The Practice of Indexing

Australian and New Zealand Society of Indexers Conference

15–17 October 2009
Sydney, NSW

Table of contents

History of Indexing in Iran: past, current and future, by Dr Roya Baradar and Sepideh Fahimifar

Practical Indexing: what we say and what we do, by Glenda Browne

Social Networking for Indexers, with Glenda Browne, Jan Wright and Tracy Harwood

Creating a large author index from a list of references, by Andrea Jordan and Donald Jordan

Metatopic Menace, by Kay Schlembach

Trends in indexing — ebooks and tagging, by Jan Wright

External links

Indexing in the '60s, ff., by Hazel Bell – video on YouTube
https://www.youtube.com/watch?v=5fnyUsRXmZ4&spfreload=10

The Digital and Institutional Environment, by Mal Booth
http://frommelbin.blogspot.com.au/2009/10/miscellaneous-connections.html

History of indexing in IRAN: past, current and future

Roya Baradar[1] , PHD. At ALZAHRA university.
Sepideh Fahimifar[2] , M.A. Student. At ALZAHRA university

Abstract

Indexing has been present for over 30 years in IRAN's History. The development of indexing in Iran, simultaneous with and impacted by its growth in developed countries, occurred under the influence of such important factors as rising number of Persian thesauruses, the evolution of ICT, and the growing functionality of indexing among users, and etc.

In this paper, along with describing the history of indexing in IRAN, various methods of indexing, different kinds of indexes, and numerous numbers of Iranian thesauruses will be discussed as well. Finally, we will also have a look at IRANDOC, one of the most operational indexing centers in Iran.

The research in this paper is based on various data collection systems and sources such as library and survey research methods, and interviews with experienced professionals.

.

Keywords: **indexing, thesaurus, history of indexing, IRAN, IRANDOC.**

Introduction

With the growing information overload, the need for resource organizations has been felt more than ever. Indexing is considered one of the resource organizing methods. Modern Indexing in IRAN initiated in 1970's with establishment of various indexing document centers. IRANDOC has been well known as a pioneer in indexing in IRAN. It has published the first book about thesaurus.

Indexing has, so far, gone through significant developmental changes such as the improvement of thesauruses and the growth of indexing centers, indexing databases, and public indexes.

The major problem facing the process of indexing in Iran is, however, lack of indexing association. Hopefully, with the establishment of a national information system, including a national indexing system, as one of its main subdivisions, the existing problem could be solved. A national indexing system would ultimately lead to a national indexing association that would facilitate professional the best systematic communication among indexers.

[1] rbaradar@alzahra.ac.ir
[2] Sepidehfahimi2004@yahoo.com

The stand of indexing in Iran is the major topic of this article. In order to achieve the best result, the survey method has been used here. Data collection technique is utilized by performing interviews with a few indexing professionals. The statistical population is consisted of 11 experienced indexing professionals. The methods adopted in this article towards indexing improvement in Iran, is based on the opinions of these professional.

A Review of Indexing in IRAN in the Past

Adopting various book Classification methods in IRAN goes back to the Samaniyan period (741-969). During this period, different kinds of books were gathered in the libraries' designated rooms. Our predecessors had, certainly, thought about finding ways to guide the users to the particular sources they were seeking. Apparently, the concerns regarding guiding the users to the right source/s, such as a particular book or page/s of that book have always existed. The ancient form of using index was "figure of speech" or a kind of abstract form. Later another method was adopted to help the users, as in Beyhaghi history book, and that included a brief introduction. The second phase of the progress led into sorting the resource contents out to separated parts: as in Jamy's Haftorang book. The categorization of the book contents was, usually, appeared at the back of the books. The third phase was the content sorting into different chapters with definite subjects. Up to this time, numbering the books' pages wasn't applied; therefore designation of chapters could, at least, lead the user to a fairly accurate page with the information he was seeking. In addition to the use of chapters, the authors decided to divide the chapter to the subchapters. Techniques improved and eventually, the first book including a table of contents, similar to modern books, Hejleh alarayes, was published (Emami, 2006).

In Iran, the first unpublished book about indexing was written by Shamsodin Adib Soltany. Although the book's title was "Guides to Book Publishing", the author had discussed indexing in detail.

The first Persian books including index were, most probably, those books which were published abroad. After that, Iranian authors tried to add indexes to their books, basically based on their personal preference. Most of these indexes were the names index or location index.

Up to 1967-1970, there wasn't any sign of advancement in the principle and structure of indexing. However, gradually, indexing principles and standards were initiated. The first book included indexing rules were published around 1976 e.g. OPEC history and Shoobiye movement (Azarang, 1976)

Will Durant in his famous work, the History of Civilization, talks about a scholar who had a vast number of books in front of him, not knowing which one to choose. We, modern people, are also in the same situation and faced with the similar problem. We are daily

exposed to avalanches of information being generated every minute; we are confused and unable to decide what to adopt from. To comply with the growth of information traffic, various Iranian information centers with a goal of accumulating, categorizing and disseminating information were established. A list of these centers is as follows: Scientific Information Database as an Iranian Academic center for education, culture and research (SID), Institute for Iranian Contemporary Historical Studies (IICHS), Agricultural Scientific Information & Documentation Center (ASIDC), Islamic revolution's document center, Islamic Republic of Iran Broadcasting (IRIB), Iranian Research Institute for Scientific Information and Documentation (IRANDOC) and etc.

Among these centers, IRANDOC has been known as the foremost indexing center in IRAN (Bagheri, 2004). It has begun resource indexing since 1969 in according to documented policy. In 1972, IRANDOC published the most historic book about indexing. This book," The Rules of Thesaurus Construction" was written by Ali Sinayi. It was about the rules of keyword selection, the rules of keyword coordination and usage of those keywords. The first Persian thesaurus was published in 1977 by the Plan and Budget organization. It was the Persian version of Macro Thesaurus: a Basic List of Economic and Social Development Terms, published in 1972 in English by the Organization for Economic Cooperation and Development (OECD). Since then, few exclusive thesauruses were published by various information centers, such as Cultural Development Thesaurus, NAMA Thesaurus: Science and Technology Policy Information Exchange System, UNESCO Thesaurus, International Thesaurus of Cultural Development: Sub-Saharan Africa and etc (Bagheri, 2005). The newest thesaurus was published by IRANDOC in 2008 that is entitled "crisis management thesaurus"

In 1992, the first public index "NAMAYE" for the entire available articles was published by Cultural and Information Institute of Index, affiliated to the Institute of Public Libraries. This index has been published on a monthly journal basis, using DDC system to arrange the topics. Users could basically search titles, subjects and keywords of articles. Since 1999, it has been made available both on CDs and as printed copies (Kats, 2007). In this way, users can have access to the entire texts of articles. Recently, in addition to NAMAYE Journal Co, ILLA Software Company is also trying to index journals.

In short, with a glance at indexing history in IRAN, one could conclude that Iranian professionals had recognized the importance of indexing from centuries ago and initiated ongoing establishment of indexing centers.

The Current Stand of Indexing in Iran

Nowadays, with the rising information explosion, indexing books, articles, documents, theses and etc plays a significant role for any user. Indexing can lead users to the

information they seek, directly, accurately, and easily. Although, many books published in Iran include different indexes, Iranian publishers despite the growing need for indexing, are not required by law to include indexes in their books (Nourozi, 2001). Due to few thesauruses, indexers complain of using the natural language to index of books, whereas the use of controlled vocabulary will make easier publishing of the national index. With regard to article and document indexing, various centers have tried to guide the user to the particular information they are seeking. There are many indexing centers in Iran with the most experienced indexing professionals. We interviewed a few professional indexers in the following centers:

SID, IRANDOC, IICHS, Islamic revolution document center, IRIB, National Library and Archive of IR of Iran (NLAI).

The majority of these centers use co-ordinate indexing and thesauruses, subject heading, and specific dictionary to index.

At first we will present a brief description of these centers. Then, the stand of indexing in IRAN with an emphasis on professional opinions will be explained.

The Institute for Cultural Research and Studies has organized and cataloged a vast number of valuable historical documents collected during and after the victory of the Islamic Revolution in Iran. In December 1996, it was replaced by the [3]Institute of Iranian Contemporary Historical Studies (IICHS), a proficient research center devoted to the study of contemporary Iranian history. The process of indexing is declining in this center and the post coordinate indexing is applied instead. [4]Scientific Information Database (SID) of Academic Center for Education, Culture and Research intends to provide intensive resource for university professors, students and researchers. The span of this scientific database includes:

- Abstracts index of scientific and research papers (Persian, English)
- Abstracts index of MS And PhD. theses (Persian, English)
- Entire text science papers published in Iran.
- Index of research papers published by Iranian scientists and researchers in international journals
- Directory of Iranian researchers along with titles of their research papers
- Introducing scientific collaborations between universities and industry

This center has undertaken the protection task of resources. Accordingly, the authors' keywords are used in their databases. As a result, there isn't any need for re-indexing. [5]Islamic revolution's document center was founded as a research institute of Islamic

[3] http://iichs.org/index_en.asp?id=696&doc_cat=25
[4] http://www.sid.ir/en/about.asp
[5] http://irdc.ir/fa/about.aspx

Republic history of IRAN in order to accumulate, organize and disseminate documents related to EMAM KHOMEYNI movement. This center tried to publish their thesaurus with an effort to apply relation between keywords in their software such as BT, NT and etc. This procedure helped to satisfy the users' particular information needs. However, making such a useful thesaurus was ceased due to the lack of budget and its time consuming nature. In National Library and Archive of IR of Iran, the indexing procedure is divided into two categories. National library use not only "NAMAYE" as a indexing source but also it index human science periodicals originally, and in national Archive, all governmental and historical documents are indexed. [6]IRIB have tried to index the audio-visual materials as well.

[7] IRANDOC origin could be traced back to CENTO. Dr. Marvin Smith, then CENTO Secretariat, suggested a national documentation center to be set up in Iran. By December 1967, Dr John Harvey who was a Fulbright Scholar presented a plan for creation of Iran Scientific Documentation Center and Tehran Bibliographic Center to the Ministry of Higher Education. These two centers were officially inaugurated in October 1968. In 1981, following the dissolution of the Institution for Research and Planning in Science and Education (IRPSE), IRANDOC became an independent organization affiliated with the Ministry of Science Research, and Technology

In 1983 the Institute was disbanded and IRANDOC was brought under the supervision of the Deputy Minister Office for Research. In 1991 IRANDOC Charter was revised and approved by the Council for Development of Higher Education. The Minister of Higher Education signed the Charter. Following the approval received from Higher Education Development High Council in 1991, IRANDOC continued its activities as a research institute. IRANDOC's major activities include: research, training, and information services. IRANDOC official name was changed to Iranian Research Institute for Scientific Information and Documentation, and its status was changed to that of a research center.

This center has been known as a foremost indexing center in IRAN. IRANDOC's initiated indexing procedure in Iran in 1969. At the first, the UDC system was used for indexing materials, because of inadequacy of UDC, the indexers created new controlled glossary. Based on this glossary, almost 400 resources were indexed. In order to compile specific vocabulary regarding science and social science, at first, indexers extract the main keywords, and then they compared them with specific dictionaries. This resulted in selected descriptors. Based on the resulted feedback, a book on thesaurus in 1972 was compiled. At the time, IRANDOC even tried to compile the indexing policy. Consequently, in 1974 an indexing office was formally inaugurated in IRANDOC. In 1980, the indexing of

[6] http://www.irib.ir/English/AboutUs/index.php
[7] http://irandoc.ac.ir/english/about_us.htm

MA/S and BA/S and PHD students' theses was started. At the same time indexing of public and private organization periodicals began. The assigned tasks of this office include;

- Cataloging none-book materials and finding the solutions for quick indexing
- Accumulating and indexing Iranian students' theses in Iran or abroad.
- Accumulating and indexing periodicals published by government organizations, or institutions and etc

Natural language and controlled language vocabularies are both used for indexing of resources. In addition to the thesaurus, indexers use the exclusive dictionaries. The first index regarding theses and other none book materials were published by IRANDOC. This index includes personal and subject indexes along with retrieval numbers. Gradually, indexing of BA/S theses was eliminated. Nowadays, the main task of this office is to index theses and articles (motaref, 1999)

Indexing professionals believe that the document centers as a whole, have tried to index resources but up until now there has not been any indexing association. Some of these centers, such as IRANDOC have tried to index all the articles and theses, but due to the work pressure and few subject thesauruses, they were not able to use semantic indexing. As a result, there isn't any relation between words such as NT, RT, BT and the entire text of some articles has to be presented to the users. Unlike IRANDOC, SID provides the users with the entire text of scientific and research papers without their indexing. Instead it uses the keywords that were selected by authors. Consequently, the user probably would not have access to some of information resources, because there isn't any relation between synonyms keywords. Therefore, these centers can and should establish a permanent communication and coordination among themselves in order to solve these problems such as establishing indexing association. The Institute for Iranian Contemporary Historical Studies (IICHS) has proposed the rules for Qajar (around1792-1926) command indexing. These rules are useful for performing searches and finding documents. However, there are few centers applying them. Moreover, some similar history documents are indexed with different methods in different centers. For instance, there are totally different indexes for the same documents in centers such as: National Library of Archives of IR of IRAN, Institute for Iranian Contemporary Historical Studies and Islamic revolution's document center.

In fact, despite useful attempts made by these centers, there are still major obstacles in the way of IRAN indexing system. Khatampour as a supervisor of organizing in Islamic revolution's document center, Mirza Hoseyin as a supervisor of indexing in IICHS and Moshayedi as a head of periodical SID, all believe that, "the indexing budget that is allocated to nongovernmental information centers aren’t sufficient and the affiliated information centers is trying to save and return some of the allocated funds. Heads of the

major nongovernmental organizations lack enough knowledge of the benefits of indexing. Consequently, the process of indexing is declining in these centers ". Dehsariyi as an indexing supervisor of IRANDOC believe that," Lack of any innovation in published indexing books is the reason for not being in accordance with today's indexing progresses. In addition, lack of advanced indexing workshops is one of the major problems that indexers are faced with". Karimiyan as a head of the National Archival Arrangement and Description, Ghahremani as a head of the indexing in National library, Ghadimi, as a indexing professional in IRIB and Mirza Hoseyin believe that" the majority of the indexing professionals have attained their indexing experiences through working in the related field for long years. Therefore, after the retirement of experienced indexer professionals, novice indexers should spend much time to achieve experience.

In general, some of the similar problems stated by professional are as follows:

- Few current subject thesauruses
- Lack of indexing association
- Lack of a comprehensive national policy of indexing
- Lack of national indexing standards
- The software problems of indexing databases

Consequently, on one hand, there are progresses and advancements in indexing field in different centers, such as indexing of theses in IRANDOC, presentation of entire text of scientific and research articles in SID and etc, whereas on the other hand, there are problems that were explained in detail before. As a consequence, the solutions to these problems can lead to an appropriate plan for the future of indexing in IRAN, based on the opinions of the indexing professionals.

A Plan for Indexing Development in IRAN

In order to solve the existing problems we are faced with, in the indexing field, experienced indexers recommend establishing a National indexing center and association. They believe, "Through establishing a national indexing center, a beneficial cooperation and coordination among the indexing centers would be created. Moreover, this center would be able to compose a national policy of indexing, as well as a national indexing system. In this policy, the national indexing standards with an emphasis on international indexing standards might be described. National indexing center could establish the National indexing association and invite the indexers from different indexing centers to exchange their knowledge and experiences in order to compile a national indexing policy and standards. Accordingly, the establishment of this association can improve the systematic communication among indexers and will result in compilation of homogenous indexes. Apparently, the establishment of a national indexing database would be

subsequent to the compilation of national standards. That means that various centers conducting the same approach can organize resources such as theses, articles, and documents in accordance with each other. Also, the National indexing center can establish communications with other indexing centers abroad, in order to obtain the latest information and to exchange experiences.

In summary, these are some of the other solutions:

- Published a special journal in "indexing"
- Back of the book indexing necessity by publishers
- Establishing exclusive and specialized thesaurus committee
- Compiling exclusive subject indexes
- Obliging publishers to use indexes
- Educating indexing professionals

Conclusion

Indexing in IRAN has achieved numerous advancements. The majority of information centers in Iran, are trying to index resources. Also many proper thesauruses have been published. However, Indexing of articles, historical documents, and theses is and has been more popular than indexing books in IRAN, solely because cataloging of books is what these centers are interested in. On the other hand, when it comes to publishers, there aren't any obligatory rules for indexing books in Iran.

Despite these developments in the indexing field in Iran, there are also some problems, the most significant of which, according to experienced indexers, is lack of indexing association. In addition, other problems worthy of mentions are: few exclusive subject thesauruses, inadequacy indexing budget and finally lack of national standards. Consequently, establishing a national indexing center and subsequently, a national indexing association can, hopefully, help solving these problems.

References:

Azarang, Abdolhosein (1976), indexing styles and its accommodating with Persian Language. M.A thesis of library and information science, educational science faculty, Tehran University.

Bagheri, Masomeh(2004)." Stand of indeing and abstracting in Iranian information centers".Journal of Psychology & Education. Vol 34, No 1.p. 127-155.

Bagheri, Mansoureh(2005). Development of thesauri in Iran. Australian and New Zealand Society of Indexers International Conference

Bagheri, Mansoureh(2005). Development of thesauri in Iran. Australian and New Zealand Society of Indexers International Conference

Emami,Saber(2006)."An introduction to the evolution and transition in cataloging of old texts".FASLNAME-YE KETAB.vol16, No 4.p.117-128.

Katz, William A (2007), Introduction to reference work. translated by Hamid Mohseni (with added entries). Tehran: Ketabdar

Moataref, Mahvash (1999). Survey on information activity in IRANDOC, from establishing (1968-1998). Research report.

Mohammadi Far, Mohammad (2003).Basic Indexing. Tehran: Ministry of culture and Islamic guidance.

Nouruzi, Ali Reza (2001). Book indexing; A Guide for publishers,indexers, librarians, authors and translators. Tehran: chpar publication

Interview whit Indexer professionals were including:

Dehsariyi as an indexing supervisor of IRANDOC

Tamizi as a retired professional indexer from IRANDOC

Aghabakhshi as a retired professional indexer from IRANDOC

Karimiyan as a head of the National Archival Arrangement and Description

Sheykhrezaei as a head of audio visual document in the National Archival Arrangement and Description

Ghahremani as a head of the indexing in NLAI

Ghadimi, as a indexing professional in IRIB

Khatampour as a supervisor of organizing in Islamic revolution's document

Mirza Hoseyin as a supervisor of indexing in IICHS

Farasati as a head of IICHS and as an experienced indexers

Moshayedi as a head of periodical SID.

Ghadimi as a professional index in IRIB

History of indexing in IRAN: past, current and future

Roya Baradar, PHD. At ALZAHRA university

rbaradar@alzahra.ac.ir

Sepideh Fahimifar, M.A. Student. At ALZAHRA university

Sepidehfahimi2004@yahoo.com

Purposes

- General perspective of the stand of indexing in IRAN
- knowing indexing centers in IRAN
- Existence problems on the indexing
- And their solutions

Methodology

- survey method
- Collection data techniques: semi- structured interviews and library
- statistical population:11 experienced indexing professionals

A Review of Indexing in IRAN in the Past

History of index

- Using figure of speech like abstract
- brief introduction of book
- sorting the resource contents out to separated parts
- sorting into different chapters with definite subjects without any page number
- Using subchapters
- table of contents

First unpublished book about indexing

- was written by Shamsodin Adib Soltany
- the book’s title was "Guides to Book Publishing“
- Discussing about indexing in detail

The first Persian books including index

- At first, those books which were published abroad
- After that, Iranian authors tried to add indexes to their books, basically based on their personal preference
- Most of these indexes were name index and location index

Indexing rules

- NO sign of advancement in the principle and structure of indexing Up to 1967-1970
- The first books related with indexing rules were published around 1976

Starting of indexing in information centers

- Iranian Research Institute for Scientific Information and Documentation (IRANDOC) as the foremost indexing center in IRAN
- It has begun resource indexing since 1969 with according to policy

(cont(

- In 1972, IRANDOC published the most historical book about indexing
- the book's title was"The Rules of Thesaurus Construction"

Starting of indexing in information centers (cont)

- Scientific Information Database as a Iranian Academic center for education, culture and research (SID)
- Institute for Iranian Contemporary Historical Studies (IICHS)
- Agricultural Scientific Information & Documentation Center (ASIDC)
- Islamic revolution's document center
- Islamic Republic of Iran Broadcasting (IRIB) and etc.

Thesaurus

- The first Persian thesaurus was published in 1977 by the Planning and Budget organization
- Publishing different thesaurus
- Newest thesaurus was published by IRANDOC in 2008 that is entitled " crisis management thesaurus"

The first public index

- In 1992, the first public index "NAMAYE" for the entire available articles
- on a monthly journal basis
- Since 1999, both on CDs and as printed copies

The Current Stand of Indexing in Iran

existence indexing centers

- many indexing centers in Iran with the most experienced indexing professionals
- SID, IRANDOC, IICHS, Islamic revolution's document center, IRIB, National Library and Archive of IR of Iran (NLAI).

Brief description of these centers

- Institute of Iranian Contemporary Historical Studies (IICHS) has organized and cataloged a vast number of valuable historical documents collected during and after the victory of the Islamic Revolution in Iran
- Scientific Information Database (SID) of Academic Center for Education, Culture and Research intends to provide intensive resource for university professors, students and researchers

Brief description of these centers (cont)

- Islamic revolution's document center was founded as a research institute of Islamic Republic history of IRAN in order to accumulate, organize and disseminate documents related to EMAM KHOMEYNI movement

Brief description of these centers (cont)

- In National Library and Archive of IR of Iran, the indexing procedure is divided into two categories.
- **National library use not only NAMAYE as a indexing tools but also it index human science periodicals originally**, and in national Archive, all government and history document are indexed.
- Finally, **IRIB** have tried to index the audio-visional materials as well.

indexing in IRANDOC

- IRANDOC's initiated indexing procedure in Iran in 1969
- At the first, the UDC system was used for indexing materials
- New controlled vocabulary
- Compile the documented indexing policy
- In 1980, the indexing of MA/S and BA/S and PHD students' theses was started.

indexing in IRANDOC

- In 1974 an indexing office was formally inaugurated in IRANDOC
- The assigned tasks of this office include;
  - Cataloging none-book materials and finding the solutions for quick indexing
  - Accumulating and indexing Iranian students' theses in Iran or abroad.
  - Accumulating and indexing periodicals published by government organizations, or institutions and etc

indexing in IRANDOC

- The first index regarding theses and other none book materials were published by IRANDOC

- Nowadays, the main task of this office is to index theses and articles

Indexing in these centers

- Indexing of articles, historical documents, thesis
- Use of the coordinate indexing
- Use of the thesauruses, subject headings and subject dictionaries

The most important problem of indexing

(attitude the indexer professionals)

- Lack of the indexing association
- Lack of the back of the book index in most books
- Indexing of similar material in different centers
- Lack of a comprehensive national policy of indexing
- Lack of national indexing standards
- The software problems of indexing databases

A Plan for Indexing Development in IRAN

(attitude indexer professionals)

solutions

- establishing a national indexing center
- Composing a national policy of indexing
- establishing the National indexing association in order to exchange indexing knowledge and experiences
- Compiling national indexing standards
- the establishment of a national indexing database

Solutions (con)

- Published a special journal in "indexing“
- Back of the book indexing necessity by publishers
- Establishing exclusive and specialized vocabulary committee
- Compiling exclusive subject indexes
- Obliging publishers to use indexes
- Educating indexing professionals

- Roya Baradar ,Department of library and Information Science , Alzahra University, Iran. Roya Baradar has been obtained a BA in History of comparative culture from Allameh Tabatabaei University, Tehran, Iran. MA in Library and information Science , from the University of Tehran, Tehran, Iran, and PhD in Library and Information Science, from Delhi University, Delhi, India. She is assistant Professor in faculty of Education and psychology of Alzahra University, and head of Department of Library and Information Science.

What we say…and what we do (rules that are not rules)

Glenda Browne

www.webindexing.biz

The rule about…

- The ‘rule’ about not indexing chapter headings has led to silly indexes
- The ‘rule’ about not writing the index in the book has led to weaker indexes
- The ‘rule’ about using specific not classified **entries** should probably be reworded to say ‘Create specific not classified **indexes**’

Rule 1: Don’t index chapter headings

- Why ever not?

OK, don’t index them all as is…

Not on the Mayflower? Then Leave!

So You Want to Kill the President!

Steve Forbes Was an Alien

A Sperm’s Right to Life

Free Us, Nelson Mandela!

Mike’s Milita

From the Michael Moore FAQ

…but don’t ignore their topic

- Surely no-one would do that!

Oh dear…

- Archives text, 1987
- Introduction to the index 3) Because the CONTENTS present the main topical divisions, and serve as a form of index to chapters, these main divisions were generally **not used** as index terms. Thus there is no main entry for ‘finding aids’. Instead the reader will find entries for ‘bridging aids’, ‘descriptive inventories’, etc. thus providing quicker access to specific types of finding aids.

and again…

- Records management text, 1995
- You look up ‘indexing’ and find:

  indexing

  computer-assisted 95-6

  definition of 85-8, 93-4

- Go to page 95 and find it is in the middle of a 30-page chapter titled ‘Indexing records’. Where have the other 28 pages gone?

You find entries for them at:

computer-assisted indexing 95-6
consistency: importance in classification and indexing 63-4, 98, 102
derived indexing 97-8
full-text indexing 119-20
hierarchies in indexing 102-4
KWIC indexes 96-7
KWOC indexes 96-7
rotary indexes 94
strip indexes 94

and again…

- A broadcaster's annual report 2007-08

  In general, this index does not include topics listed in the Contents (page 1).

- So for the **Vision**, you go to the TOC, but for the **Code of Conduct**, you go to the index.
- You find **News** in the TOC (p.86), but **news and information** in the index (pp.10, 12, 18, 45, 64, 113, 240). No overlap.

So, if it’s OK to index chapter headings, is it compulsory?

- Not every chapter will generate one index heading that is co-extensive with its content, but in all cases the contents of a chapter should be accessible for someone who looks up the topic/s of that chapter.

Extrapolation of rules is a problem in library cataloguing too

Cutter says ‘Enter books under the word which best expresses their subject, whether it occurs in the title or not’; and Metcalfe adds ‘but this is misread by some cataloguers, who almost make the rule: enter under a word which is not in the title.’

Metcalfe, John. 1959 *Subject classifying and indexing of libraries and literature*

Rule 2: You don’t have to write the book in the index

- Whenever example index entries on Index-L are wordy, someone is likely to point out that index entries should be *pointers* to information in the text; they should not *contain* the information

Vigorous writing is concise.

William Strunk Jr., "The Elements of Style", 1919

The skill of writing is to create a context in which other people can think.

Edwin Schlossberg

Why flesh out entries

- Principle of safe arrival
- Full expression of concepts/Essence of the book

Principle of safe arrival – acronyms

ABA 15–17
EAD 56–58

or

ABA (Australian Broadcasting Authority) 15–17
EAD (Encoded Archival Description) 56–58

Assume they are also indexed at their full name.

Principle of safe arrival – acronyms 2

This needs to be judged in context. In my view:

EAD 56–58 will often be enough, but

ABA 15–17 could be

Australian Booksellers Association

Australian Broadcasting Authority

Australian Business Academy, etc

and more likely needs to be qualified.

Principle of safe arrival – homonyms 1

- How would you index the following in a book about satirical Australian TV shows?

  John Howard, the actor, said ‘Sorry’ to Indigenous Australians for their treatment at the hands of the English settlers and their descendants on the mock 24 hour news channel comedy series CNNNN.

Principle of safe arrival – homonyms 2

Howard, John: ‘apology’ to Indigenous Australians

would be misleading compared to:

Howard, John (actor): ‘apology’ to Indigenous Australians

Principle of safe arrival – homonyms 3

- How would you index the place in the sentence below?
- ‘The librarians and IT representatives met in Dublin to discuss the future of metadata standards’.

Principle of safe arrival – homonyms 4

- Even if there is just one Dublin in the book, the city name alone is not enough. It would be better to use:

  Dublin, Ohio

Full expression of concepts/ Essence of the book

ancient typographers

- Block cursors, horrible memories of
- remembering the days before color separation previews
- runts as banes of existence for
- typesetting purists, *vs*. typesetting realists

Real world Adobe InDesign CS3 (Kvern and Blatner)

Index by Jan Wright, winner of 2008 ASI/HW Wilson Award

When should you omit detail?

- When it can be meaningfully made shorter
- Too much info in an index entry usually means that
  - the indexer didn’t understand the topic well enough to step back and pull out the key features.
  - the author was the indexer and couldn’t step back and see things from the user’s point of view.

The book, unchanged, in the index

Malingering, techniques for identifying, 40-43

"I heard about this thing called 'Klonopin' from a friend who has what I have", 43

interviewing clues to, 41-43

"nothing works, Doc," 42-43

tale is just too perfect, 41-42

The psychiatric interview 2nd ed.

- So, you don’t **have** to write the book in the index, but if space permits, and you think more detail would add value, go for it!

Rule 3: Use specific entries

I always wanted to be somebody, but I should have been more specific.

Jane Wagner, (and Lily Tomlin)

Problems with the specific entry ‘rule’

- While emphasising specific entry when I teach indexing, I realised that it’s not as simple as I thought…

Problems with the specific entry ‘rule’

sentencing

see also suspended sentences

discount for guilty plea, 13–14, 29–30

parity in, 38–39

purposes of, 83–84

two-stage sentencing, 29–30

- Sometimes we want to break the rule. When should we?

Complexity about the specific entry ‘rule’

- Indexes provide access to the content of a document:
  - Alphabetical (or other logical order)
  - Specific/Direct
  - Selective

Three options in specific/classified distinction

- Specific, direct (alphabetico-specific)
  - micro-blogging
- Specific, indirect (alphabetico-classified)
  - blogging: micro-blogging
- Classified
  - Blogging
- In book indexing, specific, indirect entry is usually called classified entry. Metcalfe (p.268)also groups these as alphabetico-classified entries.

Specific, direct entry is preferred in indexing

- It means users can look up a topic under it's own, specific name – they don't have to know the broader term (class) selected by the indexer.
- It means users can find information on the precise topic they are interested in – they don't have to wade through irrelevant content.
- It is easier to keep subheadings logical, and to avoid subsubheadings.
- If they want an overview of the general area their topic is in, they can search for the general area as well.

Specific, indirect (alphabetico-classified) entry causes problems because users don't always know what broader term the indexer has used.

- But why is it so hard – of course a 'collie dog' is a dog.
- Yes, but perhaps 'micro-blogging' could be classed under 'social networking' not under 'blogging'.
- And where would you look for 'tailors' in a book called 'The Complete Tradesman'?
- They are be found under 'occupations', not 'trades', and 'tinkers' can be found under 'itinerants'. (Berson, 2007)

Specific, indirect (alphabetico-classified) entry causes problems because users don't always know what broader term the indexer has used.

- Where would you look for 'computers'?
- In one early classification, 'computers' was a subheading of 'calculating machines'!

  Borko and Bernier 1978: p.14
- Is 'metadata creation' in the class 'cataloguing', or is 'cataloguing' a member of the class 'metadata creation?
- Information professionals can't agree.

  SIGCR-L discussion 29 May to 11 June 2009

Specific, indirect (alphabetico-classified) entry causes problems because users don't always know what broader term the indexer has used.

- The fatal defect of every classified arrangement is that nobody understands it except the person who made it and he is often in doubt. (Poole, quoted in Wheatley's *What is an index?* p.56).
- But in indexing this page, alphabetical arrangement fails! It is indexed at:

  Classification *v.* the Alphabetical Arrangement ……..56

  but not at 'Specific' or 'Alphabetical'.

(It's clearly an emotive subject. The index to Wheatley's *How to make an index* has the entry 'Classification within the alphabet, Evils of, 58, 67)

Specific, indirect (alphabetico-classified) entry may be effectively used:

- To give an overview of the topic (eg, in legal indexes)
- Where the specific terms are not so likely to be consulted, eg, names of made-up case studies; ‘Group Functions’ under ‘Public Service Commission’
- Where narrower terms may not be as important to users – where specific is **too** specific
  - Knight (pp.98-99) quotes Metcalfe’s example of the topic ‘Model T Ford Motor Car’ and concludes that in general literature ‘the indexer might have to invert it into “Ford motors, Model T” and that it is not necessary to descend to the *infima species*
  - He also concluded that ‘circular flying saucers’ could be inverted as ‘circular’ can be treated as a mere aspect or qualification. (Could also be classified just at ‘flying saucers’).
- As a supplement to specific entries, especially to avoid the need for cross-references

Specific, indirect (alphabetico-classified) entry for a topic overview

- Mainly in legal indexes
- '…is it in the purpose of an index to pre-arrange the concepts dealt with in the text? Or is that not the function of a table of contents, while an index is a more democratic instrument, giving rough equality of access to all the concepts? (Green, 1989)
- 6 lecturers at the Law School in Cardiff were given 6 topics that were found at different levels of subdivision in the index to the Law Quarterly Review. None were able to find all of them.
- And pity the poor beginner who doesn't know that 'defective premises' are to be found at:

  Torts: negligence: duty of care: defective premises

How would you index this?

…or this

Narrower terms not important or not known

- fusilli
- rotini
- spiral pasta
- pasta

Specific entry sometimes demands indexer knowledge and user knowledge and interest

- AC Foskett p.27: Whatever system we choose to use, there are two persons who must find it usable: the person responsible for the input, ie the *indexer*, and the person trying to obtain an output, ie the *user*.
- *Non omnia omnes possumus*: we cannot all be omniscient (or, we can't all do everything. Virgil)

Sometimes narrower terms **are** important

At the cookery and food indexing panel discussion at the 2007 ANZSI conference Tricia Waters presented a reading from Anne Tyler's novel *The Accidental Tourist*.

Rose's brothers were helping her to unpack the shopping. 'Rose stood on a stepstool in front of a glass-fronted cupboard, accepting the groceries which her brothers, Charles and Porter handed up to her. 'Now I need the N's – anything starting with N?' she asked. 'How about these noodles?' Porter asked. 'N for noodles or would that be P for pasta?' 'It's E for elbow macaroni, Rose retorted, you might have passed those up earlier'.

Specific, indirect (alphabetico-classified) entries as a supplement to specific entries

sentencing

see also suspended sentences

discount for guilty plea, 13–14, 29–30

parity in, 38–39

purposes of, 83–84

two-stage sentencing, 29–30

two-stage sentencing, 29–30

- So long as we have a specific entry for ‘two-stage sentencing’, it is quicker for the user to double it as a subheading under ‘sentencing’ than to add a cross-reference leading from the broader term to the specific entry.

Specific, indirect (alphabetico-classified) entries as a supplement to specific entries

- ‘hospital libraries’ at ‘libraries, hospital’ as well as at ‘hospital libraries’
- ‘chicken noodle soup’ as a sub under ‘chicken’, ‘noodle’ and ‘soup’ as well as under its specific name
- ‘MS-Access’, ‘MS-Excel’ and ‘MS-Word’ as subheadings under ‘software’ as well as being as entries in their own right.

Specific, indirect (alphabetico-classified) entries as a supplement to specific entries – problems with

- When you choose to index a topic in its own right and as a subheading of a broader term, it's best to keep the entries in one place with a see reference to the other until the last minute. This avoids problems such as:

Genres in film, 94-109
analysis of genres, 96-97
definition of word, 93, 94
westerns, 101-102
westerns, 97, 99, 101-102

Subset of example from Stauber, p. 164. Indexed as 'Subtopic categorical subheadings', with a reference from 'Classified subheadings'

Classified entry (not co-extensive with topic) may be effectively used:

- When the specific entry is just too specific
  - rotini
  - North Queensland
  - corporate libraries
  - Here the broader term is easy to find, as it is a component of the narrower term. Eg, if you find nothing at ‘corporate libraries’ you are likely to then try ‘libraries’
- When indexing at different levels of granularity
  - ‘indexing’ for whole chapter, and ‘strip indexing’ for a page.

Classified entry (not co-extensive with topic) doesn’t work when:

- It stops you going straight to the topic of interest, especially when you can’t quickly tell what is relevant
  - If a discussion of ‘micro-blogging’ is indexed at ‘blogging’, and there are 5 references there, it won’t be clear which of those 5 locators refers to micro-blogging (rather than blogging in general).
  - Yellow Pages has a category ‘Editorial Support Services’. This is a class entry for ‘Indexing’ (one of the editorial support services) but most callers are looking for broader-services.

Too classified Too specific Just right

Classified entries in periodical indexing

- ‘book reviews’, ‘editorials’ etc. Best in addition to specific entries, but may be used alone to save time and space.

Classified entries in database indexing

- Formal classification schemes may be used in bibliographic database indexing.
- Lancaster (2003: pp.270-271) suggests that keyword search (supplemented with a synonym ring) works relatively well for specific terms, and that intellectual indexing effort is best directed towards indexing showing broader classifications. These are useful for:
  - Broader searches, eg, ‘toxins’, places
  - Eliminating irrelevant hits for terms which have different meanings in different fields, eg, ‘sentencing’, ‘mass’, ‘irrigation’.

Classified (hierarchical) vs categorical (concept) groupings

- Overviews for biographies, eg headings such as ‘Early years in France’, ‘In Germany’ and ‘Voltaire and social reform’ (Wellisch, 1st ed, pp. 148-149).
- Where grouping is considered more important than specific entry, eg titles of books under the name of the author.
- Making interesting connections. Gather threads you discover that weave their way subtly through the work. Do Mi’s rule – if you pick it up, pick it all up.
  - serendipitous discoveries
  - indexer confusion.

How to get specificity ‘just right’, like Goldilocks

- Audience of book – book on pasta, or book on simple cooking for students.
- Specificity of book.

How to get specificity ‘just right’, like Goldilocks

- Quality indexing has multiple access points – people like a TOC-style broader entry to lead them elsewhere.
- Double entry and cross-references mean never having to say you’re sorry.
- Check specific article and quote him

But the key point is…

- …specific entry is still the FIRST choice. Classification is then fine as a SECOND choice.

Social Networking for Indexers – Panel Discussion – with responses added

Glenda Browne

Jan Wright

Tracy Harwood

Overview

- This panel session will discuss the electronic systems through which indexers communicate ‘en masse’.
- Most indexers are on at least one mailing list, but many haven’t tried the more social sites, or have tried them, and found them wanting.
- Here we will look at the information and social needs indexers have, and ask the audience for thoughts on which resources serve those needs.

Social networking participation

- Forrester research: http://blogs.forrester.com/.a/6a00d8341c50bf53ef0120a5175967970b-popup

In praise of social networking

Where there is much desire to learn, there of necessity will be much arguing, much writing, many opinions; for opinions in good men is but knowledge in the making.

John Milton (1608 - 1674)

- The numbers in the following slides are the number of people in the audience (total 22 people) who used the lists, sites, etc that we were asking about. The conference social networking survey received responses from 30 indexers and will be presented as a separate document.

Survey – Which mailing lists are you on?

Mailing lists	Numbers
Index-L	8
General indexing (eg, IndexCafe)	3
Specialised indexing (eg, web indexing)	4
Indexing software	10
Other professionals (editors, tech writers, librarians)	13

Survey – Which social networking tools do you use?

- Mailing list/Social site
  - Ning
- General social networking sites
  - Twitter
  - Facebook
  - LinkedIn

- Numbers
- 6

- 4
- 14
- 3

Survey – Which social networking tools do you use?

- Social bookmarking
  - Delicious
- Book-related
  - LibraryThing
  - Shelfari
- Traditional websites
  - Society websites

- Numbers
- 2

- 6
- 1

- Not asked

Survey – Which social networking tools do you use?

- Blogs
  - reading
  - writing

Numbers

9

1

Where do you go to…?

- Where do we all go, online, to meet our professional, business and social needs?
- Get advice from other indexers about indexing?
  - Dealing with clients (eg, late payment)
  - Technical advice on specific problems
  - Indexing choices (eg, term selection)
  - Research publishers and editors
- Hear news from fellow indexers
  - Software updates
  - Websites of interest
  - New publications

Feedback

- We thought of: The mailing lists, Ning and LinkedIn. Society websites for long-term stuff, eg online newsletters.
- The audience suggested Index-L, indexing software mailing lists, society newsletters and websites, but most preferred personal contact via phone calls, emails, committee meetings and Skype. The suggestion ‘Ring Max’ was well-received, but not considered scalable.
- There were comments that we cannot know who is on the mailing lists, so have to write messages appropriate to the context.
- When a news feed on a website was suggested there was some discussion about the purpose of society websites – whether the aim is to provide news to members, or to promote members to the wider world.

Where do you go to…?

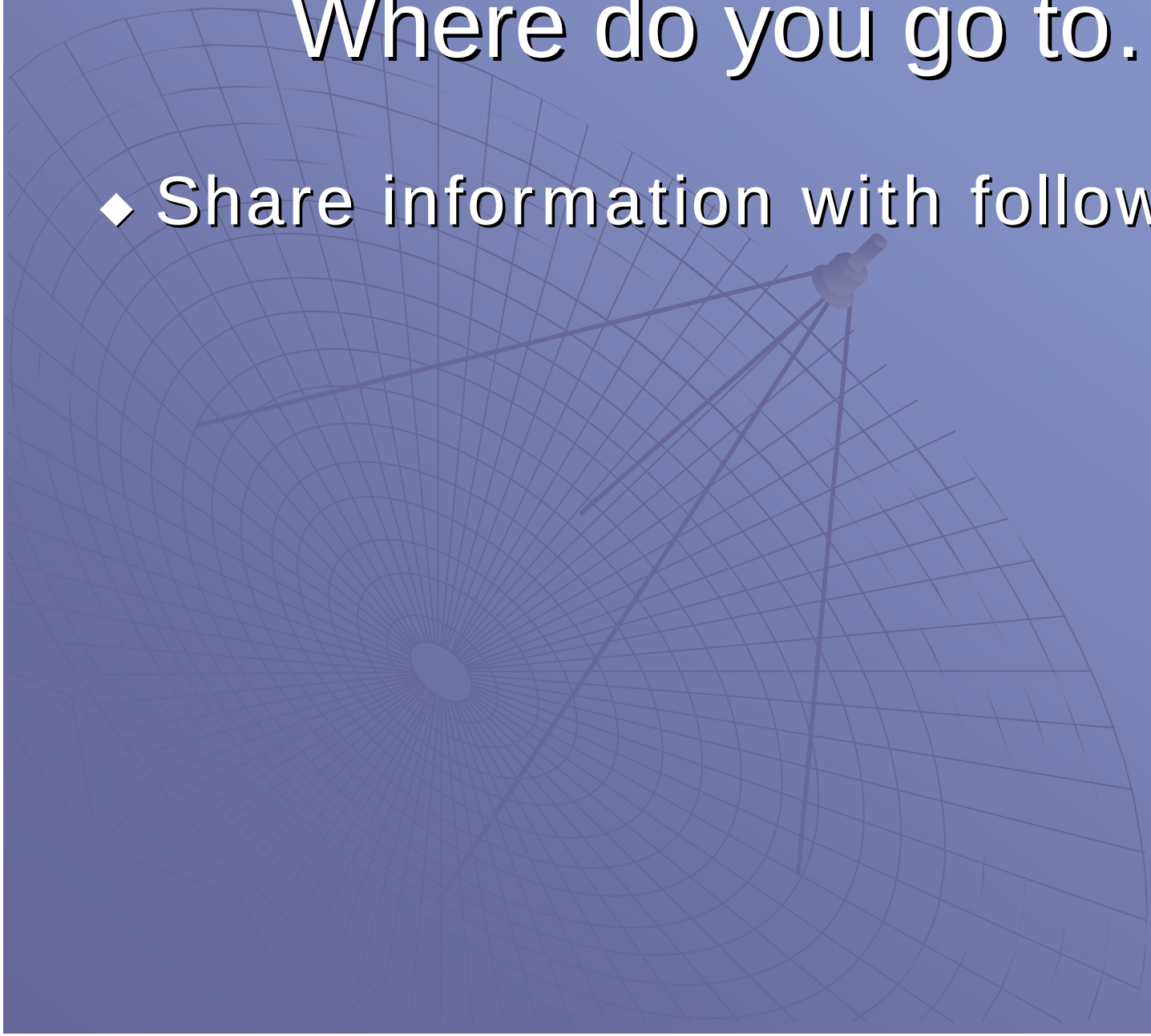
- Get opinions and information about topics related to indexing (eg, editing, taxonomies)?

Feedback

- We thought: Editors', tech writers', and taxonomists' mailing lists.
- The audience also thought about library lists, Google searches and bibliographic databases. And, of course, books.

Where do you go to…?

- Share information with followers



Feedback

- We thought: Twitter for URLs and brief announcements and as a pointer to other places. Websites. Ning and Twitter perhaps for topics less directly associated to indexing. Viral spread from one site to another. Acknowledgement of sources documents the pathway
- The audience commented on company control that would not allow use of sites such as Twitter, and mentioned a recent lawsuit that involved discussion of lawsuits on Twitter.

More on Twitter

- Twitter - #ANZSI2009, #dailyentry (these are two of the hashtags used by indexers to enhance retrieval. If everyone tweeting about our conference uses the same tag in their messages, we can easily gather content by searching for '#ANZSI2009').
- Twellow (categories of tweeters)

Where do you go to…?

- Share information with the world
  - Delicious

Feedback

- We thought: Delicious, book sharing sites, official society websites. Your own blogs.

Where do you go to…?

- Promote/advertise your services
- Display your credentials
- Make sure that your colleagues know your specialties
- Build a network of colleagues to refer and get referrals from
- Decide which colleagues you want to refer work to

Feedback

- We thought: Twitter, Facebook. LinkedIn. Mailing lists in a subtle way. Book cataloguing sites in a subtle way (including tag ‘indexed by me’).
- The audience mentioned Linked-In as being useful for contacts with permanent staff, and as a way of looking for editors.

Where do you go to…?

- Get announcements of upcoming events

Feedback

- We thought websites, Ning and Twitter.
- The audience thought of websites. A discussion started about jobs being promoted through society websites, which the Canberra editors do. (And the NSW editors do by email.)

Where do you go to…?

- See pictures of indexers you have never met
- Learn more about other indexers personally and their hobbies and interests
- Chat
- Ning

Feedback

- We thought: Ning, Facebook, Twitter
- Jan talked about her plans for Ning, and her hopes for the development of an interactive, multi-purpose website for indexers.

Where do you go to…?

- Find out about book-related matters and book lists and metadata

Feedback

- We thought: LibraryThing, Shelfari, Amazon (reviews and comments)
- We didn’t get this far in the panel discussion, so there are no more audience comments.

Where do you go to…?

- Keep in touch with friends, family and friendly indexers
- Get back in touch with colleagues

Feedback

- We thought Facebook

Statistics from conference survey (30 responses)

- See Survey Responses sheet and Summary

Searching and ‘Indexing’

- Archives
- Threads
- #tags on Twitter
- Tags in inverted commas on Ning
- Delicious – no controlled vocabulary, but site offers suggestions that aid consistency
- Not so easy though, so most people just ask the question afresh

Duplication of content

- Choosing between sites, eg, aliaINDEXERS mailing list, and Australian indexers on Ning
- Duplication of effort to cross-post
- Duplication of effort to delete duplicates
- Perhaps there will be future consolidation, and more cross-linking

Where to start…

- Difficult to choose one place to go, as they offer overlapping services. Suggestions:
- For info on all aspects of indexing (references, events, etc) undiluted with waffle, go to society websites.
- To ask questions, and to keep up-to-date with new technical and business developments
  - Index-L (or Indexers Discussion Group or Ning) for general contact and discussion
  - Mailing list for your indexing software
  - Twitter for URLs of interest.
- To keep in touch with indexers as people
  - Twitter, Ning, Facebook (unless you keep it personal), blogs
- To keep in touch with indexers professionally and to expand your professional network (eg, find out what jobs they are working on, and connect to their connections) - LinkedIn
- To record and share information – Delicious for tagging websites; LibraryThing and Shelfari for cataloguing book collections

Netiquette

- How to avoid flame wars (bitter, personal, ongoing arguments)
- When to take it off list (when discussions drag on, and only two people are posting, it is usually time to stop mailing the whole list).
- Change the subject line when you change the topic
- Promotion vs. Infomercials (this is a problem on sites such as Linked-In, when many questions are simply disguised self-promotion)
- Befriending and defriending (you need to make judgements about who to connect to on these sites, and at times you may wish to disconnect from someone).

Bewares

- Remember, replies go to the whole list
- Remember, answers are one person's opinion. Answers may be wrong. Look for consensus. Evaluate.
- Cultural and language differences. Eg, '12/5/09' is clearly December to an American...

Conclusions

- Despite all of the options now available, appreciation goes to Index-L, the Grand Old Dame of social networking, which made a significant difference when it came into our lives about two decades ago.

Creating a large author index from a list of references

Andrea L. Jordan[1] & Donald O. Jordan[2]

[1]*jordy55@gmail.com;* [2]*donjor@melbpc.org.au*

Abstract

A Name index of authors cited in the 2nd edition of a large (824-page) academic text book on psychology, together with a Subject index, was required by the publisher. A PDF file of the text, together with a list of about 4900 references in a Word file, was provided to the indexer. Each reference in the text consisted of surname(s) and year of publication, while the references list also contained the authors' initials, the title of the paper, and the publication in which it appeared. This paper gives a detailed description of the use of the Adobe Acrobat software package in conjunction with a Microsoft Excel spreadsheet to locate the references in the text and create a list associating authors' names with the text page(s) on which they occur. The time taken for the work is given, as well as suggestions for action appropriate to the various problems that arose during this project, and that could possibly occur in other projects of this type. It is aimed at new indexers, and also experienced indexers who may not have previously used the software in this way.

Method

Overview

The references quoted in the text contained the authors' surnames and year of publication *(e.g. Smith, Jones & Huynh, 1990, or Black et al., 2003*). The References list also contained the authors' initials, title of their paper, and the publication in which it appeared.

Using Microsoft Word's search-and-replace functions, the authors' names were extracted from the References list and sorted alphabetically. This list was then pasted into a Microsoft Excel spreadsheet and Excel's advanced filter function was used to create a list of unique records.

Moving from reference to reference in the original Word document, each reference was searched for in the final-draft PDF files of the publication. The page(s) on which the reference appeared was then entered against the author name(s) in the spreadsheet. This constituted the bulk of the work by far.

When this process was completed for every reference, the Excel data were sorted and pasted into Word, and the data transformed into proper index entries using simple search and replace functions.

The procedure used is described below, written as a set of instructions for brevity. The bullet symbol (•) represents a space.

Step 1. Prepare files for use

1A. Extract names of authors from reference list

These instructions are valid as long as the references are presented in the following format:

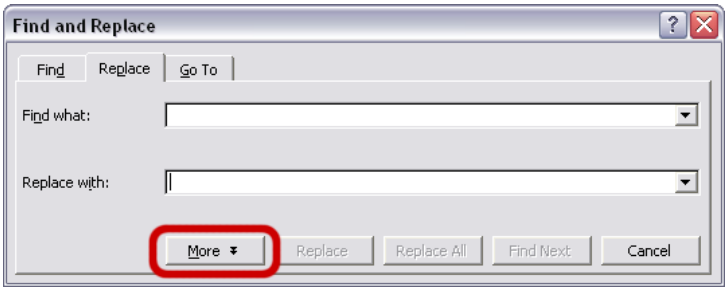
> Smith, A., Nguyen, T., & Chang, D. (2008). [Details of paper, chapter or book]

Create a list containing only authors' names

1. Open the list of references in Microsoft Word and ensure that all references are formatted correctly.

   > Ader, R., & Cohen, N. (1985). CNS immune system interactions: Conditioning phenomena. Behavioral and Brain Sciences, 8, 379–426.

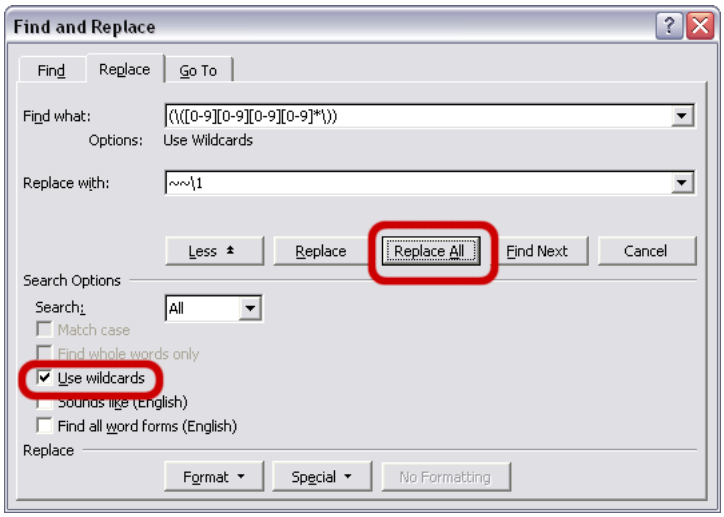
2. Place the cursor anywhere in the document (it is important only that no text is selected), and, on the **Edit** menu, choose **Replace…** to open the **Find and Replace** dialog box.
3. If the **Use wildcards** check box is not visible, click **More**.



4. Select the **Use wildcards** check box.
5. Enter the following string in the **Find what** box: (\([0-9][0-9][0-9][0-9]*\))

   This finds: open parenthesis, four digits (=date), any number of characters, close parenthesis. The extra characters (represented by the asterisk) following the date cover cases such as (2001a) and (2007, January).
6. Enter the following string in the **Replace with** box: ~~\1

   This represents: double tilde, "Find what" text. A double tilde was used by us as a unique "marker" term, but anything not used elsewhere in the document will do.



7. Click **Replace All**. Double tildes will appear before the date parentheses in all references (as long as there are no formatting errors in the references, of course).

   Ader, R., & Cohen, N. (1985). CNS immune system interactions: Conditioning phenomena. Behavioral and Brain Sciences, 8, 379–426.
   becomes
   Ader, R., & Cohen, N. ~~(1985). CNS immune system interactions: Conditioning phenomena. Behavioral and Brain Sciences, 8, 379–426.
8. To delete all of the text from the marker term to the end of the paragraph, continue working in the **Find and Replace** dialog box.
   Enter the following string in the **Find what** box: ~~*^13
   Enter the following string in the **Replace with** box: ^p
   Click **Replace All**. All text from the double tilde to the end of the reference should be deleted in every reference, leaving only a list of authors.

   Ader, R., & Cohen, N.

In simple search and replace, the combination ^p is used to represent a paragraph mark. But when Use wildcards is checked, ^13 must be used. A full list of codes and wildcards can be found online at http://office.microsoft.com/en-au/word/HP051894331033.aspx

9. To delete any trailing spaces, continue working in the **Find and Replace** dialog box. Uncheck the **Use wildcards** checkbox.
   Enter the following string (where • is a space) in the **Find what** box: •^p
   Enter the following string in the **Replace with** box: ^p
   Click **Replace All**. Any space before a paragraph mark will be deleted.

Separate authors

The next set of instructions will lead to each author name being extracted into a paragraph of its own. Steps 10–13 are all executed within the **Find and Replace** dialog box.

10. Enter the following string in the **Find what** box: •^$.,•&•
    Enter the following string in the **Replace with** box: ^&^p
    Click **Replace All**. This inserts a paragraph mark after the ampersand leading the final author in a multi-author reference.

    Ader, R., & Cohen, N.
    becomes
    Ader, R., &
    Cohen, N.

11. Enter the following string in the **Find what** box: ,•&•^p
    Enter the following string in the **Replace with** box: ^p
    Click **Replace All**. This deletes the trailing comma and ampersand after the penultimate author in a multi-author reference.

    Ader, R., &
    becomes
    Ader, R.

12. Enter the following string in the **Find what** box: •^?.,•
    Enter the following string in the **Replace with** box: ^&^p
    Click **Replace All**. This inserts a paragraph mark after the comma separating authors in a reference with three or more authors.

    Adams, H. E., Wright, L. W.
    becomes
    Adams, H. E.,
    Wright, L. W.

13. Enter the following string in the **Find what** box: ,•^p
    Enter the following string in the **Replace with** box: ^p
    Click **Replace All**. This deletes the trailing comma and space after author initials in a reference with three or more authors.

    Adams, H. E.,
    becomes
    Adams, H. E.

Sort names into alphabetical order

14. Check the list of authors to ensure that no errors crept in during the search and replace process. Delete any heading(s) at the beginning of the document. Place the cursor at the beginning of the document and, on the **Table** menu, click **Sort**.

15. The default sort settings should not need to be changed (Sort by Paragraphs; Type: Text; Ascending; No header row). Choose **OK**. The author names will be sorted alphabetically.

Ader, R.
Cohen, N.
Adams, H. E.
Wright, L. W.

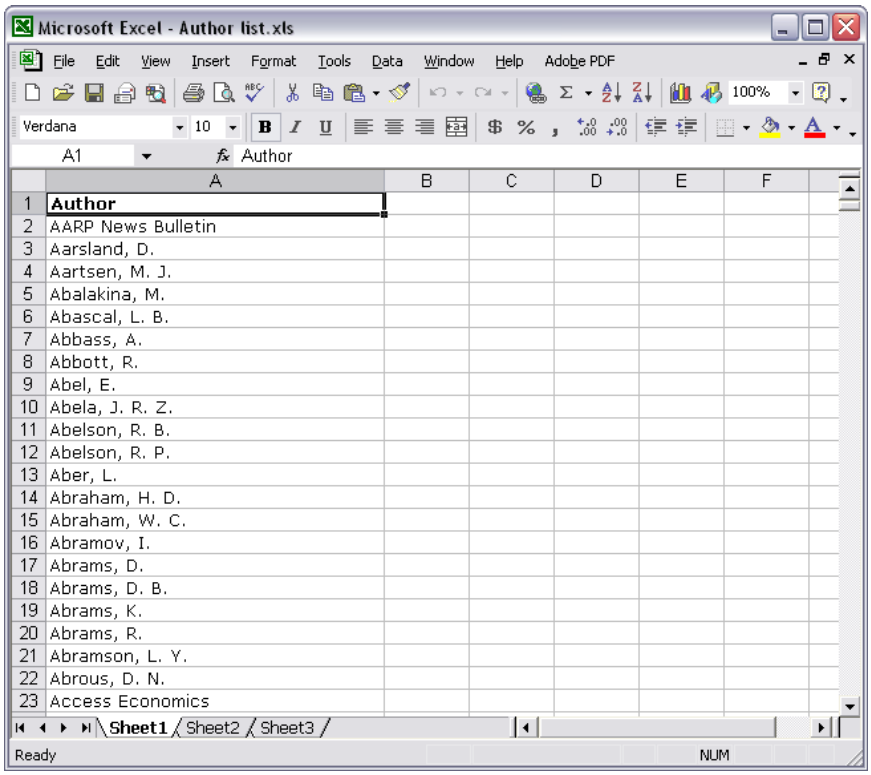
becomes

Adams, H. E.
Ader, R.
Cohen, N.
Wright, L. W.

1B. Prepare Excel spreadsheet with author names

Paste the list of names into Excel

16. Select all of the text in the Word file (Ctrl+A) and copy to the clipboard (Ctrl+C). Create a new Excel document (in Excel, open the **File** menu and choose **New**), place the cursor in cell A2 and paste in the text (Ctrl+V). Type a heading in cell A1 for column A (e.g. 'Author').



17. Use Excel's advanced filter function to create a unique list of records. Selecting cell A1, click on **Data**, point to **Filter** and choose **Advanced Filter…** Click **Copy to another location**. Place the cursor in the **Copy to** box and type "B1" (this will place your new list into column B). Click Unique records only and choose OK. The original list will be copied into column B, but without any duplicates.

18. Compare the data in columns A and B to check that the filtering has worked. When you're satisfied, delete column A (select the column by clicking on the A at the top, right-click anywhere within the column and select **Delete**).

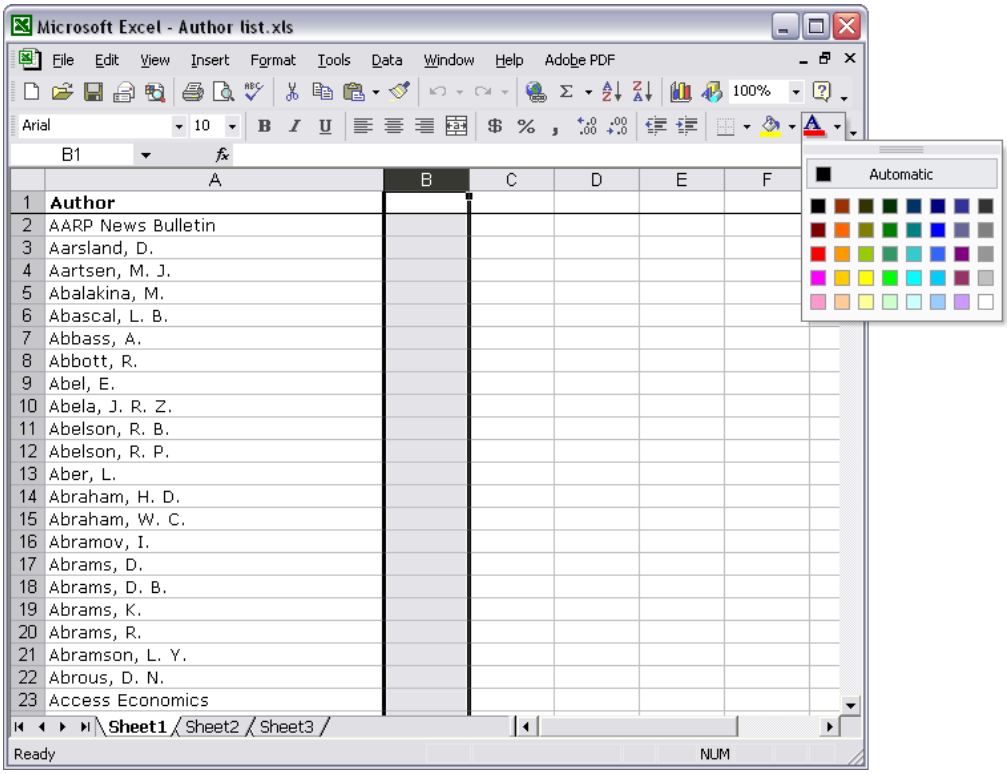
19. **Save** the file.

Prepare Excel file for use

20. In the same file, place the cursor in cell B2. Open the **Window** menu and choose **Freeze panes**.

    This keeps row 1 and column A visible as the indexer moves around the spreadsheet.

21. Select column B (click on the "B" at the head of the column) and use the Font Colour dropdown menu to choose a colour (blue is nice) to differentiate the first page number from the author name.



As there may be several page numbers for some authors, which will be placed in the columns following the name, doing this helps the indexer to be sure that the first page number for an author is being entered in the B column.

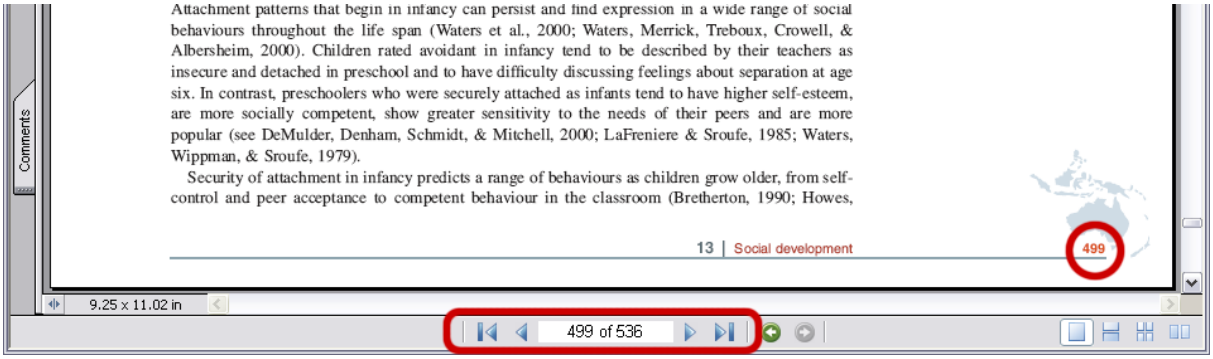
22. To choose the direction in which the cursor travels after 'Enter' is pressed, open the **Tools** menu, point to **Options** and click the **Edit** tab. Ensure the **Move selection after Enter** checkbox is ticked and that the direction **Right** is selected.

    This helps where more than one page number for an author is being entered at the same time.

1C. Prepare PDF file and Word Reference list for use

23. If text chapters have been sent as separate PDF files, create one PDF file that includes all chapters. To do this, open the first chapter in Adobe Acrobat. In the **Document** menu, point to **Pages** and choose **Insert**. Navigate to the next chapter and instruct Acrobat to insert it after the final page in the open document. Save the document with a new name. Repeat with the remaining chapters.

24. Ensure PDF page numbers (shown at the foot of the Acrobat window) match page numbers printed on pages. To do so, it may be necessary to add blank pages to the beginning of the document.



*To add blank pages to a PDF document, create a PDF file containing the number of blank pages required. Then, open the document to be indexed in Adobe Acrobat and, in the **Document** menu, point to **Pages** and choose **Insert**. Navigate to the file containing the blank pages and instruct Acrobat to insert it before the first page in the open document. The page numbering at the foot of the Acrobat window should then match the page numbers printed in the document.*

25. Open the original References file in Word. Select all of the text (**Ctrl+A**) and colour the text (once more, blue is nice) using the **Font Colour** dropdown menu. Save the file.

    The indexer will be able to turn the references to black as they are successfully indexed. With a long list of references it is useful to be able to see at a glance what references have and have not been indexed.

Having completed all these steps, the indexer will have three files ready to use: a list of authors' names in an Excel file, a PDF file of the whole text, and a list of references in a Word file.

Step 2. Find references in text and note page numbers against author names

1. Open the three files to be used during the indexing process: the text in Adobe Acrobat (not Acrobat Reader), the references list in Word, and the Excel spreadsheet containing the list of authors' names. Arrange the windows next to each other so that all three files can be seen, with Acrobat in the left pane, Word in the centre pane, and Excel in the right pane.

   *To open the PDF file in Acrobat rather than Adobe Reader, open Acrobat from the Windows Start menu (or equivalent), then choose **Open** from the **File** menu.*

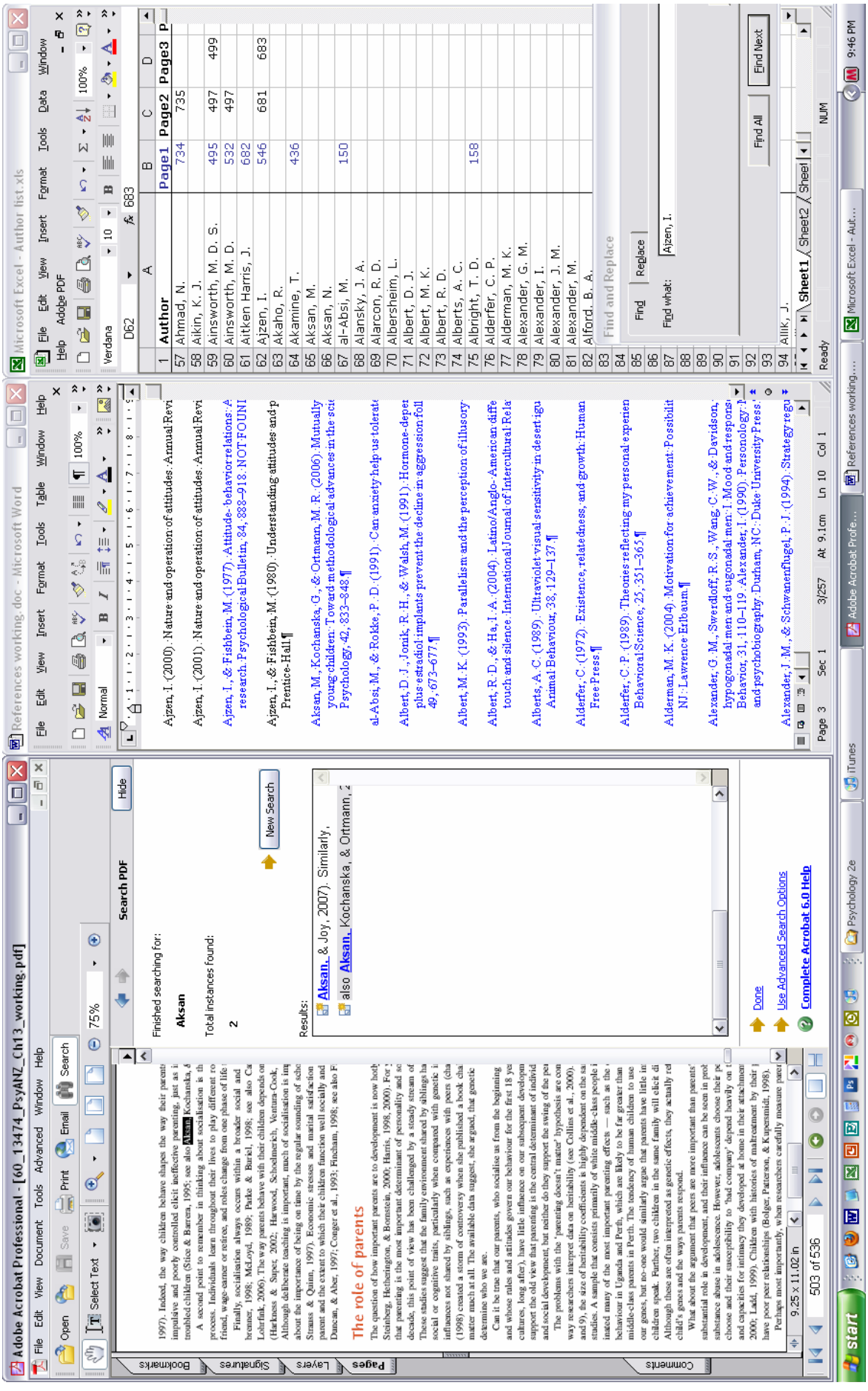
2. In the PDF file, magnify the text using the **Zoom** function as needed for it to be legible (75% was enough for this particular job).

3. In the spreadsheet, press **Ctrl+F** to open the **Find and Replace** box. This only needs to be done at the beginning of each work session.

4. In the References list, highlight the first surname in the list (no initials) and press **Ctrl+C** to copy the name to the clipboard.

5. In the Acrobat pane, click on **Edit**, then **Search** to open the **Search PDF** box. Choose **In the current PDF document** and tick **Whole words only** and **Case-Sensitive**. This also only needs to be done at the start of each session.

Appearance of windows while working



6. Place the cursor in the **What word or phrase would you like to search for?** box and press **Ctrl+V** to paste in the first surname. Then click on **Search**. A list showing all the appearances of that name (or word) in the PDF text will appear. Among the items should be the reference(s) being sought. Click on it to have the page appear with the reference highlighted on it, and note the page number. It is possible to check that the correct page is open by hovering the cursor over the selected reference and seeing that it shows the same page number. (If it doesn't, try clicking the reference again.) If the reference appears on more than one page in the text, enter the page numbers for each citation in the spreadsheet.

   If the reference cannot be found, it may be missing from the text, or there may be some inconsistency between the Reference list and the text (e.g. a spelling error or the name may be hyphenated if it occurs at the end of a line in the text). Try searching without Case-Sensitive checked or without Whole words only checked, or try searching for just part of the name. For a multi-author reference, try searching on the second author's name.

7. Place the cursor in the **Search for** space in the **Find and Replace** box of the spreadsheet and press **Ctrl+V** to paste in the primary author's surname. Then click on **Find Next** and that surname (there may be more than one) will appear. Check that the initials beside the surname are the same as in the References list and then select the first available cell to the right. Type in the page number and press **Enter**. The number pad is used for this, while the left hand hovers over the **Ctrl**, **C**, and **V** keys.

8. Return to the References list and highlight the name, including initials, of the second author and press **Ctrl+C**. Go to the spreadsheet, click on **New Search**, highlight the name that is in the **Search for** space, then press **Ctrl+V** to paste in the second author's name (with initials). Now click on **Search** and that author's name will appear (no need to choose from several surnames, as the initials have been included). Enter the page number as for the primary author.

9. Repeat this process for all authors of the publication. Having done this, select the author or authors in blue in the References list and return the colour to black using the **Font Colour (Auto)** button on the Formatting tool bar. This shows that the reference has been found and all authors allotted the appropriate page number.

10. If a reference in the list is not found in the text, leave the author's name(s) in colour and type 'NOT FOUND' beside it as insurance against the colour being accidentally returned to black.

Step 3. Prepare final index

3A. Change spreadsheet data into index entries

Sort page numbers row by row

Microsoft Excel can be used to sort data in all kinds of ways, but its capabilities are all based on sorting data within columns in a hierarchical way. Since the data are arranged in rows in this job *and* the rows need to be sorted independently of each other, Excel's built-in sorting functions cannot be used.

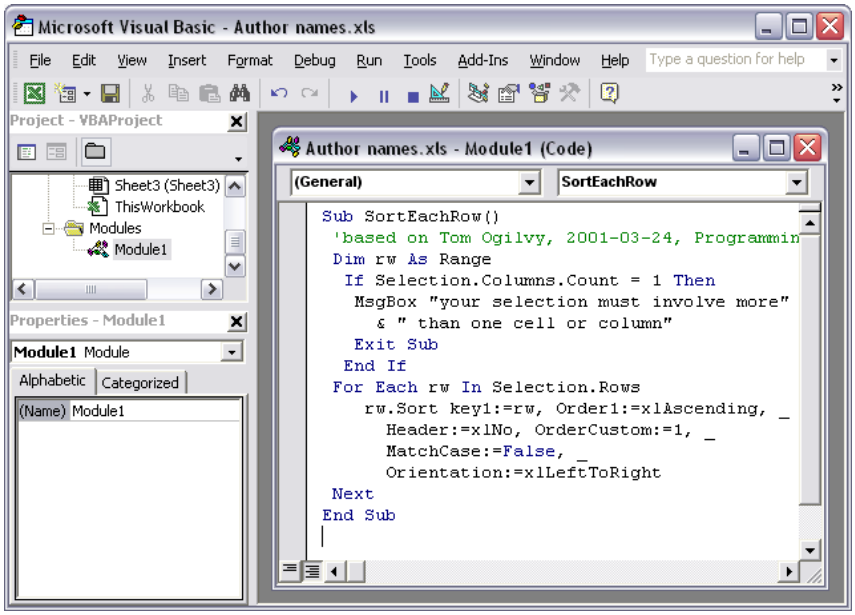
To address this problem, a programmer (Tom Ogilvy) wrote a short macro (set of commands) that can be applied to a table of data in Excel to sort each row independently of the others. The macro is available in a slightly modified form (constructed to work on a selection rather than a specific range) on David McRitchie's Excel Pages (http://www.mvps.org/dmcritchie/excel/sorting.htm).

The following points explain how to incorporate the macro into the Excel file and run it on the page number data.

1. Open the Excel file, click **Tools**, point to **Macro** and click on **Visual Basic Editor**. The Visual Basic window will open. From the **Insert** menu, choose **Module**. A small module window will open within the Visual Basic Editor.

2. Type the following into the module window:

```
Sub SortEachRow()
 'based on Tom Ogilvy, 2001-03-24, Programming
 Dim rw As Range
  If Selection.Columns.Count = 1 Then
   MsgBox "your selection must involve more" _
     & " than one cell or column"
   Exit Sub
  End If
For Each rw In Selection.Rows
    rw.Sort key1:=rw, Order1:=xlAscending, _
      Header:=xlNo, OrderCustom:=1, _
      MatchCase:=False, _
      Orientation:=xlLeftToRight
 Next
End Sub
```

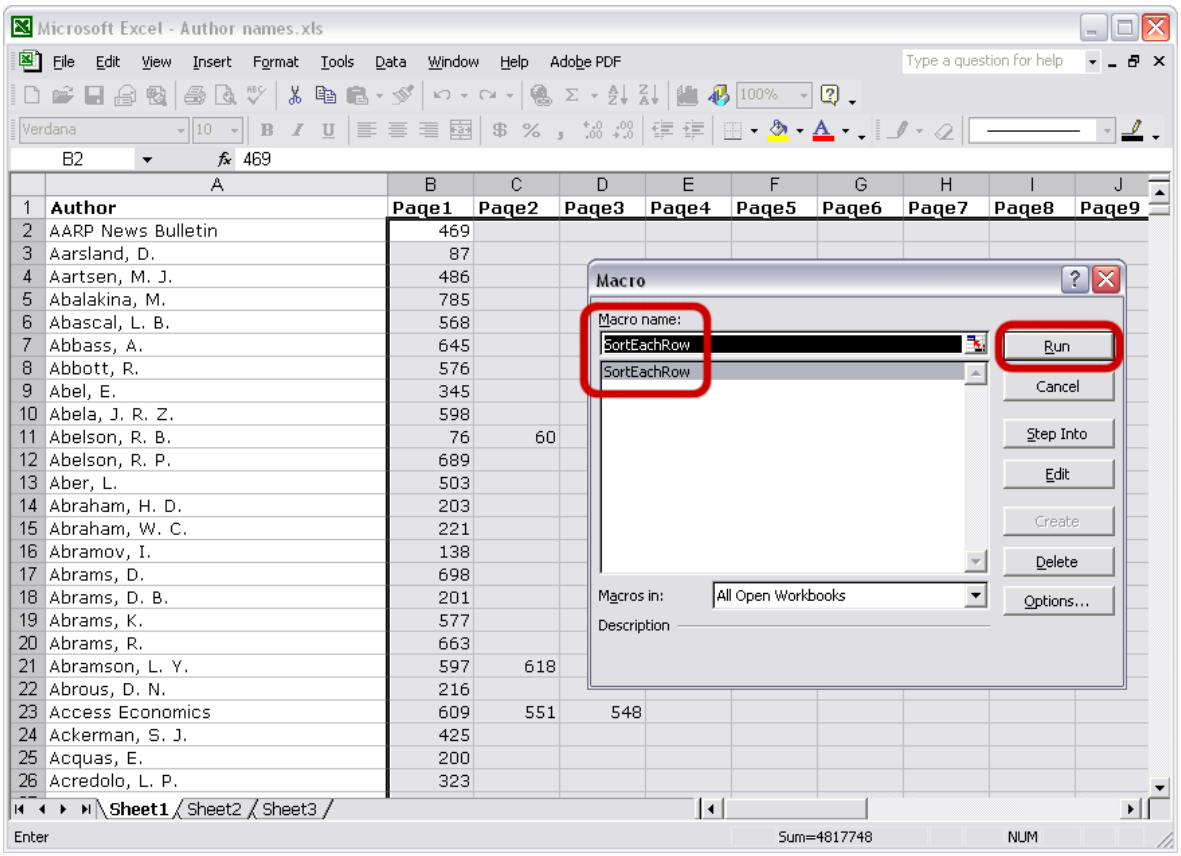


It is easy to make errors retyping text like this. It may be easier to copy and paste it from McRitchie's web page.

3. From the **File** menu of the Visual Basic Editor, choose **Close and Return to Microsoft Excel**. The macro is now saved in the Excel document and ready to run. Beware that running the macro cannot be undone. **Make a backup copy of the file *and* save before the next step is performed!**

4. Select the area to be sorted. This should be from the top left page number (cell B2) to the bottom right page number. DO NOT include the author names or column headings, lest these also be sorted.

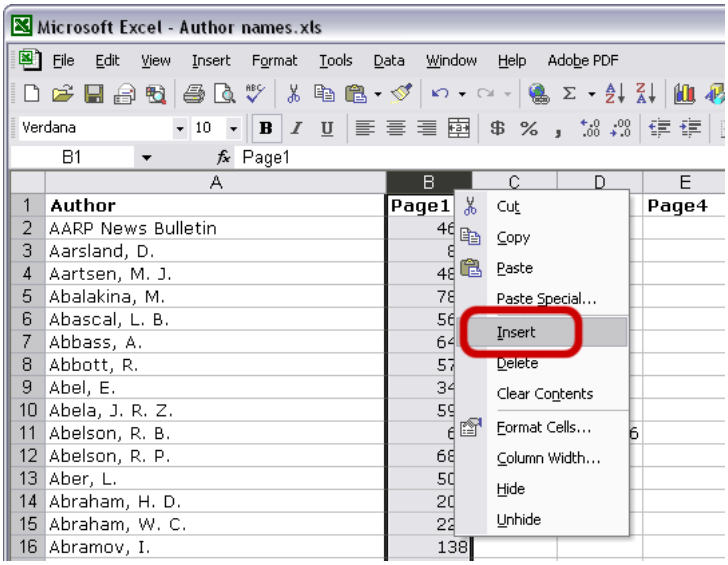
5. From the **Tools** menu in Excel, point to **Macro** and click on **Macros…** Ensure that the macro name **SortEachMacro** is selected, take a deep breath and click **Run**. Excel should now take a moment to sort all the page numbers, row by row.



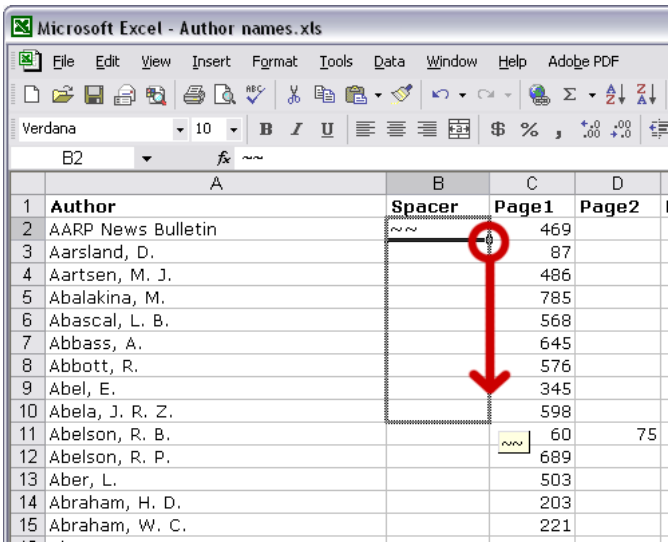
6. Check the file carefully. If all looks well, save the file. If anything goes wrong, close the file without saving it, reopen it and try again. The most common mistake will probably be selecting the wrong area for sorting.

Prepare data and paste into Word

7. To indicate the gap between the author names and the page numbers for the editor, and to provide a unique term in that position for cleaning up the document later, it's necessary to add a column between the author names and the page numbers. Right-click on column heading B and choose **Insert**.

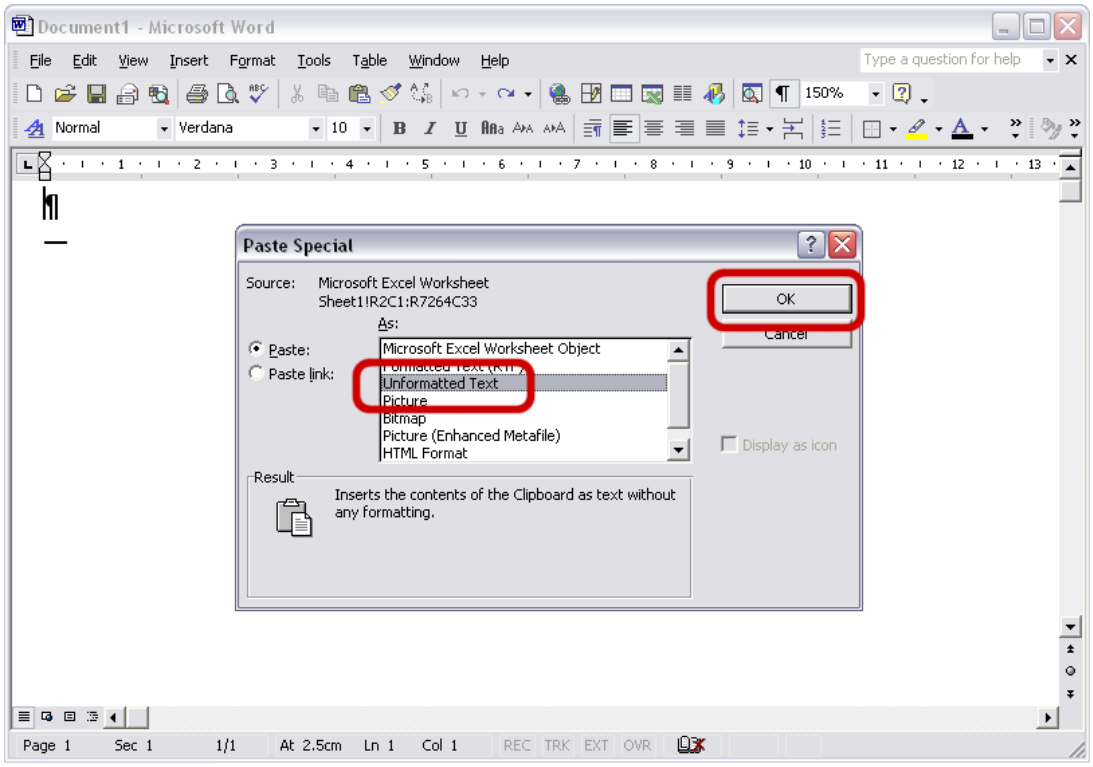


8. Type a heading at the top of the column and then enter a unique term in each cell (a double tilde was used for this job, for no particular reason). The easiest way to do this is to type "~~" into cell B2, select it and then drag the handle at the lower right corner of the cell to copy the cell contents down the entire column.

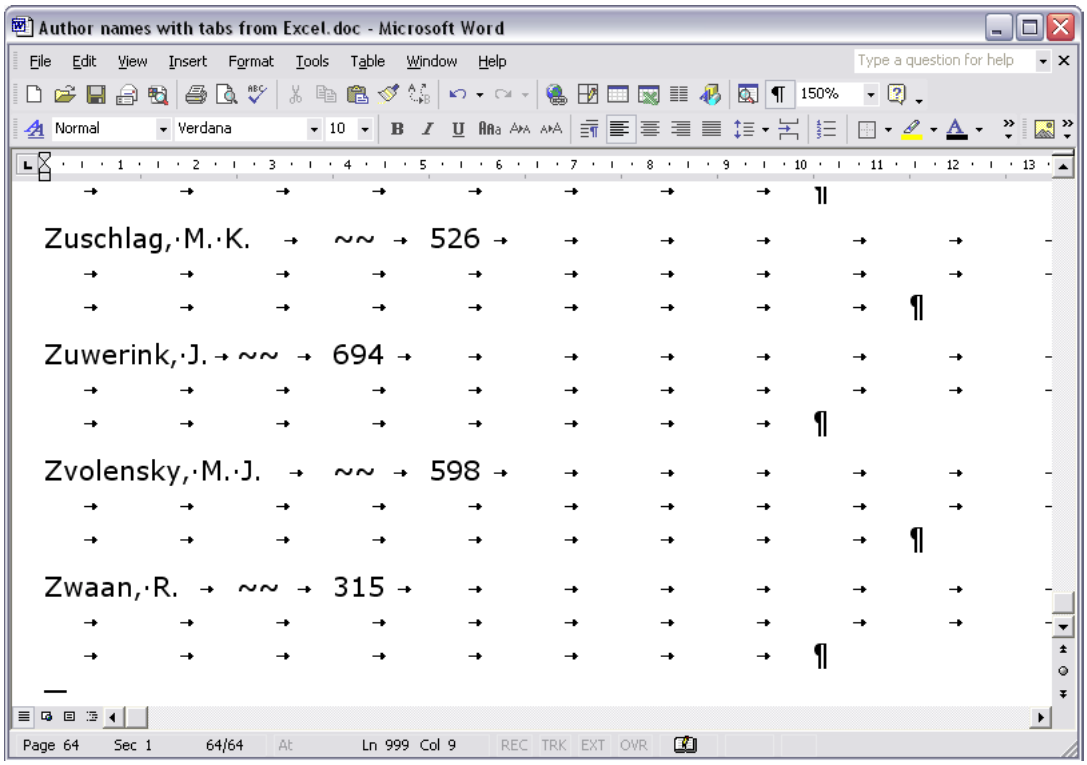


	A	B	C	D
1	**Author**	**Spacer**	**Page1**	**Page2**
2	AARP News Bulletin	~~	469	
3	Aarsland, D.		87	
4	Aartsen, M. J.		486	
5	Abalakina, M.		785	
6	Abascal, L. B.		568	
7	Abbass, A.		645	
8	Abbott, R.		576	
9	Abel, E.		345	
10	Abela, J. R. Z.		598	
11	Abelson, R. B.		60	75
12	Abelson, R. P.		689	
13	Aber, L.		503	
14	Abraham, H. D.		203	
15	Abraham, W. C.		221	

9. Omitting the column headings, select all the data in the spreadsheet.
10. Open Word, and create a new document. (In the **File** menu, click on **New…** It may be necessary to choose **Blank Document** from the **New Document** pane on the right-hand side.)
11. In the **Edit** menu, click on **Paste Special…** Choose **Unformatted text** and click **OK**.



12. The data will be pasted into the new Word document simply as text separated by tabs. Save the new document.



3B. Prepare final document for submission to publisher

Turn tabs into required text

Steps 13–16 below are all executed in the **Find and Replace** dialog box (open the **Edit** menu and select **Replace…**).

13. To delete the tab characters on either side of the unique spacer character:
    Enter the following string in the **Find what** box: ^t~~^t
    Enter the following string in the **Replace with** box: ~~
    Choose **Replace All**. The tab characters on either side of the spacer characters should be deleted.

    Abelson, R. B. ~~ 60 75 76
    becomes
    Abelson, R. B.~~60 75 76

14. To remove most of the tab characters trailing the final page number in each index entry:
    Enter the following string in the **Find what** box: ^t^t
    Enter the following string in the **Replace with** box: ^t
    (This finds pairs of tabs and replaces them with a single tab.)
    Choose **Replace All**. Keep pressing **Replace All** until Word indicates that it “has completed its search of the document and has made 0 replacements”.

15. To delete the last tab character trailing the final page number in each index entry:
    Enter the following string in the **Find what** box: ^t^p
    Enter the following string in the **Replace with** box: ^p
    (This finds a tab character in front of a paragraph mark and deletes it.)
    Choose **Replace All**. All of the trailing tab characters should now be deleted.

16. To turn each remaining tab character into a comma followed by a space:
    Enter the following string in the **Find what** box: ^t
    Enter a comma and a space in the **Replace with** box.
    Choose **Replace All**. For authors followed by more than one page number, those page numbers should now be separated by spaced commas.

    Abelson, R. B.~~60 75 76
    becomes
    Abelson, R. B.~~60, 75, 76

Prepare final document

The list of index entries should now be complete. Check it to make sure nothing has gone awry. Make any changes needed to meet the publisher's style requirements and submit!

Discussion

Outcomes

The finished Name index contained 7,265 names and 10,270 individual page numbers, and took 94 person-hours to complete. That equates to an overall average of three entries per minute, where an entry is either a name or a page number. During this time, as well as making the entries, we also had to set up the software, check the index, and compile lists of references cited in the text but not in the References list and of mismatches of dates and name spellings between the text and the References file. The time of 94 hours represents about 70% of the estimated time to do the job manually, although that estimate was probably rather on the low side.

Requirements for success with this method

- The References list supplied by the editor must be complete and accurate, with individual references consistently formatted, as errors in formatting can throw errors in the search-and-replace functions. In this job, some time had to be spent editing the References list before it could be used.
- Names and dates in references cited in the text must match those in the References list. Consistency of spellings of surnames is essential. If they do not match, they could be missed. The only way to ensure that all references in the text are picked up would be to go through the text manually, checking off references in the list as they are found, as well as recording page numbers. However, this would come at a considerable cost increase.
- References cited in the text must be in the References list, although by chance we found 21 that were not and reported these to the editor. However, we cannot know how many others there were. There were also references in the References list that were not cited in the text. The publisher suggested that this could be due to the adaptation of the book from the original American edition that had itself gone through several editions, as well as this being the second ANZ edition, and that errors like this have just accumulated with each edition and would be difficult to find now.
- When searching for a name in the text document, the name entered in the PDF search box must have the exact form as in the text citation (e.g. 'Smith') or it will not be found. If a comma is left after the name (e.g. 'Smith,'), that reference will be found only if that is the name of the primary author (where there are three or more authors). A 'Smith' or a 'Smith & Jones' reference will not be found and an erroneous conclusion drawn about the number of references with that name as the author or primary author.
- When searching for a name in the spreadsheet, the name entered in the Excel search box must have the exact form as in the reference (e.g. 'Smith, G. F.') or it will not be found. A space before or after the name (easy to miss) or a lack of a space between the initials (if

the name in the spreadsheet has a space) will cause this to happen, and an erroneous conclusion might be drawn about the presence of the reference or author.

- Artefacts introduced during the layout process may make it difficult to find some citations. For example, an author's name may have been hyphenated between one page and the next, or a citation might have been split across pages. Where a citation could not be found, we found it useful to search for just part of it, which sometimes allowed us to pick them up.

Problems inherent to author name indexes in general

Many challenges encountered during this job are inherent to producing author name indexes, rather than being associated with our particular method. Some of these are:

- The name of a single author may be given in different forms in different references (e.g. *Evans, R.* and *Evans, R. J.*). These were shown as separate authors and a note of the inconsistency made for the editor.
- Citations in the text may not be detailed enough to identify the correct reference (e.g. different authors with the same surname cited without initials). An informed judgement could be made where possible, but this is risky and not recommended. The author(s), through the editor, could be asked for the information, but the editor's production schedule may not allow this.
- References may be cited inconsistently between text and reference list (e.g. *Danziger, 1999* in text but *Danziger, 1998* in the References list). The page number was allotted to the author in the Reference list and a note of eight of these was made for the editor.
- Author names may be spelled inconsistently between text and Reference list (e.g. al Ghazi in text but al-Ghazi in references). A case-by-case decision might be needed as to whether to report such errors to the editor (we reported two).

In paragraph 18.116 of The Chicago Manual of Style (University of Chicago Press,15th edition, 2003), it is recommended that "those unfortunates whose names are concealed under *et al.* in text" be not indexed. In the case of this book, there was a number of instances of *et al.* As we found no ambiguity about dates, we felt confident in putting that page number beside the names of the other authors in the reference, too. This also maximised the ability of readers to find citations for co-authors.

Acknowledgements and thanks

Grant Skennerton and Gabrielle Brady, the two members of the editorial staff of John Wiley & Sons Australia who were involved with this project, were extremely supportive and understanding regarding the magnitude of the task faced by the authors. To help with the presentation of this paper John Wiley and Sons Australia have donated a copy of the book, and for this we are extremely grateful.

The authors would also like to thank the Organising Committee of this conference for accepting this paper for presentation. It is hoped that the method described will be of value in various ways to the indexing community.

Metatopic Menace

2009

Kay Schlembach
Managing Partner, Potomac Indexing, LLC
www.potomacindexing.com

introduction

- Who I am and why I care
- Humbled
- Disclaimer
- Remember: no rules, only contexts

© 2009 Kay K. Schlembach. All Rights Reserved

naming the menace

© 2009 Kay K. Schlembach. All Rights Reserved

terminology

- Metatopic(s) – main topic of whole document

- Local main topics – main topic of a particular section or sections

© 2009 Kay K. Schlembach. All Rights Reserved

audience considerations

- Kids
- Search engines
- Usability
- E-books

© 2009 Kay K. Schlembach. All Rights Reserved

indexer considerations

- The metatopic will drive the structural development of the index

- The ever present problem with metatopics is that there is the temptation to index the entire document under a single overarching topic, or ignore the direct indexing of the metatopic altogether.

© 2009 Kay K. Schlembach. All Rights Reserved

taming the monster: sketching

- Linear approach, such as outline

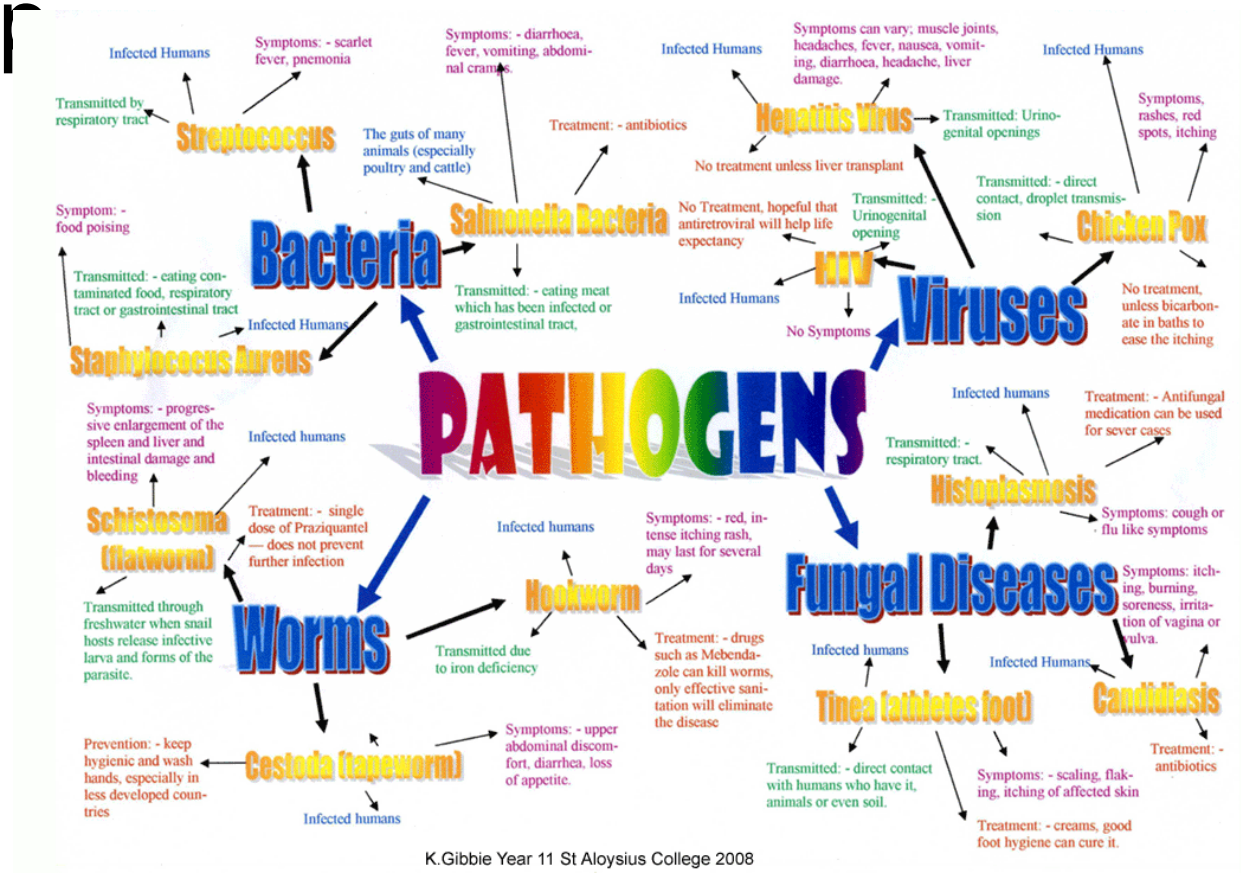
ORIGINS OF AGRICULTURE

- AGRICULTURE COMPARED TO HUNTING-GATHERING
  - Advantages of Agriculture
    1. More efficient use of land
       agriculture: 1 sq km supports 50 people
       hunting-gathering: 25-30 sq km support 5-6 people
    2. More stable food source through year (with storage)
    3. More free time in non-critical seasons
  - Disadvantages of Agriculture
    1. Malnourishment
       farmers often deficient in protein
    2. Labor intensive in critical seasons
    3. High risk if crops/herds fail
- IDENTIFYING DOMESTICATES IN ARCHAEOLOGICAL RECORD
  - Plants
    1. Seeds are bigger in size
       example: teosinte to maize
       example: domesticated sunflower is at least 5 mm
    2. Seed coats are thicker
    3. Found outside natural range of distribution
  - Animals
    1. Horns change in morphology
       example: sheep
    2. Body size changes
    3. Changes in coat or fur
    4. Age-sex distributions in skeletal collections

(www.muskingum.edu/learningstrategiesdatabase.htm)

taming the monster: sketching

- Graphical approach, such as concept map



© 2009 Kay K. Schlembach. All Rights Reserved

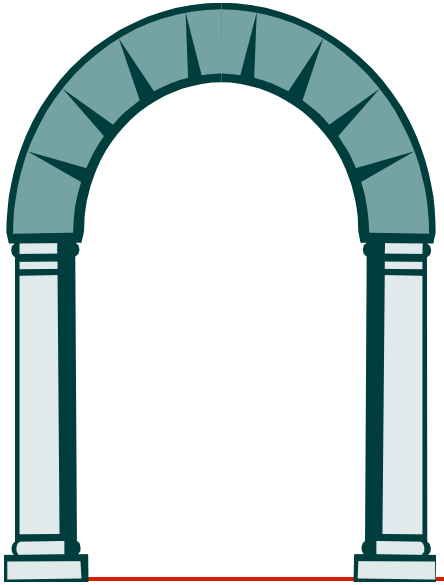
taming the monster: indexing

- Traditional approach
- Neo-Traditional approach
- Table of Contents approach

© 2009 Kay K. Schlembach. All Rights Reserved

traditional approach

- User knowledge assumptions by indexer
- Subheading stand-alone test…



Latin America

misconceptions and stereotypes of, 1–4

nations and territories, 10*m,* 11*t*

reasons to study, 4-7

see also specific nations and regions; specific topics

© 2009 Kay K. Schlembach. All Rights Reserved

table of contents approach

Latin America
agriculture and agrarian development, 252–287
as cultural region, 1–14, 186–216
climate and weather, 49–73
economic development, 381–399
health and nutritional development, 399–423
Iberian heritage, 96–120
landform regions, 16–48
mining, manufacturing and industrial development, 288–325
misconceptions and stereotypes of, 1–4
nations and territories, 10*m,* 11*t*
natural regions, 74–94
political change, 121–150
population and demographics, 346–356
race and ethnicity, 151–185
reasons to study, 4-7
religion, 217–249
urbanization, 328–344, 357–379
see also specific nations and regions; specific topics

© 2009 Kay K. Schlembach. All Rights Reserved

more scholarly example

Capital punishment. *See* Death penalty

Death penalty.

dissonance regarding, 3, 85-89, 137-140, 417

as ritual sacrifice, 422-435

symbolic role of, 15-17

see also capital crimes; executions; *specific state jurisdictions;*

specific topics

Death penalty

alternatives to, 116-134

American's attitudes towards, 79-114

arguments for/against, 403-419

capital crimes, 36-78

capital trials, 319-361

cognitive dissonance regarding, 3, 85-89, 137-140, 417

Constitutional issues, 189-246

deterrence, 136-174

executions, 363-401

international, 22-34

vs. life imprisonment, 176-189

race, 248-274

as ritual sacrifice, 422-435

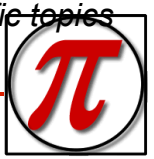
sex (see sex of offender)

social class, 275-317

symbolic role of, 15-17

in United States, 1-20

see also specific state jurisdictions; specific topics



© 2009 Kay K. Schlembach. All Rights Reserved

textbook example

child development

American cultural context for, 14–21, 425, 513, 529–530

critical issues in, 5–11

defined, 4–5

historical view of childhood, 11–14, 446–447

individual differences, attempts to address, 21–24

see also specific modes of development, such as cognitive development;

specific periods of development, such as early childhood; *specific topics*

© 2009 Kay K. Schlembach. All Rights Reserved

textbook example

child development
- conception, 80–84
- prenatal development and pregnancy, 89–107
- childbirth, 107–118
- newborn, 118–124
- infancy and toddlerhood, 129–238
  - cognitive development, 167–200
  - emotional development, 205–213
  - language development, 184–200
  - physical development, 131–163
  - social development, 205–238
- early childhood, 243–337
  - cognitive development, 269–295
  - emotional development, 308–315
  - language development, 282–287
  - physical development, 245–265
  - social development, 299–335

child development (cont)
- late childhood, 339–439
  - cognitive development, 367–399
  - emotional development, 406–418, 436–437
  - language development, 379–383
  - physical development, 341–363
  - social development, 403–439
- adolescence, 443–553
  - cognitive development, 485–509
  - physical development and health, 448–481
  - social and emotional development, 513–551
- American cultural context for, 14–21, 425, 513, 529–530
- critical issues in, 5–11
- defined, 4–5
- genetic influence on, 63–80
- historical view of childhood, 11–14, 446–447
- individual differences, attempts to address, 21-24
- *see also specific topics*

© 2009 Kay K. Schlembach. All Rights Reserved

two distinct, interlaced metatopics

Demographics

- categories for. *See specific demographic categories, such as age, race or sex*
- Census data controversies
- and educational attainment. *See specific criteria for educational attainment, such as high school or college*
- metrics for
- research on

Educational attainment

- criteria for. *See specific criteria for educational attainment, such as high school or college*
- by demographic category. *See specific demographic categories, such as age, race or sex*
- as national policy issue
- research on
- theories on

© 2009 Kay K. Schlembach. All Rights Reserved

two distinct, interlaced metatopics

Demographics
- Census data controversies
- measurement of
- metrics for
- research on

Demographics, categories for
- age
- disability
- geographic location
- home ownership
- household configuration
- income and wealth
- parent occupation and employment status
- race
- sex
- student occupation and employment status

Demographics, and educational attainment
- less than 9th grade
- high school drop-out
- GED
- high school graduate
- some college
- Associate's degree
- Bachelor's degree
- professional degree
- Master's degree
- Doctorate degree

Educational attainment
- as national policy issue
- research on
- theories on

Educational attainment, criteria for
- less than 9th grade
- high school drop-out
- GED
- high school graduate
- some college
- Associate's degree
- Bachelor's degree
- professional degree
- Master's degree
- Doctorate degree

Educational attainment, by demographic category
- age
- disability
- geographic location
- home ownership status
- household configuration
- income and wealth
- parent occupation and employment status
- professional degree
- race
- sex
- student occupation and employment status

© 2009 Kay K. Schlembach. All Rights Reserved

next steps

- Bell, Hazel K. 1998. *Indexing Biographies and Other Stories of Human Lives, 2nd ed.* Society of Indexers, 1998
- Browne, Glenda and Jermey, Jon. 2007. *The Indexing Companion.* Cambridge University Press, 2007
- Mulvany, Nancy C. 2005. *Indexing Books, 2nd ed.* University of Chicago Press, 2005
- Olason, Susan C. 2000. "Let's Get Usable! Usability Studies for Indexes." *The Indexer*, vol 22, no. 2, pp. 91-95, 2000
- Smith, Sherry L. and Kells, Kari J. *Inside Indexing*. Northwest Indexing Press, 2005
- Stauber, Do Mi. *Facing the Text.* Cedar Row Press, 2004

© 2009 Kay K. Schlembach. All Rights Reserved

Trends in indexing

ebooks and tagging

The end of Indexing?

- The end of the trail... no!
- Opening up new areas... yes!

Too much information?

The Petabyte Age, Wired, July 2008, p. 106-107

1 terabyte	A hard drive that holds 260,000 songs
20 terabytes	Photos uploaded to Facebook each month
530 terabytes	All the videos on YouTube
600 terabytes	Ancestry.com's geneology database
1 petabyte	Data processed by Google's servers every 72 minutes

Terabytes and petabytes

- Visualizing this amount of data
- Searching this amount of data
- Indexing this amount of data

Terabytes

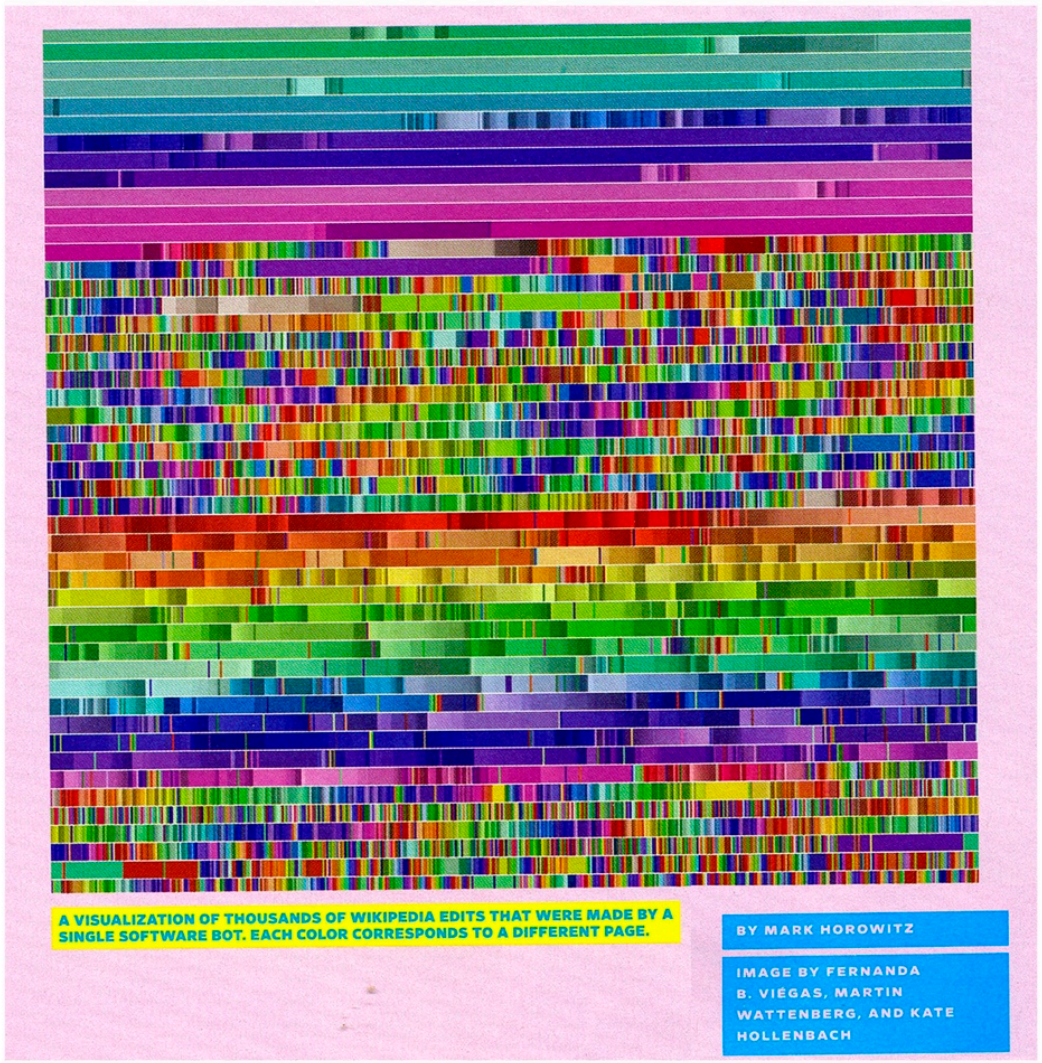
Fernanda Viegas, creator of Many Eyes:

"You can talk about terabytes and exabytes and zettabytes, and at a certain point it all becomes dizzying. The real yardstick to me is how it compares with a natural human limit, like the sum total of all the words you'll hear in your lifetime. That's surely less than a terabyte of text. Any more than that, and it becomes incomprehensible by a single person, so we have to turn to other means of analysis: people working together, or computers, or both"

Wired, "The Petabyte Age," July 2008, p. 121

- We need to think in human terms
- We need to know it is changing every minute.

Data changes every minute



Indexers

- Where do we fit in?
- Will there still be books?
- Will there still be a need for human analysis?

Yes...

- But life is going to be different
- Books, as separate items not connected and updated by the net, need indexes
- Books, as part of the net, still need indexes

Index

[SYMBOL] [**A**] [B] [C] [D] [E] [F] [G] [H] [I] [J] [K] [L] [M] [N] [O] [P] [Q] [R] [S] [T] [U] [V] [W] [X] [Y] [Z]

above line frames [See also inline frames.]
above line objects
Absolute Colorimetric rendering intent
absolute page numbering 2nd
absolute positioning
accessibility 2nd
accuracy of screen color [See color management.]
accursed files
Acrobat [See Adobe Acrobat; PDF graphics and files.]
actions for buttons
activated TOCs and indexes
active publication windows
ActiveWorkspace.xml file
Actual Size 100% view
AD, in text variables
Add Anchor Point tool
Add button (dictionaries)
Add Page button 2nd 3rd
Add Pages To pop-up menu
Add path operation 2nd
Add Point tool 2nd 3rd
Add Tag button 2nd
Add to Autocorrect List dialog box
Add Untagged Items command
AddGuides example script 2nd
AddPoints example script
Adjust Dashes command
Adjust Gaps command 2nd
Adjust Scaling Percentage option 2nd 3rd 4th
Adjust Spacing Automatically... option
Adjust View Settings preferences settings
adjusting layouts 2nd
AdjustLayout example script
Adobe Acrobat
    compatibility of PDF files
    Portable Document Format [See PDF graphics and files.]
    transparency and
Adobe Bridge
    file status
    launching
    metadata from 2nd

Bodies of information

- We will always need subsets or bodies of knowledge indexed with precision

How to think about the situation

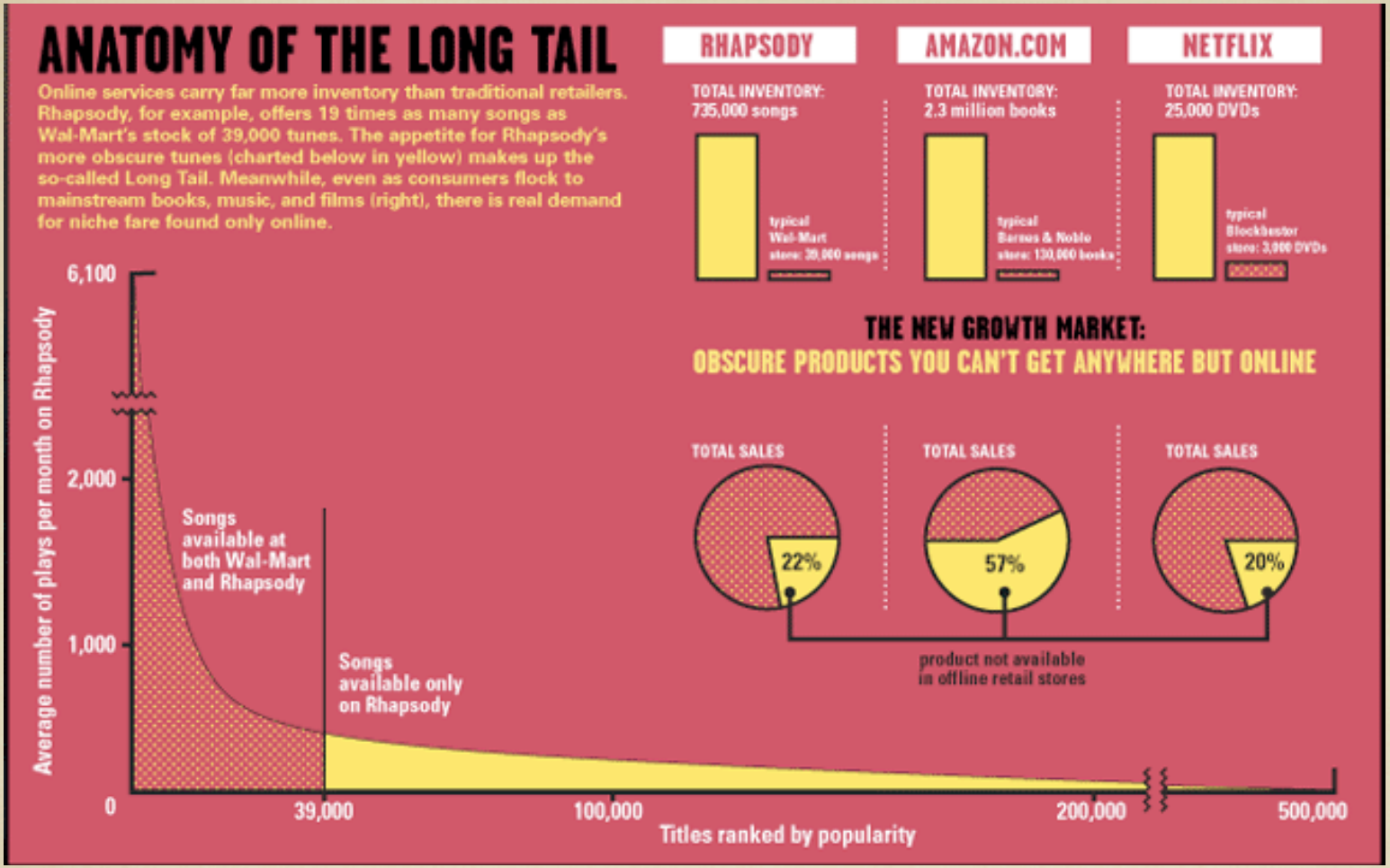
- Books as a separate, and perhaps a "niche" commodity
- Our skills as going beyond books to "contained universes of information"
- Finding out about our competitors, and how we be the ones on the next level up in theory and thinking
- Cleaning up the tagging and personalized indexing online

Books

- Books will always be around
- Books are always something special
- Some, due to rarity or special interests, will become subject to "long-tail" economics
- Some will become free

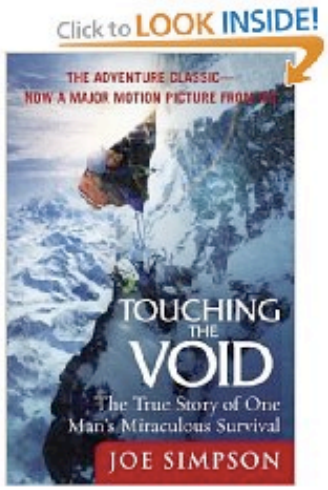
What is this Long-Tail stuff?

- "The niche-based culture catching up to the hit-based culture" (Wired, April 2008, p. 31)
- Sales of bestsellers can be matched by sales of rarer or specialized materials, if you can carry enough of both to make a business



- Wired, "The Long Tail," by Chris Anderson, October, 2004

Touching the void



Share your own customer images

rch inside another edition of this book

Tell the Publisher!
I'd like to read this book on Kindle

Don't have a Kindle? Get yours here.

Touching the Void: The True Story of One Man's Miraculous Survi (Paperback)

by Joe Simpson (Author) "I was lying in my sleeping bag, staring at the light filtering through the red ar dome tent..." (more)

Key Phrases: belay seat, belay plate, boot snagged, West Face, Siula Grande, Bomb Alley (more...)

(155 customer reviews)

List Price: ~~$14.99~~

Price: $10.79 & eligible for **FREE Super Saver Shipping** on orders over $25. D

You Save: $4.20 (28%)

Special Offers Available

In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.

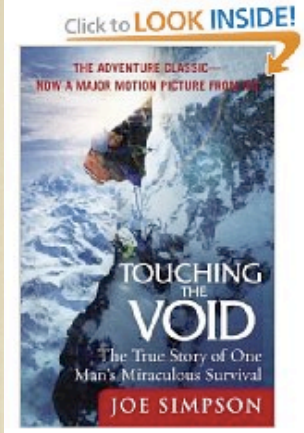
Want it delivered Wednesday, September 30? Order it in the next 1 hour and 35 minutes, and **Shipping** at checkout. Details

37 new from $5.09 **154 used** from $0.32 **3 collectible** from $9.99

Formats	Amazon Price	New from	Used from
Hardcover, December 31, 1988	--	$64.85	$7.43
Paperback, December 31, 2003	$10.79	$5.09	$0.32
Audio, Cassette	--	$20.10	--

Show 11 more formats

Free



Share your own customer images

rch inside another edition of this book

Tell the Publisher!
I'd like to read this book on Kindle

Don't have a Kindle? Get yours here.

Touching the Void: The True Story of One Man's Miraculous Survi (Paperback)

by Joe Simpson (Author) "I was lying in my sleeping bag, staring at the light filtering through the red ar dome tent..." (more)

Key Phrases: belay seat, belay plate, boot snagged, West Face, Siula Grande, Bomb Alley (more...)

(155 customer reviews)

List Price: ~~$14.99~~

Price: $10.79 & eligible for **FREE Super Saver Shipping** on orders over $25. D

You Save: $4.20 (28%)

Special Offers Available

In Stock.

Ships from and sold by **Amazon.com**. Gift-wrap available.

Want it delivered Wednesday, September 30? Order it in the next 1 hour and 35 minutes, and **Shipping** at checkout. Details

37 new from $5.09 **154 used** from $0.32 **3 collectible** from $9.99

Formats	Amazon Price	New from	Used from
Hardcover, December 31, 1988	--	$64.85	$7.43
Paperback, December 31, 2003	$10.79	$5.09	$0.32
Audio, Cassette	--	$20.10	--

Show 11 more formats

Ebooks in the recipe

Electronic books and e-readers could prove a resilient niche that, for Amazon alone, may generate more than $1.2 billion in retail-level sales by 2010 despite potential competition from smartphones

Sony reports sales of 400,000 Sony Readers and 4 million books from its online store

(e-Book Readers May Become More Than A Niche By Amy Gilroy -- TWICE, 3/4/2009)
http://www.twice.com/article/262485-e_Book_Readers_May_Become_More_Than_A_Niche.php

Success in publishing

- Make everything available
- Make it available in many formats
- Cut the price -- oh dear
- Help me find it -- indexes or tags?

Book Publishing Trends

- In theory p… away
- Index… cted to the …
- Boun… xes
- Onl… books (… worlds of information) need indexing or tagging too

The question is, will people still pay money to publish them, and for us to index them?

Book Sales

Book Sales May be Much Greater than Previously Thought

The Book Industry Study Group has issued a report that says book sales have been seriously underreported. The study, "Under the Radar," says that approximately 63,000 publishers with annual sales of less than $50 million generate aggregate sales of $14.2 billion.

Jeff Abraham, executive director of BISG, says that while some of that revenue is represented in current industry sales estimates—which puts revenue at between $23.7 billion and $28.5 billion—a significant portion of the revenue is not. Abraham says the discrepancy between the findings of "Under the Radar" and historical industry measurements is that the study tracked sales from companies whose main business is outside of book publishing. The study also found that the majority of sales made by midsized and smaller publishers are made outside of traditional bookselling channels such as non-book wholesalers who serve outlets such as sporting goods stores.

-- By Jim Milliot, Publishers Weekly. http://www.PublishersWeekly.com as reported on www.bookstatistics.com

BOOK SALES, NUMBER OF

Executive summary:-)

Company	2008 FY North American Sales (Books, Media plus cofee, etc:-)
Barnes&Noble / B. Dalton[1]	$4.52 billion (Down from $4.68 billion in 2007)
Borders / Waldenbooks	$3.11 billion (excludes international)
Amazon Media (excludes electronics, services - books, includes books, music, DVDs)	$5.35 billion (Amazon International stores sold $5.73 billion in 2008, more than North America) for the first time.
BN.com	$466 million (Down from $477 million in 2007)
Total	$13.7 billion

HTTP://WWW.FONERBOOKS.COM/BOOKSALE.HTM

AAP results

http://www.publishers.org/main/PressCenter/Archicves/2009_September/PublishingSalesJuly09.htm

US Publishers Net Sales 2007	July sales	
Adult Hardbound	$88.7 million	6.9% increase in July; -15.5 for year
Adult Paperback	$124 million	9.0% increase in July; - 11.2 for year
Audio books	$11.7 million	3.5% increase in July; -29.9 for year
E-books	$16.2 million	213.5% increase in July; 173.9% for year

Australian book sales

- BookScan:

  2008's $1.214 billion book sales represented a 2 per cent increase on the previous year

http://www.theage.com.au/cgi-bin/common/popupPrintArticle.pl?path=/articles/2009/06/20/1244918231827.html

Australian publishing

2008 $2 Billion	Output up 12%	Publishing companies = 4,000

http://www.booksellerandpublisher.com.au/articles/2009/09/13413/

New Zealand Publishing

- In 2008, participating Publishers turned over $209,659,314
- In 2008, participating Publishers published 1,986 new New Zealand titles
- Reference and Education are the biggest publishing categories
- Publishers claim the industry is growing

--- 35% of publishers say their business is growing

--- While 12% say it's declining

http://bpanz.org.nz/wp-content/uploads/2008/10/2008-bpanz-colmar-brunton-book-publishers-report.pdf

Who is buying what?

- 57% of book buyers are women yet women purchase 65% of the books sold in the U.S.
- Mystery books are the most popular genre for book club sales, with 17% of all purchases of mystery books coming directly from book clubs
- Generation X consumers buy more books online than any other demographic group, with 30% of them buying their books through the Internet
- 21% of book buyers said they became aware of a book through some sort of online promotion or ad
- Women made the majority of the purchases in the paperback, hardcover and audio-book segments, but men accounted for 55% of e-book purchases

http://www.bowker.com/index.php/press-releases/567-bowker-publishes-first-consumer-focused-research-report-for-book-industry

Bowker's 2008 U.S. Book Consumer Demographics and Buying Behaviors Annual Report

Latest ebook stats

E-book sales statistics for July 2009 have been released by the Association of American Publishers (AAP) via IDPF. Trade e-books sales were $16.2 million for July 2009, a 213.5% increase over July 2008. IDPF reports calendar year to date revenue is up 173.9% for the year.

http://thecite.blogspot.com/2009/09/july-2009-e-book-sales-statistics.html

Antipodean eBooks

- Kindle is launching its international edition. You can place an order for it now, as of Oct. 7.
- LexCycle, the company behind Stanza, said it had 55,000 Stanza iPhone/iPod Touch users in Australia and New Zealand. --
- Don Grover, chief executive of the Dymocks Group, says that his e-book website sales exceeded his two-year projections in its first six months, with current sales topping 2000 a month.
- Central Book Services has set up its own e-book website, carrying 70,000 international titles, and has launched a new, cheaper e-reader, the Eco-reader, costing less than $500 and linked to the website. The Eco Reader can handle an impressive number of formats.
- Fictionwise, one main supplier for the Stanza app on Iphones, has problems distributing outside of the US and Canada.

http://www.booksellerandpublisher.com.au/articles/2009/07/12471/
http://www.theage.com.au/cgi-bin/common/popupPrintArticle.pl?path=/articles/2009/06/20/1244918231827.html
http://www.theage.com.au/cgi-bin/common/popupPrintArticle.pl?path=/articles/2009/06/20/1244918231827.html
http://jkontherun.com/2009/04/21/fictionwise-e-book-sales-restricted-outside-us-the-real-story/

Kindle book sales

- For books available in both formats (275,000 titles including nearly all the bestsellers) Kindle units account for 35% of unit sales.
- Josh Bernoff, an author, says that sales in Kindle format for his book in 2009 equal all of his sales in 2008.
- Sales figures for the Kindle are hard to get, but the Forrester blog is estimating that 800,000 have been sold. TechCrunch says that 400,000 Kindle 1's were sold, and 300,000 Kindle 2's were sold as of April.
- The New York Times reports that eBooks sales are really 1.6 percent of total book sales

http://blogs.forrester.com/groundswell/2009/05/an-authors-perspective-on-the-surge-in-kindle-sales-35-of-units.html
http://www.techcrunch.com/2009/04/16/300000-kindle-2s-sold-to-date/

What are Kindle books?

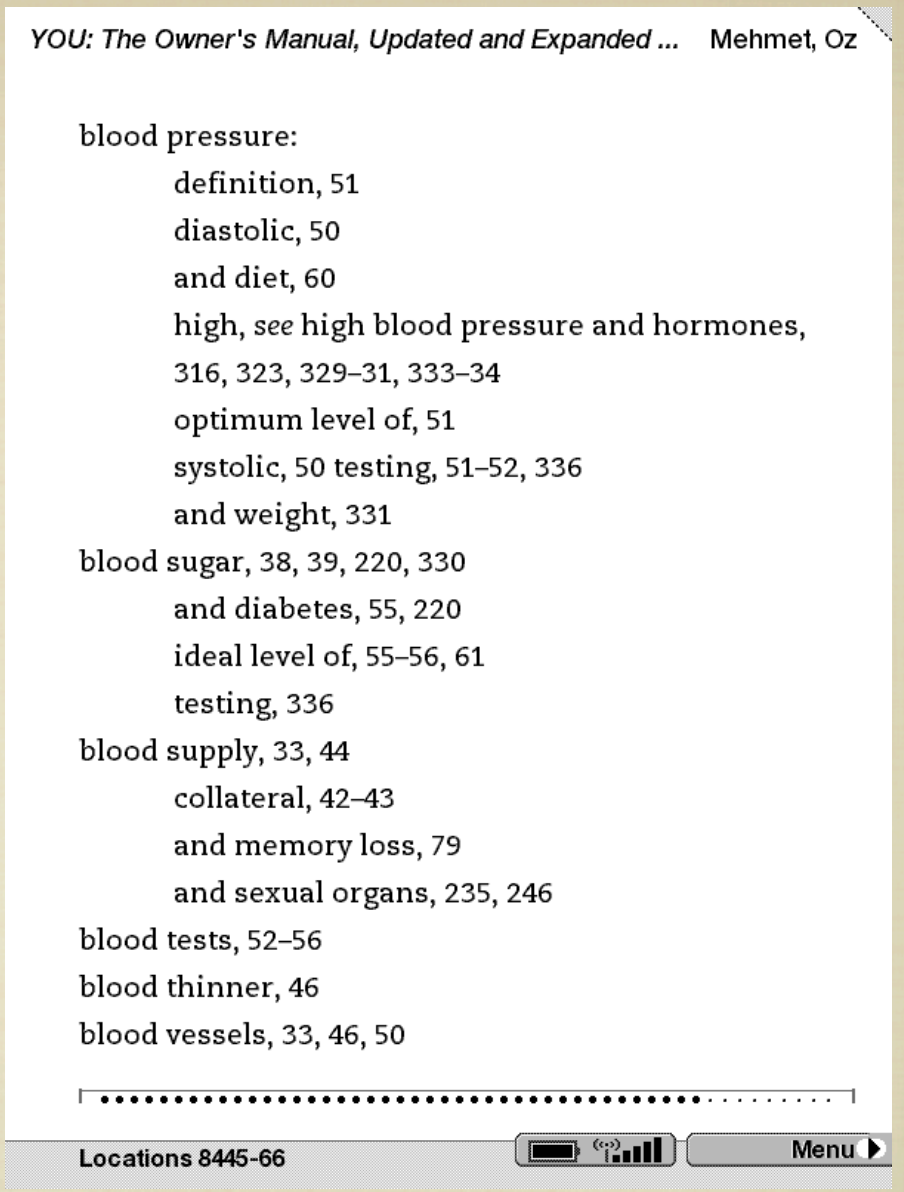
- Kindle books are published in the AZW, Epub and Mobi standards
- These standards allow publishers to produce and send a single digital publication file through distribution and offers consumers interoperability between software/hardware for unencrypted reflowable digital books and other publications.

Why indexes matter in ebooks

- Most ebook indexes are inactive
- You must use the search, and it searches all of your downloaded books
- This is not good, as Jan Zlendich reports

http://kindlereader.blogspot.com/2008/07/grammar-girl-redux.html#links

A non-working index is nearly useless



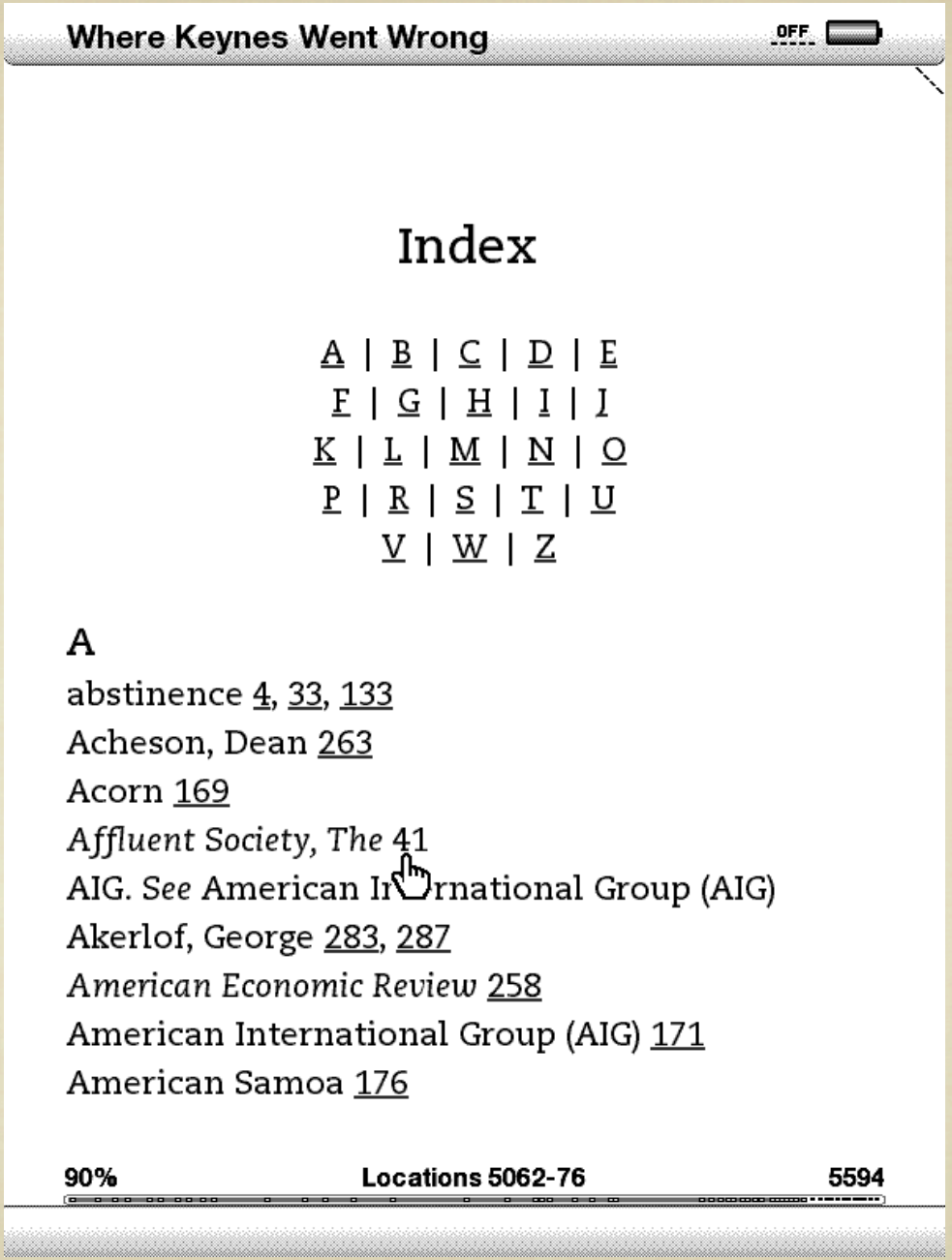
blood pressure:
definition, 51
diastolic, 50
and diet, 60
high, *see* high blood pressure and hormones, 316, 323, 329–31, 333–34
optimum level of, 51
systolic, 50 testing, 51–52, 336
and weight, 331
blood sugar, 38, 39, 220, 330
and diabetes, 55, 220
ideal level of, 55–56, 61
testing, 336
blood supply, 33, 44
collateral, 42–43
and memory loss, 79
and sexual organs, 235, 246
blood tests, 52–56
blood thinner, 46
blood vessels, 33, 46, 50

Kindle Indexing

- Questions about process
- Joshua Tallent of www.kindleformatting.com

- Joshua Tallent, www.kindleformatting.com

- Joshua Tallent,
- www.kindle formatting.com



> the Religion of the Future should concern itself with... [public] needs... with public baths, play grounds, wider and cleaner streets and better dwellings.[28]

Harvard (and Keynesian) economist, John Kenneth Galbraith, seconded this thought in books like *The Affluent Society* and *The New Industrial State*. In Galbraith's version of Keynesianism, society needed to stop saving and spend more, but on public goods, not on tasteless, mass produced, mass advertised consumer goods.

Would Keynes have backed this position? Perhaps in part. He might have agreed that society has a greater need for public goods than consumer goods. But the whole emphasis on goods, public or private, would have struck him as unduly materialistic, part of what he called the

> Benthamite Calculus,

Embedded indexing for Kindle?

- Sadly, no
- { XE “index entry:subhead” }
- {{{index entry:subhead}}}<a name=anchorwithuniqueid>

Make new friends... and keep the old

- Programmers, coders, macro writers
- Learn CUP processes, embedded processes, DTP programs, ebook standards
- Create interactive chains of indexers who know each other's specialties.
- We can't afford to buy all of the products -- we need to cooperate to keep our own costs down
- Interactive scheduling -- filling in the cracks -- subcontracting -- intelligent partnerships

The other trend - Help me find it!

- Traditional indexes
- Online indexes
- Search
- Controlled vocabularies
- Taxonomies
- Tagging

CT's and Taxo's

- Controlled vocabularies
- Taxonomies/thesauri
- Tagging

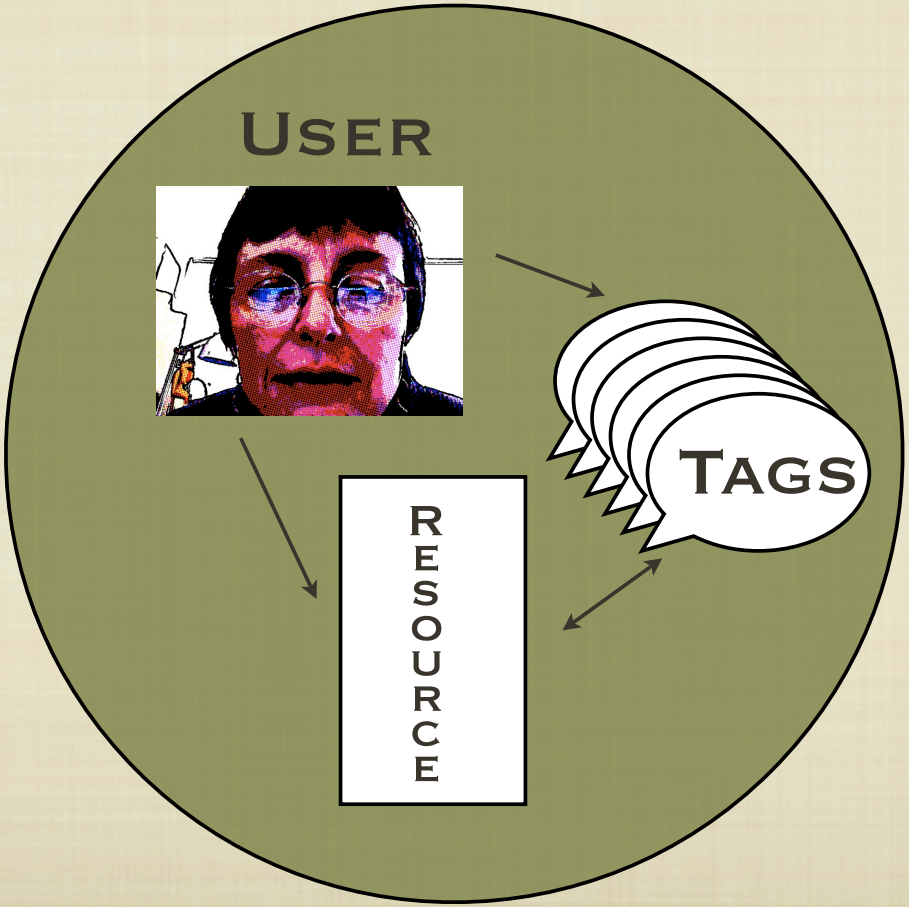
Maniacs

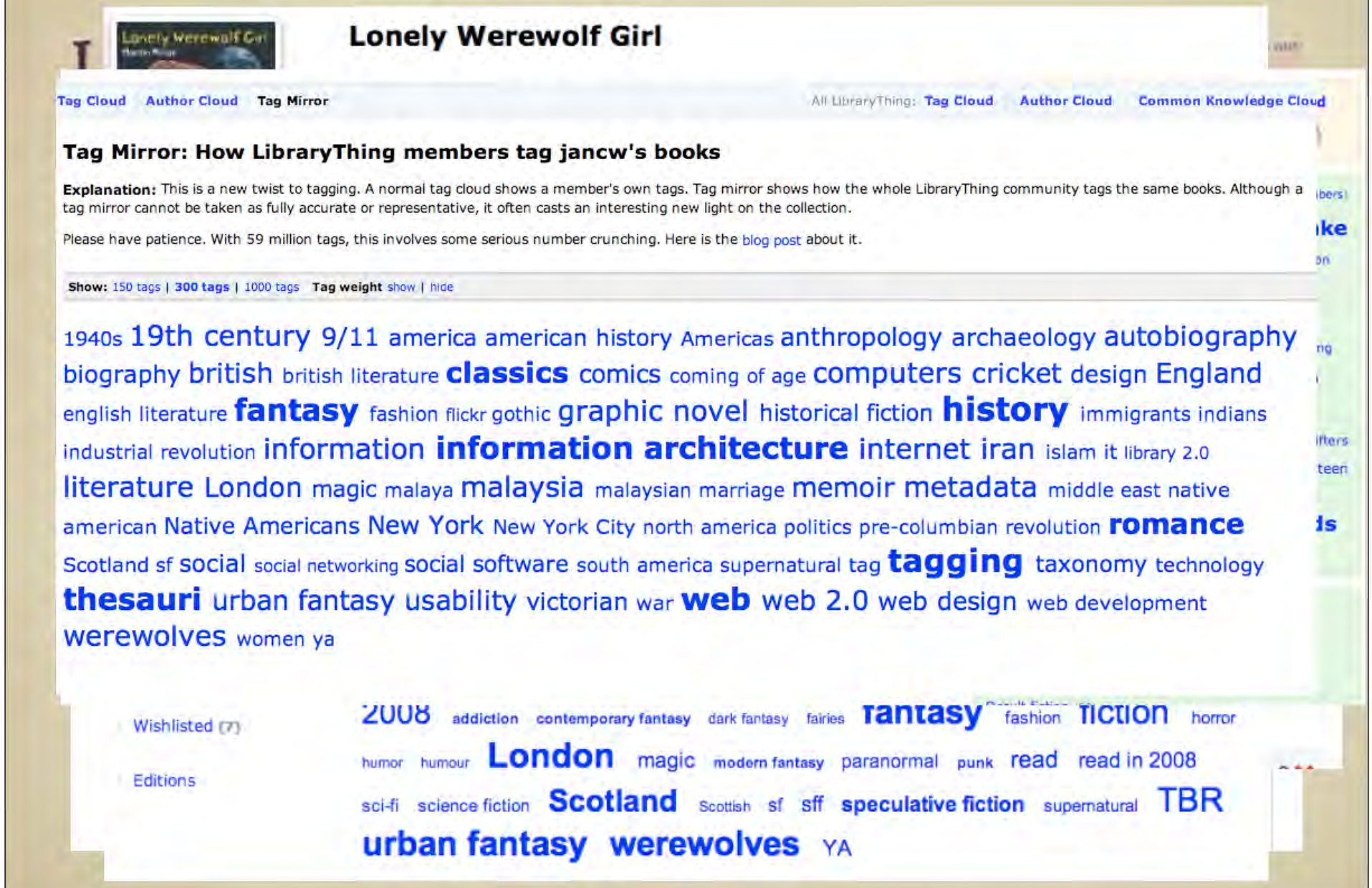
- Everyone is tagging
- On their own free time
- Using whatever they want as terms
- And going into depth in their interest areas

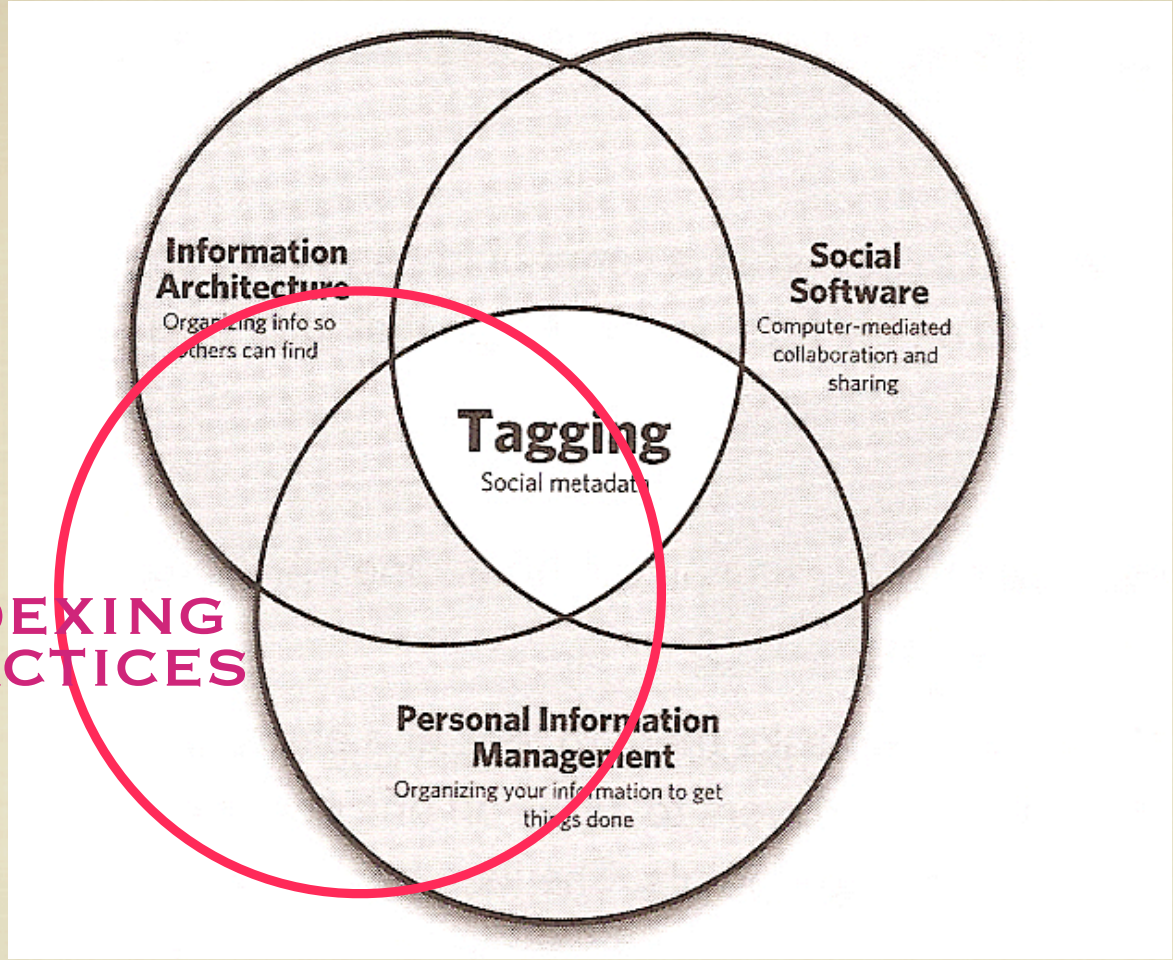
Tagging

- indexing applied to items, documents, articles
- Can be controlled in-house
- Can be public (users, readers, viewers)
- Can be precise (controlled vocabularies)
- Can be chaos (everyone makes up their own)
- Can be missing (no enthusiasm or rewards)

Tagging process




Lonely Werewolf Girl
Tag Cloud
Author Cloud
Tag Mirror
All LibraryThing:
Tag Cloud
Author Cloud
Common Knowledge Cloud
Tag Mirror: How LibraryThing members tag jancw's books
Explanation: This is a new twist to tagging. A normal tag cloud shows a member's own tags. Tag mirror shows how the whole LibraryThing community tags the same books. Although a tag mirror cannot be taken as fully accurate or representative, it often casts an interesting new light on the collection.
Please have patience. With 59 million tags, this involves some serious number crunching. Here is the blog post about it.
Show: 150 tags | 300 tags | 1000 tags
Tag weight show | hide
1940s 19th century 9/11 america american history Americas anthropology archaeology autobiography biography british british literature classics comics coming of age computers cricket design England english literature fantasy fashion flickr gothic graphic novel historical fiction history immigrants indians industrial revolution information information architecture internet iran islam it library 2.0 literature London magic malaya malaysia malaysian marriage memoir metadata middle east native american Native Americans New York New York City north america politics pre-columbian revolution romance Scotland sf social social networking social software south america supernatural tag tagging taxonomy technology thesauri urban fantasy usability victorian war web web 2.0 web design web development werewolves women ya
Wishlisted (7)
Editions
2008 addiction contemporary fantasy dark fantasy fairies fantasy fashion fiction horror humor humour London magic modern fantasy paranormal punk read read in 2008 sci-fi science fiction Scotland Scottish sf sff speculative fiction supernatural TBR urban fantasy werewolves YA



Indexing Practices

Gene Smith, Tagging: People-Powered Metadata for the Social Web, New Riders, 2008
p. 12

Everyone's learning to index

- Pew Internet and American Life Project, January 2007
- One third of U.S. Internet users (42 million Americans) had tagged some form of online content.
- 10 million Americans (7% of Internet users) are tagging content daily.

Smith, Tagging, p. 18

Antipodean tagging?

- Various research reports have indicated that at least 50% of Australian internet users participate in social networks.
- Forrester's Groundswell profiling tool suggests that 23% of Australians aged 18+ actively create content online; 31% are 'critics', providing comments and feedback online; and 50% are 'joiners' of social networks, forums and online groups.
- Australia has 16,926,015 Internet users as of Feb./2009, 79.6% of the population. If 7% of them are tagging, as they are in the US, you have 1,184,821 people tagging content online.
- New Zealand has 3,360,000 Internet users as of June 2009. If 7% of them are tagging, you have 235,200 people tagging content online.

http://governingpeople.com/Home/17847 -- Craig Thomler
http://www.internetworldstats.com/sp/au.htm
http://www.docstoc.com/docs/10350828/Internet-Usage-and-Population-in-Oceania-—-June-2009

More tagging numbers

- Delicious, YouTube and Flickr - In a study covering 2005-2007, combined data from the three social networks totals 1.8 million objects to tag, 3.1 million taggers and 12.1 million tags, of which 648,368 tags are unique.
- Per object, 2.74 tags (YouTube) to 9.31 (Delicious)
- Tags one tagger assigns, 3.33 (Delicious) to 8.79 (Flickr)
- Objects tagged per user, 0.36 (Delicious) to 2.84 (YouTube)

http://www.dlib.org/dlib/march09/ding/03ding.html

Inherent tag tensions

- Personal <--> Social
- Idiosyncratic <--> Standard
- Freedom <--> Control
- Amateur <--> Expert

Smith, Tagging, p. 15

Money & maniacs

- Publishers must pay for precise indexing
- Tagging is free, but you must motivate it
- Controlled vocabularies control tagging
- Tagging takes time
- Much tagged data may wind up as ephemeral

The results of maniacs

Dander

From Wikipedia, the free encyclopedia

Dander is material shed from the body of various animals, similar to dandruff or pet pollen. It may contain scales of dried skin and hair, or feathers. It is a frequent cause of allergy in humans.

Dander is also a Northern Irish term to describe taking a relaxed walk, with no particular purpose or haste.

- Powder down

Importance to indexers

- Important field
- building controlled vocabularies
- Building taxonomies
- gathering terminology
- Bringing standardization

And lastly...

A big thank-you to ANZSI...

Jan C. Wright
Wright Information Indexing Services
www.wrightinformation.com
jancw@wrightinformation.com