School of Information Sciences -- Faculty Publications and Other Works

School of Information Sciences

6-1992

A Day in the Life of a Database Producer.

Carol Tenopir
University of Tennessee - Knoxville

Recommended Citation

A day in the life of a database producer

Abstract:
A typical day at a database production company is presented. The descriptions are based on visits to H.W. Wilson Co, Information Access Co, and UMI/Data Courier. The process of database creation and production is explored, from the initial decisions regarding database content through the continuing processes of ordering and receipt of titles, indexing/abstracting, editing and quality control. Brief discussions are also included regarding production quotas and the use of home workers.

Full Text:
Five o'clock. Monday Morning. The first employee arrives in the Information Access Company offices in Foster City, California. The weekend and Monday a.m. editions of several newspapers are already waiting for her to check them in through the serials check-in system. She hurries to get them processed before the indexers start arriving at 7 a.m. Soon the parking lot will be full and the mail deliveries of magazines and newspapers will begin in earnest.

So begins a day in the life of one database producer. This one happens to be Information Access Company, but all of the major indexing/abstracting companies follow much the same pattern. Whether it is a Monday morning in a complex of five 2-story buildings near the San Francisco Bay, an eight-story fortress in the Bronx, or a converted bank building in downtown Louisville, serial checkers, indexers, inputters, editors, and proofreaders together process thousands of periodical articles every day. Their finished products are the dozens of online, CD-ROM and/or print indexes and abstracts that librarians use daily.

A TYPICAL DAY

Recently I had the opportunity to visit several major database producers to meet the people on the lines who do the intellectual and physical creating of many familiar information products. I talked to employees at all stages of the creation and distribution process, followed them around, observed them working, and even tried out some indexing myself. The H.W. Wilson Company in New York, Information Access Company (IAC) in California, and UMI/Data Courier (UMI/DC) in Louisville, Kentucky and Ann Arbor, Michigan produce many of the finding tools that are so essential in libraries. Although their products are different, and are often in competition, these companies share much in common. The "Day" I describe here is a composite of my experiences at all three companies.

COLLECTION DEVELOPMENT DECISIONS

To get a feel for how a bibliographic database is created on a day-in and day-out basis, it helps to follow the process from the beginning. Just like in a library, collection development decisions dictate what titles will be included in the "collection," or in this case, indexed in the database. For ten of the H.W. Wilson Company indexes these decisions are helped by the Committee on Wilson Indexes of the American Library Association's Reference and Adult Services Division. The information they gather helps Wilson put together a list of periodicals to be considered for addition to each database. Subscribers are then asked to vote on each title placing "primary emphasis on the reference value of the periodicals under consideration" and to balance subject coverage in the index. Other companies base their decisions on requests from customers, close contacts with professional library groups, what the competition is covering, and publishing trends in a subject.

SERIALS ORDERING

When the decision is made to add a new title to an index, an order goes to their acquisitions department. Jobbers such as EBSCO and Faxon serve database producers as well as libraries, while some titles must be purchased directly from the publisher. Just like in libraries, database producers usually must pay for their subscriptions.

If a periodical database includes over 500 magazine titles, that means over 500 subscriptions for that database alone. It means well over 500 copies, however, because the indexing company often buys more than one copy of many titles. If they own microfilming

rights, a second copy is ordered just for filming. If they own full-text electronic rights, a third copy for scanning may be purchased. The Indexing copy is just that; it goes to an indexer and doesn't get passed on to microfilming or scanning.

SERIALS RECEIVING

Although new titles are constantly being considered by all the database producers, the real daily routine begins with serials receiving. It is in the serials receiving department that the physical act of creating an index begins.

A serials receiving department looks a lot like a university's serials department. They are responsible for subscribing to, checking-in, distributing, claiming, and preparing for storage all of the periodicals indexed. In companies that create multiple index products the serials department is always busy. IAC, for example, receives approximately 300 journals every day. (Monday is the busiest day, Friday is lightest.) They have 5000 current subscriptions total, plus direct electronic feeds from many newswire services. Wilson receives and indexes almost 4000 periodicals each year (in addition to cataloging over 70,000 books).

Card check-in files, or more commonly these days, terminals to access an automated check-in system, are the heart of the serials departments. Each periodical check-in record tells what indexes include this title, whether it is a priority rush title, and how many copies should be received. Priority items are determined by the editorial staff of each index and include titles such as Newsweek or Business Week that are especially timely or heavily used in libraries. Daily newspapers are always considered rush and are indexed the day they are received at UMI/DC and IAC.

Claiming is especially important for companies that try to index materials in a timely manner. If a periodical hasn't arrived when it is expected, the serials department often contacts the publishers by phone. They then, according to an editorial supervisor, "do whatever it takes to obtain a copy of the journal. In our kind of business we simply can't tolerate a gap in a journal. We have to have every issue of a journal. If a publisher can't supply it we go through other avenues and we do whatever we have to, to make sure we have every copy of every journal we will be indexing for our products."

Most of the database producers have developed (or are developing) their own automated serials check-in systems. At Wilson it is a part of their company-wide automation system, installed in the early 1980s.

Once a magazine is checked in, a label is put on the cover of the journal with codes for what indexes it goes in and a title/issue tracking number. Issues are then distributed to the appropriate indexing department.

WHO DOES THE INDEXING?

When the check-in and distribution processes are complete, an issue is ready to be indexed. All three companies have large indexing staffs. Wilson is the largest, with 100 indexers, 37 editors, 20 catalogers, 37 abstractors, and 78 clerical assistants. Wilson requires its indexers, editors, and catalogers to have degrees in library science and preferably also a subject specialty. In other companies, the indexers may be subject specialists if the index is aimed at a speciality audience. Indexers for Legal Resource Index, for example, are paralegals, have law degrees, or have work experience in the legal field. For general interest or less technical titles, the indexers are typically either librarians or liberal arts majors. The main ingredient for success is an interest in a wide variety of subjects and satisfaction from learning about new things. The ability to concentrate, work alone, and pay attention to details are also necessary personality traits of a good indexer.

Indexers go through extensive training in the procedures of indexing, the computer system they will use, and the philosophy of the company. Wilson's training period lasts between six months and one year, depending on the individual. At IAC and UMI/DC new indexers may have had no experience or education in indexing and may have never even used the indexes. Training is therefore intense and closely supervised. At IAC the training period may last from two weeks to three months, but averages six to eight weeks. At UMI/DC, indexers are considered on probation for three months. During this time they are indexing, but at their own pace and with extensive feedback. According to supervisors at these companies, "Learning by practicing and doing it and getting comments on your work is the best way to learn how to index."

Since January 1991 UMI/DC's personnel office has administered a battery of standard office skills tests to potential indexers. They feel these tests of grammar, spelling, proofreading, and following directions are effective as a first cut in identifying people who will make good indexers.

Fat printed style manuals that cover all aspects of indexing and abstracting for each product are in constant use by new indexers at all three companies. The style manuals contain special instructions for each index, plus indexing decisions and hints for each journal title or each type of article.

Indexers typically are organized into departments based on the index they create. Within an index department they may be further separated into groups by periodical titles, subjects, regions (for newspapers) or experience. Each indexer at Wilson has an assigned list of titles in his field. The indexer is responsible for these titles, which Wilson believes helps to ensure consistency of indexing from one issue to the next.

At IAC, an indexing workgroup consists of eight to ten people; each group has a list of magazines for which they are responsible. The group may be assigned only trade magazines or only trade journals, journals specific to particular industries, only academic journals, or journals on a particular topic. A group supervisor assigns the work, but usually individuals will specialize in certain journal titles that they get to know quite well. This helps them to move quickly through a title.

Though an indexer may have responsibility for certain titles, each individual gets a variety so indexers are able to do different magazines. Within each group there are a number of individuals able to handle any subject or title. Priority items are always done first by all indexers in a group.

SELECTING ARTICLES FOR INDEXING

When an indexer takes an issue off the shelf, the first task is to review it for "indexable" articles and do the initial stages of indexing. Indexers choose which articles will be indexed based on a predetermined set of criteria that may be different for every index. General instructions at IAC say "Index every significant article," which is defined as one-half page in length or more in most cases. According to an IAC indexing supervisor, "certain categories of information don't lend themselves to indexing, so we omit them. These include, for example, letters-to-editors in most journals and columns that contain categories of random information without focusing on a single topic." ABI/INFORM includes only articles of lasting value and of at least one page in length (unless the full text is online). The required 150 to 200 word abstract is difficult to write if the article is too short. Periodical Abstracts takes reviews of any length, but generally requires one-third page or more for other types of articles. Almost all Wilson indexes include every article in every title, excluding only letters to the editor and advertisements.

THE INDEXING PROCESS

Once articles are selected, they are indexed. In a fully automated environment, the first steps of indexing are to enter basic journal and issue information into the system. In a partially manual system or when tasks of indexers and inputters are split, indexers may still fill out coding sheets or mark magazine pages for clerical inputters.

Then the intellectual work of indexing begins. The indexer checks the article title to see if it is indicative of content. If the article title is unclear, the indexer may decide to annotate the title. Depending on the index, the indexers may also add information about illustrations, captions, geographic codes, Standard Industrial Classification Codes, etc. The most important part of an indexer's job, however, is to assign the subject headings.

All of these companies maintain extensive lists of subject headings. Wilson maintains a subject authority file for every one of its indexes. Keeping these lists up-to-date by adding and deleting terms is a job undertaken by experienced people in the indexing departments. Wilson prides itself on its vocabulary, 50% of which are terms formulated in-house. IAC bases their vocabulary for most of their indexes on the Library of Congress Subject Headings, although they often must expand beyond LCSH. Approximately 75% of IAC's topical vocabulary is LC; another 25% is added by IAC indexers. For example, earlier this year IAC created a new heading for Bungee Jumping. All of these additions and changes are communicated to the Library of Congress.

Indexers quickly read or scan an article, looking for the main concepts. If it is a particularly technical or difficult article, they may have to read it completely. Experienced indexers know much of their indexes' controlled vocabulary in their heads--they can often assign appropriate terms without checking the subject authority file. For a new subject or one they have not encountered often, they must check either an online or printed version of the subject authority.

Each of these companies also maintains extensive personal and company name authority files. Every name that goes into Readers' Guide, or Periodical Abstracts, or Magazine Index (or any of the other indexes) is checked against an authority file. These files require full time staff to maintain, since every new term that gets added is verified in some external source, such as a directory. Indexing companies maintain collections of standard reference books just like a library (and may ask local librarians for help if they cannot verify a particular name or term).

ABSTRACTING

Abstracts may be written by the same person who does the indexing or by a separate abstractor. H.W. Wilson has the process completely separate; the indexers and abstractors are even in different states! Indexing for Readers' Guide takes place in the corporate headquarters in the Bronx as it always has, but abstracts that go into the Readers' Guide Abstracts version are done by professional abstractors in Cambridge, Massachusetts. That is why there may be a gap between the time indexing gets done and is loaded into the system and the time the abstract is added, although, since they began down-loading the indexed records at Cambridge, distance hasn't been as much of a problem.

For ABI/INFORM (UMI/DC's business database that includes lengthy informative abstracts) indexing and abstracting are done by different people because "they are two different skills; a person can be good at one and not the other." Abstractors must be good writers and, for technical subjects, may also need to be subject specialists.

In UMI/DC's general interest products, on the other hand, the selecting, indexing, and abstracting are all done by each indexer. The indexers have college degrees, typically in English, journalism, or a related subject.

DOUBLE CHECKING

Checks and double checks are a part of the creation process for every index. If the system is fully automated, such as UMI/DC's system for Periodical Abstracts and Newspaper Abstracts, all controlled term fields are verified as soon as an indexer completes an index record. The system makes sure each heading assigned is found in the controlled vocabulary and suggests "near matches" if something minor such as a singular/plural or word ending is different. An indexer may choose to keep an unmatched heading if it is a new concept not represented by another term in the vocabulary or is a new company name with no near match. These new headings are printed out at the end of each day and reviewed by a special editorial committee who accepts them as new valid terms or

changes the indexing to match existing vocabulary.

In addition to automatic checks, each person's indexing is checked by other indexers, or by an index editor, or, in some cases, by both. H.W. Wilson Company procedures are an example of how much time and effort go into quality control. Indexers check authority files as they index, but they do not input the records. As clerical staff input each record, the system automatically performs four kinds of validation: "format validation, which checks for specific data elements and detects typing errors; content validation, which also verifies whether subject headings have been established for "see also" references exist for "see" references; near-miss checking, which alerts the staff that a close match exists for a term in a file; duplicate checking, which determines whether a record has been previously entered."

A revisor at Wilson then manually checks each record for accuracy and spelling. Finally, an editor or associate editor checks all records that make up a print issue or update and the editor reads all paper issues and the annual.

With other indexes, the necessity for speed dictates fewer checks. For Periodical Abstracts, for example, each record is reviewed by other indexers. Indexers spend approximately six hours per day indexing/abstracting and two hours editing their colleagues' work. they make the corrections immediately online. Unless one of the headings appears on an editor's new heading list, this is the last time the record is reviewed.

In most indexes the editing/review process is not just a matter of checking for input errors or verifying spelling. Reviewers may check for accuracy of the bibliographic information and accuracy or appropriateness of the indexing. Each reviewer asks himself if the indexer understood the main focus of article and did he include subject headings that will help users find the article? Did he follow proper editorial procedures for that index, for that magazine title, or for that type of article? Is an abstract complete, accurate, well written, and in compliance with all guidelines in the style manual? Not until this review, editing, and changing process is complete are the records actually added to the database.

HOME INDEXERS

Although small database producers may hire out large portions of their indexing on a free-lance basis, most of these major companies rely on staff indexers. This does not mean that all staff indexers have to be in the office every day. IAC, for example, has a home indexing option for experienced employees who prefer to work out of their homes. Each home indexer is provided with a micro version of the indexing software, a personal computer, and the authority files, which are updated monthly on a CD-ROM. They work out of their homes, but come in at least once a week to drop off and pick up work and to attend meetings. The company believes it is an "excellent way to retain employees who have small children, wish to live at a distance and don't want to do the daily commute, or want to go back to school." IAC has over 50 home indexers on the staff, each of whom worked in-house for at least one year. There are even a couple of employees who do home indexing from out of state. They are sent magazines by courier service and they return a disk of indexing plus the magazines by courier service. The editors keep in touch by phone.

UMI/DC has over 18 contractors that do part of the abstracting for ABI/INFORM, in addition to the ten in-house, full-time abstractors. Each contractor does about 25 abstracts each week. ABI/INFORM is "gradually going to an in-house staff," however, because the editors believe "the quality of the abstracting done in-house is better." The most popular and the core titles are all done in-house.

Wilson agrees that in-house indexers mean better quality. In-house indexers can share ideas, have access to all the files and all the resources, and get immediate answers to their questions. According to a Wilson spokesman, "we have experimented with this and find we do attain better quality if all indexers are on the premises."

MORE QUALITY CONTROL

In addition to daily checks of the indexing and maintaining authority files, other quality control functions are emphasized at all of these companies. Quality Control specialists typically don't create any records, they oversee the general quality of the products and authority files. They take responsibility for consistency of editorial policy, such as how a particular subject should be indexed or how problems should be handled. They make decisions that apply for all indexers. Sometimes QC takes suggestions of new terms from indexers, sometimes they add new terms in anticipation of articles about major news events, such as summits, natural disasters, wars, etc.

QC is also responsible for correcting errors in the file. Error reports come in from the customer service department or phone calls or letters from users of the files. QC takes each of these reports and researches it. At IAC they correct errors in the master authority file so any updates in the future will reflect any changes. These error checks and correction procedures were not in place in the early days of IAC databases, so QC must also search through the databases looking for data that got into the wrong fields, for typos, and other input errors in older records. They admit that data created in the late 1970s and early 1980s had errors, because they did not have so many quality control checks. IAC believes now, however, that "data produced in the last five years is extremely clean."

QUOTAS

Some companies don't like to use the word "quotas," but recommended rates of indexing exist for nearly every index. These rates will vary widely depending on the complexity of the information being indexed, the corporate emphasis on speed vs. exhaustivity, and whether an abstract is included. For indexing only, IAC expects an experienced general indexer to create over 70 records per day. Abstractors do about 25 abstracts per day. (Together, all of the indexers produce between 3000-5000 records per day at IAC or over one million records per year.)

Wilson resists the idea of quotas, making allowances for individuals with special skills because "indexers work at different paces. If someone is a good quality indexer we make allowances" for the first year if one may not be as fast as some others. On an average, Wilson indexers do 45-50 articles per day and abstractors do 20 articles.

UMI/DC expects, after three months of probation, ABI/INFORM indexers to index 40 articles per day and to select an additional 40 articles for indexing later. ABI abstractors must do 15 per day. General Periodical Abstracts indexer/abstractors must average 55 articles per day, with incentives for exceeding that quota. Their "Century Plus Club" provides time off to indexers who break the 100 mark in a day. The entire department celebrates if they pass monthly production goals.

TOMORROW IS ANOTHER DAY

A bibliographic database is never complete. New issues to be indexed arrive every day; new magazine titles are added; records are created, reviewed, and revised; authority files are constantly updated; and new editorial decisions are made. Every day brings a new stock of periodicals and a new load of challenges. More and more bibliographic database producers are getting into full-text distribution as well. But, that is a topic for another day.