of overlapping services - even if they are competitors - BIOSIS, EMBASE, MEDLINE etc.
- The business of database production - marketing, negotiating, pricing, staffing etc.
- How Utopian and unattainable a universal indexing language is when I wrote the Dutch submission UNISIST and attended the UNISIST meeting at UNESCO in 1972.

On return to Australia, I reluctantly accepted a job in the library at ACI in November 1972 - reluctant because I liked the business side of our information world - but ACI was intriguing. In 1972 the library had its own database, skilled staff in a business library in Melbourne and a technical library in Sydney , a budget of $250,000, and access to a strong computer department that was later to develop AUSINET. Our own internal database then called LISARD, an acronym (chosen by the Systems Librarian Dagmar Schmidmaier) for Library and Information Service Automated Retrieval of Data was a bibliographic database with abstracts. [1]

Between 1972 and 1975 having seen STAIRS demonstrations in Australia and Europe , I recommended to the ACI computing department that they lease the STAIRS programs and offer external services as well as internal services, but my suggestions were rejected. In 1975 I had also seen Dialog in operation at Shell in The Hague and late in 1975 I introduced online searching to ACI with access to Dialog and Orbit and a year or two later Finsbury's Textline which later became Reuters Business Briefings.   By the end of the 1970's ACI Computer Services introduced AUSINET using STAIRS and we in the library eventually were able to move our internal database and even our catalogue to the STAIRS platform! So my progress as *a passionate searcher* ” and now *a database producer* continued.

In 1988 ACI was taken over. I was made redundant but I was invited by the State Library of NSW to establish Information Edge as a joint venture. I bought the LISARD database, renamed it EDGE, and became an Information Broker. In 1996 I acquired the State Library's shares in the business and so I not only became *a passionate searcher* but I now became *a passionate searcher with a business* to maintain.

I don't intend to bore you with a biography. Instead I would like to address a number of issues that I feel strongly about as *a passionate searcher* and I would like to explore issues that I think as information professionals we need to debate. I intend to address:

- The role of indexing and indexes today in a fulltext world. Why is indexing still important, even in a Googled world? Searching for high recall versus high precision.
- The value of controlled versus uncontrolled indexing, the use of codes, the value of weighting, the need to be able to limit searches easily to date ranges, the challenges of bibliographic searching and fulltext searching
- The business of indexing, how timely, how cost efficient
- How some of the big players index? Factiva? Promt? Derwent?   And closer to home databases available on Informit?   How valuable is mapping?
- Experiences using weighting tools and proximity operators and qualifiers versus indexes
- Internet searching - Google, clustering search engines, the Invisible Web, metasearch engines

The role of indexing and indexes today in a fulltext world. Why is indexing still important, even in a Googled world? Searching for high recall versus high precision.

The Google spiders index whatever comes their way - whether they are peer reviewed or not, whether the documents are about Uncle Bat's fishing trip, or a CSIRO fisheries experiment, and the spiders cannot type so they do not venture too far into many sites with rich contents- especially databases and other dynamic content. The same can be said of other search engine spiders. So already we have some key advantages of human indexers - they can type, they do concentrate usually on information resources with some quality. But more than that, human indexers can evaluate the material they are indexing and with that evaluation provide clues to searchers as to what they will find when they retrieve a document.

When considering searching, it is important I believe to recognise that there are different types of searches. Some require high precision – every article must be highly relevant – others require high recall – i.e. we will take some irrelevant articles as long as we can be sure we have found absolutely everything. These latter searches are vital for so called "prior art" searches – searches used in intellectual property type scenarios. Then there are searches that require a few inspiring articles say for a lecture or articles to measure media coverage where both high precision and high recall are needed over a recent time frame. Searching therefore at a professional level is not simply a question of "plug and play" – plug a couple of terms into a search box and accept the results.

The value of controlled versus uncontrolled indexing, the use of codes, the value of weighting, the need to be able to limit searches easily to date ranges, the challenges of bibliographic searching and fulltext searching

What sort of indexing are we talking about? Are we talking about Thesaurus controlled indexing? And if so how complicated is the Thesaurus? Is it a multi-volume type job like the LC subject headings? Is it a small list like the old *Economic Literature Index* - 9 pages arranged in 2 alphanumeric columns?

Big professional databases such as *Chemical Abstracts, Engineering Index, MEDLINE, Embase, and BIOSIS* all use controlled indexing. Big professional online hosts such as Factiva use controlled indexing. Some of the News Limited databases do too. But others do not. And Google does not. I wonder if the advent of online fulltext searching means that the value of controlled indexing is diminishing? Which searchers for instance can afford to purchase a copy of every relevant thesaurus? Many systems these days – Informit, Factiva, Dialog – allow cross file searching so there are even dangers relying on one thesaurus that may not be relevant for the files being searched.

Being able to date limit to me is absolutely essential. And by date limiting I am not interested in knowing when a record was indexed. I am interested in knowing when that information was first published or first amended. It is no different between knowing when a book was published or a new edition published as opposed to when it was catalogued or reprinted. It is disappointing to me when I need to search Australian databases (or WWW sites) when I find there is no provision to be able to identify easily new information. I am amazed how often I am unable, especially with collections of Australian data on the Web, to determine when the information was first released – so there is this fuzziness of how current or how obsolete some information actually is.

Of course in recent years there has been a huge growth in fulltext databases. And one of the issues for indexers is – how can you reveal in an index both major and minor concepts of importance? Let's say there is an article about "hot filling of PET containers" – that is the major concept – but there is also a minor mention – but significant to some that there is a "waisted bottle capable of hot fill that has been made for Coca Cola". How in an index, is one able to show the difference emphases of these two concepts?

Personally I am also a great believer in the use of codes. Of course we use codes when we catalogue a library collection but the classification systems are used to arrange physical collections. In a database such as COMPENDEX or DERWENT or PROMT and many others CODES can be enormously powerful search aids.

The business of indexing, how timely is the indexing, how cost efficient?

To me the key value of indexing is to lead me to material I may not find easily any other way. But I am often looking for new information in a business competitive world. What is being published about competitive products? What are the business relationships between companies or people? What is known about a particular science? Certainly I sometimes do work of an historical nature – how have certain practices evolved? But very, very rarely am I asked to do searches of a totally time insensitive nature. I would say that 97% of my searches involve seeking very recent information. So immediately this raises the issue of timeliness of indexing.

It frustrates me enormously that some Australian databases are appallingly out of date with their indexing. It seems to be a pattern that follows the well known "cataloguing backlog" which I personally was familiar with in my CSIRO days. But when many files on the major commercial providers are loaded several times daily, such as the newswires, then manual indexing of online databases that are 6 months *or more* delayed, is clearly totally unsatisfactory to most searchers and in my view simply should not be tolerated. I would go as far as saying such out of date indexing is unprofessional.

This also raises the issue of cost efficiencies. How much should it cost to index articles for a bibliographic database? Because labour is the key cost of indexing, this really boils down to how quickly an article can be indexed and how quickly will boil down to the procedures to be used. At one end of the spectrum is a haphazard "Pick a couple of terms" to the thorough detailed index of a service such as *AESIS* for stance. *AESIS* was an earth sciences database created by the Australian Mineral Foundation. It had an excellent, thorough well constructed thesaurus and in addition to subject and author index terms it also had geographic coordinate indexing so one could search the database for records relating to a geological deposit in a certain region of Australia. There is no question this was a thoroughly professional database with excellent consistent indexing.

But also it would appear that assuming a base grade librarian could perform the indexing at a salary of say $40,000, with salary associated costs of about 33% this equals $52,000 or $37.14 per hour cost. [2] This means that for a database adding say 200 records per month at one per hour, the cost of indexing alone would be $7,500 and for a year $90,000 for the indexing alone. Add the cost of the primary source material, computing, marketing, copyright fees etc this makes the cost quite significant requiring a sufficiently strong market to justify those costs. And this is a problem I believe for Australian databases. We simply do not have a big enough market to make such services economically viable.

Our experience indexing the EDGE database in Information Edge has been different. We have never, even when part of ACI, been able to afford complex detailed indexing. We currently expect an average of 3 articles per hour to be indexed. This is quite hard for some types of articles – some of the longer and more obscure management articles. But for other trade type articles experienced indexers can complete 6 or more an hour. But indexing rules have been specifically designed to

- make the database workable while
- being economically viable.

Briefly we use a combination of the following:

a)   Subject codes , e.g. financial management = 041, strategic management = 042, HR management =060 and so on.
b)   Indexers are instructed to choose any terms to represent the major concepts of an article and supplementary terms not already covered in the abstract that may be useful for retrieval, e.g. BEER MARKET SHARE and STATISTICS
c)   A database manual outlines the indexing philosophy

d) Indexers are not permitted to use abbreviations unless they are in a 4 page list of controlled terms, e.g. AUST, ROI, WWW. This list is periodically reviewed and updated lists of approved abbreviations released.

WE do not have a complex thesaurus and I acknowledge this is not as professional as databases like MEDLINE or AESIS.   However MEDLINE has a huge global market and can afford to spend more on indexing than we can. EDGE has managed to be profitable even with a small customer base.

So it seems to me timeliness and cost efficiency are both goals that should be paramount.

How do some of the big players index? Factiva? Promt? Derwent?   And closer to home databases available on Informit?   How valuable is mapping?

PROMT when owned by Predicasts was one of my favourite databases to use. I found it a fascinating and rewarding database to use largely because of the quality of its indexing. PROMT originally was an acronym for Predicasts Overview of Markets and Technology and Predicasts had taken the SIC [3] codes as a foundation and developed their own detailed classification system for products, markets and technologies and secondary attributes including countries (e.g. CC=8AUST = Australia), events (e.g. EC=6 = market data and trends, EC=01 = forecasts, trends, outlook) and the codes cascaded.

Because the codes were quite detailed it sometimes was very easy to achieve a high recall and high precision with a simple search statement such as:

s pc=2076266 and ec=6   and cc=8aust and py=1999:2003

  to answer the question

What market statistics are available on sesame seed oil in Australia for the years 1999-2003

PROMT now belongs to Gale, and as a long time user of the database it seems the costs of indexing the database have been reduced by limiting the indexing and it is not as powerful as it once was. Codes, I am told, no longer cascade and must be truncated. Dialog, Gale, Informit are database hosts that offer access to individual files. For both Dialog and Informit, it is possible to carry out multifile searching. Both services have made attempts to standardise field names but clearly cannot standardise the indexing terms. Dialog takes multifile searching further and offers to REMOVE DUPLICATES if the titles are identical. So searching for instance MEDLINE, EMBASE, BIOSIS may result in say 300 records. But after removing duplicates one may end up with a final set containing say 150 records. Quite impressive.

But even so in constructing the searches one has to be careful if using the controlled indexing terms to make sure the idiosyncrasies of each of the 3 thesauri are taken into consideration. Factiva takes a different approach and while it is distributing a lot of data that is also distributed by Dialog (and other services) in individual files, Factiva extracts the individual records and places them in source groupings that can be searched together.

From 2000-2002, I was a member of the Factiva Information Professional Advisory Board which met once a year for those 3 years. This was fascinating for me as Factiva addressed the issues of blending the Reuters Business Briefings service with Dow Jones Interactive (DJI). I was a critic of RBB because I found the indexing almost useless and the search engine so crude that only very limited search statements could be used. So I frequently downloaded hundreds and hundreds of articles to visually browse through to find the material I needed. DJI on the other hand had evolved from a STAIRS based system and had a wonderful search engine with which I once constructed a search statement with about 20 lines of text!

The management team at Factiva under the leadership of the CEO Clare Hart is impressive and very responsive to the needs of information professionals. Indeed Factiva employs

several information professionals in key roles.   The term "intelligent indexing" was adopted by Factiva to describe the approach Factiva would take to indexing in the new child to be born of the RBB/DJI joint venture. During discussions about this I was anxious as I was critical of the RBB indexing, and I dreaded losing the functionality of the DJI search algorithm.   Has it been a success?

The search engine is fine in terms of functionality. At this stage I will be equivocal about the indexing.   Recently I was searching for information about "innovative and successful corporate event management". My client is an event management company that specialises in managing events in the corporate sector. While their business is related to conference and exhibition management, similar to conference organisers that help manage ALIA conferences, their business is not dealing with public events but with private events within the corporate sector. For this search I could not use the Factiva indexing as I found the indexing was neither comprehensive enough nor was it consistent enough.   Here are a few examples that illustrate the inconsistent indexing and how some indexing terms were used for articles containing different concepts.

Save ResultsFactivaDow Jones & Reuters
 HDTHINK TANK - **CORPORATE EVENTS** - In praise of parties.
 BYBy Emma Reynolds.
 WC1,709 words
 PD31 July 2003
 SNMarketing Event
 SCMAREVE
 PG16
 LAEnglish
 CY(c) Marketing Event, a Haymarket publication www.haymarketgroup.com, for more
  information visit www.brandrepublic.com or email info@brandrepublic.com
 LP
 When the going gets tough, **corporate events** are often the first activities to be cut from
  the marketing budget. Bad move. Our expert panel tells Emma Reynolds why.
 The **corporate event industry** received a welcome boost in April's budget when
  Chancellor Gordon Brown doubled the tax-free limit on individual spend at corporate
  parties to £150.
 NS
 ncal: Calendar of Events | ncat: Content Types | nrgn: **Routine General News**
 PUB
 Haymarket Business Publications Ltd.
 AN
 Document MAREVE0020030731dz7v0001c

This article is clearly a lengthy article about CORPORATE EVENTS being useful for MARKETING within the CORPORATE sector but note the index terms above?

Â(c) 2003 Dow Jones Reuters Business Interactive LLC (trading as Factiva). All r ights reserved.   FactivaDow Jones & Reuters
 HDSPECIAL REPORT VENUES - Long-distance learning.
 WC1,530 words
 PD12 June 2003
 SNMarketing Week
 SCMKTW
 PG35
 LAEnglish
 CY(c) 2003 Centaur Communications Limited or its licensors.
 LP

The benefits to be had from **taking staff out of the work environment for training and debriefings still far outweigh the financial outlay - especially as organisers are finding new ways to get the most from your money,** says Ian Whiteling

There's a major project on the horizon and staff need to be briefed and trained to make sure it's delivered efficiently and effectively. Under the current tough economic conditions, there's a tendency for companies to prepare staff for such a job in house or close to the workplace to avoid time away from the office and unnecessary expenditure. This seems a logical step to take, but is it really a false economy? Would staff respond better away from the distractions of the office?

IN
i8395411: **Convention/Trade Shows** | iadv: Advertising/Public Relations
NS
gtour: Travel | gcat: Political/General News
IPC
United Kingdom
PUB
Centaur Communications Ltd.
AN
Document MKTW000020030616dz6c00027

Here is another lengthy article that is clearly about how successful corporate training events can be. Note the indexing terms.

HDBusiness Day (South Africa) - Massive growth in industry.
BYBy David Jackson.
WC549 words
PD30 April 2003
SNBusiness Day (South Africa)
SCMEWBUD
PG15
LAEnglish
CY(c) 2003 Chamber World Network International Ltd
LP
Massive growth in industry Events can link consumers to a brand or company, writes SUCCESSFUL mega events such as the 2003 Cricket World Cup, with the international television exposure they bring, have underlined the dramatic growth of sponsorships and events management **as an essential   ingredient in the overall marketing mix.**

The **events management industry** has mushroomed in SA and the growth in sponsorship spend has increased dramatically over the past two decades.

NS
c31: **Marketing** | c32: Advertising | ccat: **Corporate/Industrial News |**
gcat: Political/General News | gspo: Sports/Recreation | ncat: Content
Types | nrgn: Routine General News
IPC
Company News | General News | **Marketing** | Sponsorship | Sports | 71 |
71121 | 711211 | South Africa | Sub-Saharan Africa
PUB
Financial Times Information Ltd
AN
Document MEWBUD0020030501dz4u0001z

Â(c) 2003 Dow Jones Reuters Business Interactive LLC (trading as Factiva). All rights reserved.   FactivaDow Jones & Reuters

HD **Event Management** - What it Means to Make your Conference or Seminar Delegate Driven.
BYBy Glenn Baker.
WC1,551 words
PD2 April 2003
SNManagement Magazine
SCMANAGM
PG48
LAEnglish
CY(c) 2003 Profile Publishing Limited
LP

Staging a successful conference or seminar involves more than just applying the latest technology. It's about having a clear objective, an appropriate venue, expert help, and being delegate driven. Nothing motivates quite like a well-organised get-together involving key leaders. In his book Jack, former General Electric CEO Jack Welch highlights the importance that his **Crotonville management seminars** played in sharing ideas and catalysing the success of the multi-billion dollar

company.
NS
C31: **Marketing** | C315: **Conferences/Exhibitions** | CCAT:
Corporate/Industrial News | NCAT: Content Types | NEDC: Commentary/Opinion
RE
AUSNZ: Australia and New Zealand | NZ: New Zealand
RBBCM
HEADCNTY:NEW ZEALAND | INX:LOW | C::4I8395411:H::01 | C::3C31:L::04 |
BIPAZ | TOTTED
AN
Document managm0020030412dz4200060

Â(c) 2003 Dow Jones Reuters Business Interactive LLC (trading as Factiva). All rights reserved.   FactivaDow Jones & Reuters

HD **Event Measurement Conference** Slated; Intensive One-day Event for Marketing Pros
WC518 words
PD25 October 2002
ET00:06
SNBusiness Wire
SCBWR
LAEnglish
CY(Copyright (c) 2002, Business Wire)
LP
FRAMINGHAM, Mass.--(BUSINESS WIRE)--Oct. 23, 2002--Successful Shows announces the Event Measurement Conference, to assist marketing professionals with the techniques and strategies to measure the performance of face-to-face marketing efforts. EMC will be held on Wednesday, Dec. 4 at the Sheraton at Woodbridge Place in Iselin , N.J.

Skip Cox, a faculty member and president of Exhibit Surveys Inc. of Red Bank, N.J., says the Event Measurement Conference is a high impact, full-day seminar for sales and marketing executives, trade show and event managers, marketing communications managers, and meeting and event planners to gain insight and hands-on tools so that they can make informed decisions about trade show and event marketing participation.

CO
EXHSUR: EXHIBIT SURVEYS INC
IN
IACC: Accounting/Consulting | IBCS: Business/Consumer Services | ICNSL: Consulting
NS
C31: **Markets/Marketing** | C315: **Conferences/Exhibitions** | CCAT: **Corporate/Industrial News**
RE
NAMZ: North American Countries | USA : United States | USE: Northeast U.S. | USNJ: United States - New Jersey
DJIC
NND | CLT | IAFR | BW | PREL | MRK | ENGL | USE | NJ | NME | US
DJID
Newswire End Code | **Consulting Services** | Business Services: Consulting Firms | Business Wire | Press Release Wires | Marketing | English language content | Northeast U.S. | New Jersey | North America | United States
AN
Document bwr0000020021024dyao00a73

This article however does not appear to be relevant as it appears to be more about an event measurement for marketing managers rather than corporate event management. Note the index terms are the same almost as hose for corporate events in GE.

HDStage Managed ... or how to deliver an outstanding event.
BYBy Claudia Tasker.
WC3,030 words
PD7 September 2002
SNManagement Magazine
SCMANAGM
PG34
LAEnglish
LP
With a current growth rate of 16 percent annually, New Zealand 's convention and incentives industry has never looked healthier. But the trick is to be more than a statistic and deliver a memorable event. That takes good management. Here's how to go about it. Not even the events of September 11 affected the figures and the flow of business personnel heading here to conference last year. They contributed more than $260 million to our economy.

NS
C31: **Markets/Marketing** | **C315: Conferences/Exhibitions** | CCAT: **Corporate/Industrial News**
RE
AUSNZ: Australia and New Zealand | NZ: New Zealand
RBBCM
HEADCNTY:NEW ZEALAND | INX:LOW | C::4I8395411:H::01 | C::3C31:L::03 | BIPAZ | TOTTED

AN
Document managm0020021107dy970002b

Similarly this is more about marketing event management companies rather than corporate event management.   Again the same indexing terms have been used for this article as have been used for events for marketers and the GE story.

Â(c) 2003 Dow Jones Reuters Business Interactive LLC (trading as Factiva). All rights reserved.

The indexing can be examined for consistency this way:

Relevant	Calendar of events, content types, routine general news
Relevant	Convention/Trade Shows, Advertising/Public Relations
Relevant	Marketing, Advertising, Corporate/Industrial News, Political/General News, Sports/Recreation, Content types, Routine General News
Relevant	Marketing, Conferences/Exhibitions, Corporate/Industrial News, Content types, Commentary/Opinion
Not relevant	Markets/Marketing, Conferences/Exhibitions, Corporate/Industrial News
Not relevant	Markets/Marketing, Conferences/Exhibitions, Corporate/Industrial news

From this table you can see that 3 articles had common indexing terms, but 2 of those articles were not relevant. Of the relevant articles there was very little consistency in the indexing terms.

So how useful is the Factiva intelligent indexing? In my experience when there is an indexing term that exactly matches the concept of a search – e.g. Government policies relating to the IT industry and specific countries – the Factiva indexing can be useful.

But in my view, although I believe Factiva is using both human indexing and automated indexing, Factiva has other tools that are much more useful to me given that most files in Factiva contain fulltext data.   The indexing in my view suffers from the same problem that other indexes have – i.e. inconsistent application of indexing terms and too much reliance on indexing "terms" rather than indexing the "concepts".   In my view indexers need to ask the question "What is this article really about?" and then aim to index that concept consistently, rather than looking for potentially useful terms – such as the South African article above which attracted terms such as "Political/General News".

It would be unwise to leave this section without commenting on Derwent. Patents are notoriously difficult to search and those who are not experienced may make the mistake for example of searching for a descriptor term such as REFILLABLE/DE AND PACKAGING/DE assuming those terms would be fairly safe to use. Wrong! The correct way would be to search for REFILL/DE AND (PACK OR PACKAGE)/DE   Also experienced searchers would aim to use codes as well as approved thesaurus terms.

Experiences using weighting tools and proximity operators and qualifiers versus indexes

While I may give Factiva 5/10 for indexing, I give it 10/10 for other search aids. In full text files one of the indexing challenges is to represent the significance of certain terms and concepts. In a bibliographical file, one would not normally use as an indexing term BHP for an article that describes a traffic accident that occurred outside BHP House. But when searching fulltext files simply inserting the term BHP will retrieve that article.   So the challenge in searching fulltext files is to locate articles that are focussed on say BHP and not those articles that mention BHP in passing.

Factiva has retained most of the power of the DJI search capability by permitting searchers to qualify search terms to particular fields – e.g. Headline (title), Leading paragraph, Indexing fields and to use proximity operators EVENT ADJ MANAGEMENT or MANAGEMENT

NEAR2 EVENT$1. The ADJ operator can be used with numbers 1-10, NEAR can be stretched to 500 characters, WITHIN (W/N) can be used for within the same sentence, and SAME can be used to mean within the same paragraph. But more than that, Factiva allows ATLEAST (1-50) which I find extremely useful. So I was able to set up a search string of:

(CONFERENCE$1 OR EVENT$1 OR MEETING$1) NEAR4 MANAGEMENT NEAR4 (CORPORATE OR BUSINESS OR COMPANY OR COMPANIES) AND (ATLEAST4 EVENT OR ATLEAST4 CONFERENCE$1 OR ATLEAST4 MEETING$1)

Here I am looking for articles that deal with these concepts:

CONFERENCE OR EVENT OR MEETINGS MANAGEMENT in the CORPORATE SECTOR or in a COMPANY or in COMPANIES or in the BUSINESS sector and the articles should focus much more on events or conferences or meetings that on the other search terms.

Factiva also has brilliant mapping technology that enables the mapping of index terms taken from some files (e.g. Asia Pulse) to the terms used by Factiva.

Dialog also has some brilliant mapping technology that is especially useful when searching for information on drugs. One can map all terms relating to a chemical registry number for example and retrieve articles that refer to those chemicals as trade names or as chemical names or synonyms. The registry numbers are unique numbers and so clearly this powerful technology is enormously useful for high precision and high recall retrieval of anything to do with specific drugs.

Internet searching - Google, clustering search engines, the Invisible Web, metasearch engines

Having been an online searcher for so long, and having felt totally comfortable setting up quite complex search strategies in several systems, I experienced certain dismay, perhaps shock, when in 1996 it became necessary to search the Internet. It seemed that anarchy and chaos surrounded me. No longer could I control the frequency of search terms. No longer could I qualify search terms to specific index fields.

But now just a short 7 years on, it is fascinating to see the evolution of the search engines. Of course like everyone else I use Google but I still use phrase searching, I still use Boolean operators and field qualifiers in my searches using the Advanced screen and I find it intriguing to be able to limit searches for instance to PDF or PPT files in the EDU domains. How fabulous to be able to pop into a search box: BIOAVAILABILITY PHARMACOKINETICS FILETYPE: PPT SITE: EDU and find various explanations from US pharmacy educator in Powerpoint slides or to replace the SITE: EDU with SITE: AC.UK and just as easily see UK explanations.

But clearly there are exciting new developments on the horizon with the new search engines TEOMA and VIVISIMO. TEOMA has the fascinating features of providing suggested links and suggestions for narrowing the search. VIVISIMO on the other hand appears to have the full range of Boolean logic including nesting and it magically clusters the results into logical and conceptual groups.

It is a challenge however, to find the wealth of material now buried in the Invisible Web. In their *Power Searching with the Pros* workshops, Mary Ellen Bates and Chris Sherman advise to "Adopt a hunter mindset" and to be "opportunistic" to find useful material in the Invisible Web. It is no longer possible to use search engine boxes to find the wealth of material now available via the Web but not indexed by the search engine spiders. And yet we cannot afford to ignore this rich store of information estimated to be 50 times the size of the Visible Web excluding the proprietary database services now delivered on a web platform. All sorts of techniques have to be used – find special library maintained portals, datamine bookmarks, run mini searches to find sites that may have databases buried in them, use Invisible Web pathfinders and so on.

Metasearch engines in my view appear to be more powerful than they often are. Often they do not have their own indexes. But they are useful for quick overview type searches and to find appropriate terminology.

Conclusion

My career as an information professional has now spanned over 40 years. I became *a passionate searcher* when I discovered research librarianship in special libraries 40 years ago. I saw my first online searching experiment in the winter of 1971 - over 30 years ago and obtained my first Dialog and Orbit passwords in 1975,   and I began producing and indexing a database in 1972.

Indexing, if done well, continues even today to be extremely valuable. But if too slow, too finicky, or too elaborate, in my view it is useless.

My challenge to indexers is to remember your end-users and what they need - they need you:

- to alert them **quickly**   - this means indexing must not take too long - quality and speed must be balanced
- to identify the **key concepts** in the item you are indexing - this means you must understand and evaluate the content you are indexing and weight your indexing in some way
- to reveal **how recent or how** old the material is
- to suggest the **variety of terms** that can be used to describe concepts - this means be a lateral thinker, think outside the square when selecting index terms
- to indicate the **type of data -** this means use your expertise to reveal if the article you are indexing is an opinion piece, rich in data such as statistics, its geographic scope (if relevant) whether it is dense or novel and so on.

If, as indexers you can achieve those goals, to this day despite Google and the other clever clustering search engines, the metasearch engines and the mapping, ranking and weighting of some services, and the intelligent, automated indexing, despite any new innovations over the horizon - to this day in a fulltext era the best, fast, consistent, human indexers are real treasures!

September 2003

Information Edge

www.infoedge.com.au

---

[1] It is now called EDGE, belongs to Information Edge and is available online via Informit and the MEDGE subset is available on Business Australia on Disc.
[2] 200 working days a year @7 hours per day assuming 100% productivity per day
[3] Standard Industrial Classification

Dates for your
Diary...

EVENT	DATE/TIME	LOCATION
Annual General Meeting: Victorian Branch	Tuesday, 17 February 2004 – 6.30pm	Graduate House, 224 Leicester Street, Carlton
Annual General Meeting: New South Wales Branch	Wednesday, 3 March 2004 – 6.30 pm	New Hong Kong Restaurant, 44 Macleay Street, Potts Point, NSW 2011 (tel: 9356 8481)
Book Indexing Courses in ACT	Monday and Tuesday 22 and 23 March, 2004	Australian Archives, Parkes, ACT
Annual General Meeting: National & ACT Branch	Tuesday, 23 March 2004 – 7.00pm	The Brassey of Canberra, Belmore Gardens, Barton
Society of Indexers Conference	3–4 April 2004	Chester, Britain

'Impressions of a Conference'

Fiona Mottram

In September 2003 the New South Wales Branch of AusSI hosted our international conference 'Indexing the World of Information'. The Conference covered a varied program which attracted many allied information professionals, and was a landmark for one of our Society's members, Fiona Mottram, who was awarded the first Conference Sponsorship provided by the National Committee. This Sponsorship is designed to give people from remote areas in Australia who have never been able to attend an AusSI conference the opportunity to participate, by paying some of the costs of conference attendance. Our first recipient is an active member of the Victorian Branch (currently Branch Secretary). A condition of Sponsorship is a report on the conference, and Fiona has provided this personal as well as professional view from a "newbie's" perspective. For those who were there, it may bring back memories. For others, we hope it acts as an introduction and overview of an interesting Conference (note: over the next few months the Newsletter will be printing some of the formal papers of the Conference).

Friday 12th September

Registration & satchel

Looks like most people don't know anyone else either. Good, I'm not at a disadvantage & it should be easy to say 'hello' & meet others. The nametags have big print so you don't have to peer at people's chests. Well you don't anyway; they hang on a string that dangles them at navel height. When you sit down, the nametag disappears into one's stomach fat roll if you happen to possess one!

Now that's clever marketing – all these women = chocolate eaters...and inside the conference satchel there is a 'caramello koala' with an advertising sticker for Glenda Browne & Jon Jermey's business on the back. They are assuming that we'll not rip the wrapper off so fast that we won't see it.

Yep – good contents in the satchel. Apart from the choccy there is a usefully sized notebook, a pen, a highlighter and tourist maps. Not too much advertising. Well done. Better write my name on everything as we all have identical paraphernalia now.

The conference rooms look really nice, toilets are handy (insufficient numbers for females as usual) and there is a cute balcony with seating too. I approve.

Introductory address & 'housekeeping' Caroline Colton

'If humans use the system, then humans will always be required to develop, organise and maintain the system'.

OK – so there will always be the opportunity for indexing work, despite technical changes.

The conference will hopefully 'spark imagination and resolve' for those attending.

Gee, I hope so too!

An anonymous donor has provided for the cocktail party.

Hmm…either they don' t think its good business advertising to be associated with alcohol, or else they are simply generous and the conference funds don' t stretch that far. I'll be generous and assume the latter!

Keynote address by Elizabeth Swan

With some humour, Elizabeth took us on a journey of 40+ years where she progressed from falling off wooden library ladders to being a business manager of an indexing/information/research service.

As a business manager, **ES** kept referring to the need to be financially viable while doing good work and keeping the needs of the end user in mind. Use intelligence to modify methods & terms to fit with these requirements. Dating information is important to researchers and is often not a field included.

People are required to evaluate information and to make relevant information 'visible' and findable to users.

Morning tea

Yummy & plentiful and typical of more to come. My only complaint is that drinking chocolate isn' t served. So I whip up to my well-appointed room in the hotel (more posh than anything I' ve ever stayed in before) & collect my own jar of drinking choc to use every break. Queuing with all those other women at the conference in a small space while we wait for a chance at the toilet provides opportunity for conversation and swapping details. Maybe I should have had my business cards laminated? The few men at the conference must miss out on this fun!

International panel

Jill Halliday, UK *(Plummy voice rather like a younger version of the Queen).*

Indexing as a profession got a public boost in the UK when a female cabinet minister who'd had an affair with then Prime Minister John Major complained that she'd been missed out of the index in a biography about him. *(Moral: Sex gets people' s attention anywhere.)*

Currently about 900 members. Members are after work and education. The society's web site is proving to be a very important means of contact and information. The society is doing a lot of work marketing indexing education to members and also in targeting professional conferences where specialist needs would benefit from an index. More communication is also happening by email where an active 'listserve' group discusses indexing issues. However you have to be a UK society member to participate. *(Pity about the exchange rate, eh).*

Noeline Bridge, Canada

In Canada they have an Indexing and Abstracting association. Only 135 members – but this is up from the average of 90. The association does joint conferences with both the editor's society, and with the USA indexing society.

Geographical spread of indexers in Canada is a bit lonely & isolating. Executive meetings are held by 'e-meetings'. A rather quiet listserve is also available for answering members' queries.

I am surprised how small this society is.

Frances Lennie, USA

Some media coverage describing 'easy jobs to do from home with good income' included indexing as a profession. This resulted in a somewhat chaotic influx into the American Society of Indexers of people who believed this! *(Loud, extended laughter from the audience.)* Membership went from 500 to 1300 members, most of whom had unrealistic expectations. Membership has now settled at around 750.

There has been some internal shake-up going on. Some believe that society efforts have favoured the interests of freelance members over other types of members. The society is now trying to get more organisations as members, rather than just individuals. For members, the society has an all-important web site. There is a 'jobs hotline' on this. Marketing the need for indexers is aimed at corporate organisations. Three-day conferences are held, with the third day devoted to a fun excursion. *(This sounds like a great idea. I mean – it would be a good opportunity for networking.)*

Indexing for beginners is a course run by the Dept of Agriculture, and only 3% of students enrolled in it actually

complete the course! (Historically the course was set up to help farmers' wives find an alternative employment.) There is a big gap between this course and the advanced indexing course run by the ASI.

Lynn Farkas, Australia *(Amusingly, she sounds more North American than either Noeline or Frances)*

Two major concerns for AusSI:

1. Training for indexers at all levels depends on only a handful of people.

2. Membership has been declining steadily for some years. Currently now about 173 members down from a previous steady average of 190 members. People joining & dropping out again.

However Lynn believes there is a great feeling of community in the Australian society. People who have been out of the indexing profession are coming back into it again. She also notes that 'Indexers Available' is used extremely heavily.

Lynn is enthusiastic and the audience is quiet & attentive. (Well, with all those librarian-types in the group, what do you expect?)

I now feel inspired to write an article for our small – town country newspaper about being sponsored to the conference, and what indexers do etc. Yes, we can all do our bit for promoting indexing and the society.

General meeting for AusSI members

First thing to note is that this is **not** an AGM – just a meeting to take advantage of so many members being together in one place. Non-members were welcome to sit in & listen. Only 1 Queenslander was present – a new member who doesn't know any others from her state.

Office bearers at the National level were asked to stand up so people could see them.

It would have been much more useful if they' d stood up as their name was called so we knew who was who.

As the sponsored attendee, I had to stand up & be gawped at too.

Ahhh – I hadn' t realised that this was the first year the sponsorship was offered. Later on, I find that it was a very useful exercise to have been pointed out to the crowd. It breaks the ice when introducing myself, gives an instant topic

for conversation and everyone now seems to remember my name.

Lynn gave a progress report from the Outlook workshop. She is very keen to find ways for the society to promote itself and the profession publicly. Help is needed with improving the web site usability – and the web site's importance part of publicity is recognised.

Fewer people are applying for registration. Registration is no longer an entry point, but is rather peer acknowledgement of experience and expertise. However gaining registration may now be too slow due to the time taken in getting indexes published.

The 'Indexer' magazine will cost us more due to the exchange rate changes.

People were encouraged to ask questions and give ideas.

(It was a bit like an auction – if you scratched your head you were thought to be asking a question.)

State reports were given. Max McMaster from Victoria described Vic activities and the mentoring pilot project.

Caroline Colton for NSW spoke about the conference organisation, and about work being undertaken to determine the hourly rate recommended for indexers. Geraldine Triffit on behalf of ACT described their system of meetings. She also talked about interacting with other information professional societies.

Food seems to be a common denominator in attracting people to events!

It would have been beneficial to have a representative from each sub-committee and state group providing a report to the meeting, even if read out on someone's behalf.

Parallel sessions
Indexing names

I decided not to go to this as I'm still at the stage where I dive into a book to see how trickier ones are done, and thought that I'd rather learn something about database indexing. However the people who attended the Indexing names talk and discussion came out laughing and cheery.

Using metadata: a magic bullet or death by degrees by Nelly Ivanova

This talk seemed mostly relevant to people in large organisations with a need to decide how and at what cost, they index material for researchers to retrieve. It also dealt with sharing indexes and information between many users for an added benefit at a lower cost. Issues involve trust, feedback, quality, software, duplication and controlled vocabularies.

Personally, I found this after lunch session tedious. The speaker used a rapid, monotone delivery plus lots of acronyms and jargon I didn't know. Lack of humour combined with a stuffy room and a full stomach, meant I struggled to find any meaning or relevance. I even got dizzy watching the cursor swirl around on the big screen! However I did get (I think) definitions for the following terms:

> Identifier – location on & offline of the resource.

> Interoperability – sharing between systems.

> Metadata – Structured data used to describe resources/documents and the source itself. Includes title, subject/s, description and universal resource identifier.

> Portals – different web sites leading to information.

I'm not totally sure about 'harvesting data' and 'reverse harvesting', but I believe it was to do with finding and removing or else finding and adding, information from/to the database. Someone did later explain this to me over dinner but I admit to being distracted by food and noise in the restaurant at the time.

I did like one quote from the speaker… '*Create once, use many times*'.

Whole-of-government thesauri & high level subject thesauri: trends & issues by Sherrey Quinn

This talk was about government initiatives in delivering information to clients via the Internet given the increasing importance of the web 'as a market place'. In fact provision of the internet at public libraries has had a big impact on people being able to access government via the web. Huge volumes of information have become available but in a chaotic manner. Consequently, there has been 'a re-discovery of interest in controlled vocabulary' in order to make government web sites more useful and to develop standards for resource description.

Sherrey spoke about the following aspects:

1. How to get the citizen's language viewpoint without oversimplification. And don't forget lots of people have trouble spelling even the terms they use themselves! In addition, people often do not understand the hierarchies within government departments.

2. People expect an instant gratification in finding information.

3. The Thesaurus of Australian Government Subjects' (TAGS) and the Australian Governments Interactive Function …? (AGIFT). *It was interesting that in Canada, an officially bilingual country, every subject thesaurus naturally had to be in both French and English.*

4. Provision of controlled vocabularies for agencies lacking a specialised thesaurus.

5. The politics in choosing terms and not offending different sectors of the community. For example people 'refugees', 'asylum seekers' or 'illegal immigrants'? However even apparently innocent terms such as those relating to rabbits, can be a problem. To some they are a feral nuisance, while to others they are a booming new meat industry!

6. Different ways of directing people to information – using authority files, thesauri, taxonomy, classification schemes (e.g. browse trees), synonym rings and antology. (Antology as in antonyms).

7. 'On site' tours and training on how to use government web sites.

My librarian's heart sang when Sherrey referred on several occasions to the Library of Congress Subject Headings. Familiar territory, though I hadn't realised LCSH had been going since 1895.

Panel: Indexing software
Cindex – Frances Lennie

A software that literally looks like index cards on the screen. Costs US$525. There is a listserve for users. (I think!)

After Glenda Browne introduced Frances, she then accidentally took off with Frances' notes, to the audience's delight.

SKY Index – Michael Wyatt

Michael demonstrated 'a few of his

favourite features' such as a split screen look (similar to spreadsheets), ability to edit 'on the fly'. There is a mailing list for users and it costs much the same as Cindex does.

Our laugh on Michael occurred when the timer indicating the end of his session buzzed before he'd even got onto the speakers' rostrum!

Macrex – Max McMaster

Max is an excellent salesman. He did not put up visuals on the screen (I think because Macrex's MS-DOS look makes it appear old-fashioned) but instead talked about the features of the latest version 7.18B. It is compatible with all windows platforms and can be used in HTML format for web site indexing. Costs AUS$830.

We all laughed heaps when Max described it as 'the package to use if you want to win the Indexer's medal'!

Max then adjourned into the foyer where he went into more detail about Macrex. I own and use Macrex version 6.

Product reviews by exhibitors

I bombed out on these. I'd had enough for the day.

Cocktail party

This was rather fun and good for meeting people and relaxing before dinner. I enjoyed sitting out on the balcony chatting during this time. I'm not good at balancing drink & food politely while standing up in a press of people. I left before the end in order to test out the hotel's swimming pool. The pool is to be highly recommended!!!!! Set in a garden on the 8th floor, it has views of city lights. I lazed in the warm water rejuvenating and watching bats and aircraft pass overhead. Please can there be another conference here!

Dinner

This was a great lark and I got to spend time with people I hadn't yet got to talk to, plus some I'd already got to know a bit. It involved the best Chinese food I've ever eaten, and did not resemble the 'standard' Chinese food I am used to. There was a fun trivia quiz and I was in the team that came last! Honestly, the rubbish some people fill their heads with!

I was surprised that only 33 or so people attended the dinner. If you can possibly afford it, I highly recommend it as a way of relating to other people at the conference! Thank you very much to the Vic Branch, which sponsored my attendance at the dinner. I would not have made it without their support.

The only downside was that the rich food affected my digestive system in a negative away, and it took me 3 days to get back to normal. For those who wondered at my occasional dashes out of the conference rooms on the Saturday, you now know why!!!!

Saturday 13th September

Indexing the future of information: Glenda Browne

What can I say? Well, like quite a few other people, I missed noticing that Saturday's first session started half an hour earlier than Friday's did. So unfortunately I didn't attend Glenda's talk 'cos I was still mooching about in my hotel room. People coming out were very pleased with Glenda's contribution, and buzzed with conversation. My apologies Glenda!

Parallel sessions
Health thesauri

I chose not to attend these talks as I had already heard Mary Russell's excellent talk on the women's health thesaurus at an activity run by AusSI Vic Branch. Consequently I went to hear the other topics.

Developing an indexing product: 'Wine Diva' by Caroline Colton

I found Caroline's talk extremely interesting. In many ways, it was a description of one person's journey to take control of her employment and economic future. She displayed adventure, independence and courage in developing an internet directory designed to encompass the entire Australian wine industry. She structured the directory to reflect the structure of the wine industry, which is based on regions. She saw success would only be achieved if the directory could pass a comprehensiveness test, as this is how she perceived other directories have failed. Consequently basic listing is free, so everyone is encouraged to list. Additional advertising or descriptions are paid for. She also gets a percentage of sales of books via the site and can sell data such as mailing lists. Not to mention a lot of free wine samples and plenty of travel fun!

Caroline was closely supported and assisted in her work by friends and relatives with skills she didn't have herself. She also did a thorough business analysis. While I may not personally want to do what Caroline has done, I think she is a good example to follow in terms of the manner in which she has operated. It can be applied in whatever direction one wishes to progress. However Wine Diva turns out, Caroline can count successes, though I really hope that she receives the economic benefits she is striving for.

Have a look at <http://www.winediva. com.au>

Ensuring an intranet meets staff needs by James Robertson

Intranets seem to be a microcosm of the internet. Heaps of useful information may be there if only one can find it! Need for controlled vocabulary, trustworthy quality of content and relevance to the users in relation to organisational goals are all issues. James strongly believes in finding out what the workers at the coalface need from an intranet, rather than what management thinks they ought to need. He uses structured investigation methods – focus groups, workplace observation, process & knowledge mapping, review of existing situation, interviews and use of search engine logs – to determine what those needs are. (You don't just ask people what they want!) When the intranet meets the needs of the users, they are its greatest advocates.

Look at his web site <http://www. steptwo. com.au> for articles about these topics. James also recommends doing internet searches about 'information architecture' and 'knowledge management'.

James managed to make it all interesting and easy to follow, with humour thrown in.

Indexing for online users 2: User centred information architecture and menu systems

I decided I'd had enough of online information and wanted to get a better look at ideas relating to back of book indexing. As a result I did not attend this talk or those 2 following it by Derek Jardine & Bruce Mcleod.

Two steps forwards, one step back...indexing...biographies of Elvis Presley by Susan MacDougall

You just had to laugh through this talk as Susan cleverly divided her topic

under subheadings, many of which were titles of Elvis songs, for example 'Suspicious minds', or what I think are lyric lines from Elvis songs. She has an unabashed passion for Elvis and combined it with her indexing passion to make a study of indexing of biographies over a 30-year period. Naturally she easily justified her choice to use Elvis biographies for comparing indexes – Elvis is someone who has had more than 1000 biographies written about him. Susan also has access to friends with libraries of his biographies, in addition to her own collection, which made research easier. She even drew a parallel between Presley books and the Bible… 'they are now the subject of more serious study and you keep returning to re-read them'!

Susan found that over time, indexes have improved. This is partly because the time lapse allows more serious, scholarly biographies to be written and these are much more likely to have a quality index. However the lighter style of book is less likely to have an index, and any having indexes were usually shorter than the ideal for the book's size. All the same, she concludes that electronic advances has not diminished indexing, but that indexing societies still need to promote the importance of quality indexes.

Susan also provided a very clear handout of her talk.

Classified and specific indexing for legal publications by Lorraine Doyle

Legal indexing has been a system used over several centuries. In court, a lawyer may need to look up a counter argument on the spot if an opponent introduces a new argument during the case. A quality index is required to meet that urgent, on-the-spot need. Further, law books are not books to be read, they are reference books, so are used primarily via the index.

Traditionally, legal material has been indexed under a fairly cumbersome classified system. There are few abstract concepts included. And it makes problems with cross-referencing. Legal indexing is specialised because legal terms have special meanings that must be fully understood. In Victoria, there is a 'Table of classified index headings' to use. Committee meetings are required before categories can be changed, as there is a big history in indexing legal documents – in Victoria since the 1860s. Additionally, much

indexing is done for loose-leaf services, which have supplementary indexes until they are consolidated.

Currently, there is pressure to move to a combined style of index with legal books. Modern lawyers rely on computers and CDs, so do not commit as much to memory as has been traditional. Students are trained more to direct entries. Also, the government is committed to using 'plain English'. However law is a field where debate and consistency are strong, so change is slow!

Lorraine recommends the UK indexing Society's booklet for people interested in indexing legal books. She suggests that people can start indexing legal journals to get a feel for the area with a lower level of exactness required. Legal indexing is very labour intensive and there is much less publishing pressure to limit the size of the index. 'You can index to your heart's content'.

Lorraine's talk was very interesting. There were obviously quite a few people in the audience who worked for legal publishers.

Unfortunate omission: editing and indexing women out of the Anzac legend by Caroline Viera Jones

Caroline's talk really opened my eyes to the manner in which influential people can sanitise, censor and bias the reporting of history. And if the information isn't in the text, then it can't be indexed! The official history of WWI could have told in detail what it was really like on the front, in hospital lines and back home. However much of this was sacrificed due to political pressures, nationalistic beliefs & friendship between publisher and prime minister, which dictated that the work should be on a grand scale and should polish the Australian male hero image. Our Anzac legend, indeed. Hence references to 'Australian stragglers', 'clashes between surgeons and matrons on hospital ships', 'domestic history' 'mateship between nurses and soldiers' and 'female heroism' were edited out. Whole chapters about women's involvement were removed, and women were excluded from the official indexes. Three volumes about medical history during the war were published in such small numbers that copies were rare. Yet the main volumes were common in people's homes.

Caroline's talk was rivetting and I enjoyed it immensely.

Education & indexing talks by Anna Gifford & Steven Haby

Again at a Vic Branch activity, I have heard Anna's clear & informative talk, so I opted not to attend these talks.

Principles and practice 2: Indexing by the book

Alan Walker & Max McMaster have a practiced, easy, ad lib comedy act *(was that a question or were you just waving?)* that simultaneously manages to teach plenty and invite discussion.

Alan began by describing life as an indexing student in the 1960s, but then progressed to discuss indexing abstract ideas, things that have a phrase rather than a name, 'how specific is specific', see also & see references and natural language use. He wanted people to consider whether practices are traditional and generally accepted, or are they rules? In a particular situation, what is best for the user?

Max talked about indexing the Queensland Tertiary Assessment Courses (QTAC) Guide in a manner completely different to his previous versions. He changed from specific entry to classified entries. He also did 3 separate indexes. (An alphabetico-specific index, a classified index, and a general index for topics such as accommodation, application etc) This was on the basis that users were year 12 students who were not familiar with academic courses and found specific entry to be unfriendly. The impact of an 'unfriendly' course guide could actually affect enrolments as students couldn't find the course they were looking for!

The general discussion then turned to questions such as 'would you include the Pope in an index about soccer*? One audience member said ' Yes I'd include it because I'm a Catholic and I don't know what might happen if I didn't!' She then blushed heavily.*

The consensus was index it, and if it becomes merely a passing mention, then delete it. But retain it if it becomes a major issue.

Are we making indexing rules that readers don't understand? Do the readers care?

Max & Alan have opposing views over the use of the ampersand (&) in indexes.

Indexing about people who rise in rank over the course of the book: Index

under the highest rank attained if you are including ranks. However ranks are not generally indexed.

Plenary: Education for indexing: an international panel
Frances Lennie, USA *"I don't index dental books"*

This woman is excellent.

Frances reckons that indexing is something which can have the basic elements taught, but that 'the rest follows by cultivation'. She compares indexing to painting by numbers and works of art. You can follow the rules, but there is also the requirement for a creative element that lifts the index to excellence. A good index means that you can find what is required within 30 seconds. Software does the chores of indexing. Students need to understand these principles, but no longer need to memorise them. Students need to build up their confidence in analysing what to select for the index, and must learn when to edit entries to improve the index's quality.

Qualities of a good indexer are: creative, a writer rather than a reader, active, organised, well-read, work disciplined, good at comprehension and précis work.

There is no right way to index a book. Skills used to index one type of format can be applied to other types. Differences are relatively small. Indexers should keep on track professionally: analyse their own work, go to conferences & society activities, participate in listserve discussions and peer reviews.

Realities of indexing can include: being used as a free consultant, changing deadlines, training clients and cost cutting!

Jill Halliday, UK *"I'm just showing you how this works, I'm not trying to teach you how to index"*

Jill demonstrated the training software now sold by the Indexing Society in the UK. There are a series of self-test units, followed by tests that are marked by examiners and that lead to accreditation. The training course has lots of highlighted anecdotes, examples and hints. It is well set out. Help is available by phone and email, plus there are 1-day workshops for students and recent graduates. This training system (available by CD-Rom & mail) illustrates a big commitment to training

and support of indexing students. This system cost 15,000 pounds and will take some time for cost recovery.

Lynn Farkas, Australia
Australia only has limited face-to-face training available. Distance education in indexing is only available from overseas and is high cost. AusSI has not the resources for interactive CD-Rom & web-based education. However it may be able to tap into government resources for vocational, skills-based training. Partnerships with organisations such as museums may be a possibility. A workbook option is a possibility. (idea from the National Library's Kinetica workbook for library staff around Australia doing inter-library loans). Perhaps there should be hands-on workshops at conferences?

Self-education can be assisted by using the web to get started in looking at related information-profession fields.

If people have ideas, can they please pass them onto the education committee!

By the end of the conference, people simply looked too tired to respond much to Lynn's plea for ideas. Lynn must have been tired too as she got tongue-tied. Instead of saying 'user pays basis', she came out 3 times in a row with 'user bays pasis'.

And if you want a chance at winning the door prize, you have to stay until the very end to be in the draw!

Overall impressions arising from the conference that might help others.

1. *Indexers come in all age ranges, so it's obviously **never** too late to start!*

2. *As one of the relatively younger attendees (and I'm 42!), I could be in a position to get a lot more work in future, as at least half of those at the conference will theoretically be in semi-retirement or retirement within the next 10 years. Sorry if I have offended anyone here.*

3. *Be confident and keep at it. Don't let self-doubts hinder you. Plug away at getting opportunities for work. They will happen in time. Don't be intimidated by all those indexers with more experience – there are plenty around with even less than you have too.*

4. *Get the experience now if you can – in the future, there will be*

opportunities for teaching indexing work. Already there are too few people teaching indexing skills.

5. *There will always be a need for indexers as the human element in making information links. Be prepared to be flexible and adventurous in following work opportunities.*

6. *Indexers come from a variety of backgrounds. Don't feel left out if you haven't yet met anyone with a similar background. Besides, you can always turn it to an advantage if you are unique!*

7. *Feel good about a your passion for patterns, order, meticulousness, hierarchy and pedantry. You are neither freaky nor alone nor uncreative. But never mind if such passions don't translate into areas such as your wardrobe organisation or housekeeping skills.*

8. *If you really don't understand database indexing stuff, you certainly are not the only one. However it does seem mostly like it's the same old ideas parading under different jargon terms, so you could probably do it ok anyway. If it really leaves you cold don't panic. People seem to have enough work to keep them going without being a database whiz.*

9. *Given our population size, Australia seems to be doing pretty well for indexers.*

10. *Going to a conference does heaps for your indexing passion. It makes you itch to get back home and get stuck into work with some new perspectives and enthusiasms It also gives you some contacts in what is a relatively isolated profession.*

11. *If you have something interesting you can give a paper on at a conference, give it a go. Many of the speakers were nervous beforehand and apparently felt lacking; yet their performances were terrific! If you are a bit behind in presentation technology, there is obviously a club you can join.*

12. *Good speakers tie in comments heard from previous speakers with their own talk. This makes links and adds relevance for the audience.*

13. *If you could be eligible for the sponsorship next conference, jump in and try for it. The conference is valuable, fun and confirming.*

14. If you want to ask a question at a conference, it is standard practice to get up from your seat while you ask.

15. Writing a post-conference report, even if only for yourself, is very useful for establishing what you learnt into your head!

16. Be glad you are Victorian. Vic Branch runs excellent activities, which was reflected in the papers given at the conference – Mary Russell, Anna Gifford & Max McMaster's contributions. If you can't get to conferences, then make it a priority to attend Vic Branch activities. Cumulatively over the year, it's a bit like a conference and has the same high standard.

17. The conference has given me a sense of direction and a feeling that although I travelled to it alone, I am going home in company.

Comments about the conference organisation

1. There was no table devoted to promoting AusSI. Quite a few conference attendees were not AusSI members, yet they were not wooed to join us. The registration table has a confusion of items on it, so could not be said to be advertising AusSI. I did not see any membership forms being flaunted. We should have had a stand that hit people in the eye when they arrived. It was AusSI's conference and we talk about promoting the Society, but where was the action at a prime opportunity???? I would have liked to be able to direct non-members to a promotional stand as in conversation I encouraged some to join. More about AusSi needed in the conference satchels too. I was shocked at this lack of publicity.

2. Most people I spoke to wanted more talks relating to back of book indexing.

3. I would be great to have had a behind-the-scenes tour of somewhere close & related, such as the Powerhouse museum.

4. That idea of a workshop/s at the conference sounded like a good idea.

5. Basically I loved the conference and the venue. It was well organised and ran to time. I came away inspired and addicted to attending more!

From the Literature

Glenda Browne

The Quilt Index

The Quilt Index <http://www.quiltindex.org> aims to be a comprehensive index to American quilts. It is built on four state quilt documentation and digitisation projects. Nine hundred quilts can be searched by pattern name, quilter's name, location made, date, and collection or documentation project. Pattern names include 'baby blocks', 'ocean waves', and 'wedding ring'. (This information came from the Scout Report v.9 n.42 <http://www.scout.wisc.edu>.

< http://www.judywebster.gil.com.au/methods.html >

Lynn Farkas's conference handout listing websites for professional development led me to Judy Webster's site, which specialises in information on genealogy. It includes a section on 'Indexing Rules and Methods' with hints specifically for genealogical indexers.

Infotoday Newsbreaks

In 'OCLC project opens WorldCat records to Google' <http://www.infotoday.com/newsbreaks/nb031027–2. shtml>, Barbara Quint reports on an OCLC/Google project to extract a two million record subset of the most popular and widely held books on WorldCat (cooperative library catalogue) and make them available through Google. Searches on Google will retrieve the records and link through OCLC to library holdings. One issue still to be resolved is the likelihood that the library holdings will not rank highly as they are 'thin' records with few keywords to contribute to linguistic frequency rankings.

Quint also notes that Google executives are talking with investment banks about a possible float of Google. One option being considered is holding an online auction of shares. I can imagine quite a lot of 'mum and dad investors' (or should that be 'nerdy young adult investors'?) would be quite pleased to put in a bid.

Journal of China Society of Indexers v.1 n.1 2003

The table of contents of the *Journal of China Society of Indexers* lists the following topics:

Inaugural statement on the first issue
Modern index is database
Make a good job of index and become Chinese technical pioneers
Widely used and brilliantly Prospective image index
Outlook for automatic classification technique
Statistic language model used in text information retrieval
Information quality and information retrieval teaching
Index in ancient China
A decade review of the China Society of Indexers
Creation and development of index organisations throughout the world.