THE INDEXER

October 1988

Journal of

the Society of Indexers

and of the affiliated

American & Australian & Canadian Societies

Vol. 16 No. 2
Pages 73–144

ISSN 0019-4131 £7.50

THE INDEXER

Journal of the Society of Indexers and of its affiliated American & Australian & Canadian Societies

VOLUME 16 No. 2 OCTOBER 1988

CONTENTS

Founded 1958. Published twice yearly in April and October. Regular circulation in more than 60 countries. Annual subscription £15.00, US $25.00 p.a.

Abstracted in *Chemical Abstracts, Information Science Abstracts, INSPEC,* and *Library and Information Science Abstracts;* indexed in *Aslib Information, Book Review Index,* and *Library Literature;* available on major online hosts; contents page reproduced in *Documentaliste, Information Hotline,* and *IREBI.*

Advertising rates and notes for contributors available on request from the Editor.

The Society of Indexers is a non-profit organization founded in 1957 as an autonomous, professional body for indexers. Membership is open to any person or organization interested in the promotion of the role of indexers in the organization of knowledge, and of standards and techniques for all forms of indexing.

Back-of-the-book indexing with the nested phrase indexing system (NEPHIS)

James D. Anderson and Gary Radford

NEPHIS, Craven's Nested Phrase Indexing System, transfers one more indexing procedure, the creation of individual index entries, to computer algorithm, permitting the human indexer to concentrate on the intellectual task of analyzing text and naming its important features. Experience at Rutgers University has shown that novice indexers can learn NEPHIS quite quickly and can use it to produce acceptable indexes.

Back-of-the-book indexing

Back-of-the-book indexing requires five basic steps:

1. determining the features of probable interest in each paragraph or page of the book;

2. naming these features in succinct and meaningful statements;

3. converting these statements into index entries, each with a key term in the lead position and in most cases a subheading to provide context;

4. sorting and formatting entries into an alphabetical index, with suppression of redundant repetition; and

5. displaying relationships (syndetic structure) by:

(a) adding cross-references to link multiple terms that represent the same concept when only one term is used; or

(b) insuring consistent posting of referents under each equivalent entry if all are used; and

(c) adding cross-references to link terms representing narrower, related, and sometimes broader concepts.

In traditional practice, steps (2) and (3) are often merged.

Prior to these steps, of course, the overall design of the index must have been determined, with respect to:

1. exhaustivity (average number of terms or entries per page);

2. term specificity versus genericity (closeness of fit between terms and the concepts they index, and indirectly, the size of the index vocabulary, which in turn will determine the number of entries or referents per heading);

3. entry information and syntax (use of contextual subheadings);

4. indexable matter (textual elements indexed); and

5. extent of syndetic structure.[1,2]

These decisions will determine the overall size of the index.

Role of computers

No one has yet been able to automate or computerize all the steps in creating a back-of-the-book index. To be sure, it is possible to extract keywords from text by computer algorithm, but entirely too many keywords result. So far, the application of term frequency or location criteria has not resulted in term selection similar enough to human selection to be considered useful for back-of-the-book indexes.[3,4] The resulting plethora of keywords, together with the frequent lack of meaningful contextual subheadings, leads to entirely too many referents under each keyword, frustrating and defeating the user who seeks substantive discussion of particular concepts. Furthermore, the lack of consistent subheadings prevents meaningful subarrangement under lead terms to permit rapid scanning of printed index arrays. And although vocabulary management *may* be applied to terms extracted from text, it usually is not, resulting in the scattering of entries referring to the same or similar topics and the absence of syndetic structure linking equivalent terms, to say nothing about broader, narrower or related terms.

Information scientists skilled in the indexing process and aware of the qualities of good indexes have, however, created computer-based systems which enable indexers to transfer some of the more routine steps in back-of-the-book indexing to the computer. For many years, indexers have routinely left sorting and final formatting of index entries to the computer, using a wide variety of programs,[5] but they have continued to 'handcraft' each individual entry, including both lead term and subheading. We have recently experimented with the use of Timothy Craven's NEPHIS (Nested Phrase Indexing System)[6] for the creation of individual index entries for back-of-the-book indexes.[7] We believe the results have been good, and we think it is now time to share our experience with indexing colleagues.

Nested Phrase Indexing System (NEPHIS)

NEPHIS is a coding scheme plus computer algorithm for converting a summary statement into one or more index entries consisting of lead term and subheading. Here are two examples of summary statements:

1. Role of Story-telling in History of Human Communication.
2. Reference to Elliott (J. H.) on Role of Manuscripts in Human Communication.

The preferred statement construction makes maximum use of nouns and prepositional phrases. Adjectives are used when the preferred construction is awkward, e.g., 'human communication' rather than 'communication by (of, among) humans'. The second sample statement reflects a decision to index all reference citations not only by the name of the person cited, but also by the subject of the reference.

These summary statements were created by Gary Radford, a doctoral student in the Rutgers PhD program in Communication, Information and Library Studies. Prior to this assignment, he had no experience, education, or training in indexing. After an initial period of experimentation, he became quite proficient at creating one or more summary statements for each paragraph of text.

Once summary statements have been created, the next step is to add codes to indicate desired lead terms. At the beginning, Radford coded summary statements after they were created, but he quickly mastered the use of codes, and thereafter, he coded as he created the statements. Three codes are used:

@—To prevent a term in normal lead position from becoming a lead term in an index entry. Ordinarily, the first term in each statement and the first term in each 'nested phrase' will automatically become a lead term. These potential lead terms can be 'turned off' by placing a '@' in front of them. For example, the statement '@Ambiguous <Concepts>' will create an entry under 'Concepts' but not under 'Ambiguous', i.e.:

Concepts
  Ambiguous —.

< >—To mark additional lead terms and associated phrases that are 'nested' within a statement. The term or phrase within a set of brackets will be treated as a unit. Additional terms or phrases may be 'nested' within a longer nested phrase. The first term in each nested phrase, consisting of all words up to a '?' or a '>', will become a lead term in an index entry unless 'turned off' by a preceding '@'.

?—To indicate connectives, usually prepositions connecting terms or phrases. There are two kinds of connectives. 'Forward-reading' connectives precede nested phrases; 'backward-reading' connectives come at the end of nested phrases, i.e., within nested phrase brackets. Backward-reading connectives are used to indicate reciprocal relationships that should result in entries reading in both directions. For example, the statement 'Information? and <Behavior? and>' will create two entries:

Information
  and Behavior

Behavior
  and Information

Similarly, the statement 'Capitalism? compared to <Communism? compared to>' will create the entries:

Capitalism
  compared to Communism

Communism
  compared to Capitalism

Note that in this example 'compared to' is treated as a connective.

To create index entries, the NEPHIS program takes each lead term, one at a time. It drops any forward-reading connective immediately preceding the lead term. Following the lead term, on the second (and/or third) line of the entry, come additional terms in this order: remaining terms in the same nested phrase or any backward-reading connective at the end of the lead term's nested phrase; any preceding terms in a larger nested phrase, if there is one, otherwise any preceding terms in the statement as a whole; any following terms in a larger nested phrase, if there is one, otherwise any following terms in the statement as a whole. If there are multiple nested phrases, this process proceeds successively through each larger nested phrase until all elements are processed.

Backward-reading connectives are ignored unless they conclude the lead term's nested phrase or a larger nested phrase which contains the lead term's nested phrase.

Within subheadings, the end of a nested phrase is marked by a full stop if the phrase does not end with a backward-reading connective *and* it is followed by additional terms from a larger nested phrase or the statement as a whole.

If a three-level format is used, additional terms from the lead term's nested phrase, if any, go into the second level. If there are no additional terms in the lead term's nested phrase, terms from the next larger nested phrase go into the second level. All remaining terms, if any, go into the third level.

Terms preceding the lead term that are not connected to the lead term with a forward-reading connective are followed by a dash. For example, the statement 'History? of <Human <Communication>>' produces the following entries:

History
  of Human Communication

Human Communication
  History

Communication
  Human —. History

In the third entry, 'Human —.' is followed by a dash because in the original statement it is not connected to

the lead term 'Communication' with a forward-reading connective. In the second entry, 'History' is not followed by a dash because in the original statement it is connected to the lead term by the forward-reading connective '? of'. This procedure is designed to indicate inverted adjective-noun combinations.

NEPHIS entry creation: some examples

All NEPHIS codes were used in the following statement. The numbers over each term are placed there to facilitate explanation.

```
   1     2          3          4          5
@Role?  of  <Imagination?  and  <Communication?
 6      7     8              9
and> >? in <Human <Consciousness> >
```

Term 1 '@Role' will be skipped over as a lead term, since it is 'turned off' with a '@'

Term 2 '? of <' is a forward reading connective and as such is not eligible for lead position in an entry.

Term 3 '<Imagination?' is the first lead term. It will be followed in a subheading by the remaining terms within its nested phrase, terms 4 and 5 '? and <Communication?' and a concluding full stop (period). The backward-reading connective, term 6 '? and>', will be omitted since it is neither part of the lead term's nested phrase nor a larger nested phrase, but is part of a nested phrase *within* the lead term's nested phrase. Next comes any term or terms preceding the lead term's nested phrase within the next larger nested phrase, which in this case is the complete statement. These are terms 1 and 2 '@Role? of <', but term 2 '?of <' is dropped since it immediately precedes the lead term. Next come terms following the lead term's nested phrase within the next larger nested phrase or, in this case, the complete statement since there is no larger nested phrase. These are terms 7 through 9 '? in <Human <Consciousness> >', resulting in the entry:

```
Imagination
  and Communication.
    Role in Human Consciousness
```

Term 5 '<Communication?' is the next lead term. Since its nested phrase concludes with a backward-reading connective, term 6 '? and >', the subheading begins with this connective and the preceding term in the next larger nested phrase, term 3 '<Imagination?'. Since the forward-reading connective term 4 '? and <' immediately precedes the lead term, it is dropped, and since 'Imagination' concludes a nested phrase and will be followed by terms in a larger phrase, it is followed by a full stop. The next larger nested phrase is the entire statement. Preceding terms come before following terms, so we have term 1 '@Role?' followed by terms 7, 8 and 9 '? in <Human <Consciousness> >'. The forward-reading connective, term 2 '? of <', is omitted since its 'parent' term 3 '<Imagination?' has already been placed in the entry. The resulting entry is:

```
Communication
  and Imagination.
    Role in Human Consciousness
```

The next entry begins with terms 8 and 9 '<Human <Consciousness> >'. This is considered a single term since it is not interrupted by either a '?' or a '>'. The immediately preceding forward-reading connective, term 7 '? in <', is dropped. The next larger nested phrase is the entire statement, so its remaining terms 1 through 5 '@Role? of <Imagination? and <Communication?' follow the lead term as a subheading. Since a backward-reading connective is used only when a term within its nested phrase is in lead position, term 6 '? and>' is ignored. The resulting entry reads:

```
Human Consciousness
  Role of Imagination and Communication
```

The final entry begins with term 9 '<Consciousness>'. The first term in the subheading comes from the next larger nested phrase, i.e. term 8 '<Human <'. Since there is no backward-reading connective connecting 'Consciousness' to 'Human', the term 'Human' is followed by a dash, indicating an inverted noun-adjective relationship. Since there are no terms within the nested phrase beginning with 'Human' that follow the lead term nor a backward-reading connective at the end of that nested phrase, the dash is followed by a full stop. The remaining terms 1 though 5 '@Role? of <Imagination? and <Communication?' precede the nested phrase that begins with term 8 '<Human <', so they conclude the entry, resulting in the following:

```
Consciousness
  Human —.
    Role of Imagination and Communication
```

The sample summary statements given above were coded as follows:

1. @Role? of <Story-telling>? in <History? of <Human <Communication> > >
2. @Reference? to <Elliott (J. H.)>? on <@Role? of <Manuscripts>? in <Human <Communication> > >

Following computer processing, they resulted in the following formatted, but unsorted, index entries.

```
Phrase 1:
  Story-telling
    Role in History of Human Communication
  History
    of Human Communication.
      Role of Story-telling
  Human Communication
    History.
      Role of Story-telling
  Communication
    Human —.
      History. Role of Story-telling
```

Phrase 2:

Elliott (J. H.)
Reference on Role of Manuscripts in Human Communication

Manuscripts
Role in Human Communication.
Reference to Elliott (J. H.)

Human Communication
Role of Manuscripts.
Reference to Elliott (J. H.)

Communication
Human —.
Role of Manuscripts. Reference to Elliott (J. H.)

NEPHIS cross-references

One-way 'see' references and two-way reciprocal 'see also' references may be coded as follows:

Attorneys? See <@Lawyers>
Library science? See also <Information science? See also>

The first statement will create the entry:

Attorneys
See Lawyers

The second statement will create two entries:

Library science
See also Information science

Information science
See also Library science

From strings to index entries

The coded NEPHIS strings were processed by 'IOTA: Information Organization based on Textual Analysis', a suite of programs developed at Rutgers for the design and creation of small or experimental textual databases.[8] NEPHIS is only one of many approaches to indexing and classification supported by IOTA. Others include keyword plus context, permuted keyword, subject headings (e.g., Library of Congress), enumerative classification (e.g., Dewey or Library of Congress), faceted indexing and classification, and variations of PRECIS.

To use IOTA, a record structure must be set up to represent the organizational unit. For a back-of-the-book index, this is usually a page, although it could be a paragraph, column or other unit. For other applications, it could be various types of documents or portions of documents, e.g. periodical articles or monographs. Fields are created within the record for surrogates and subject descriptions. For a typical back-of-the-book index, the only 'surrogate' is a page number, and this can be the same as the record number, so for a back-of-the-book NEPHIS index, the record structure needs only fields for NEPHIS statements, with as many fields as the number of NEPHIS statements expected for each page or other organizational unit. Fields for additional statements may be added at any time as needed. Figure 1 illustrates a list of fields.

```
rec\ Record #
st1\ Statement 1
st2\ Statement 2
st3\ Statement 3
```

Figure 1. A list of fields, with 3-character field tags and field names, as set up by the IOTA fields specification program.

NEPHIS statements may be entered into records using the IOTA input program, which automatically supplies record numbers and field tags, or any wordprocessing program may be used as long as a 'pure' ASCII file is created, without any invisible wordprocessing characters. If the IOTA input program is used, suggested 'default' punctuation should be ignored, replaced by the NEPHIS codes discussed previously. Figure 2 shows sample records for three organizational units, each having fields for three NEPHIS statements. The NEPHIS statements used for this and following figures are those already used as examples.

Once NEPHIS strings have been input, random access records are created in preparation for subsequent classification, indexing, and/or searching. Field tags are verified and errors noted, and a proof file is created (figure 3). Classified sorting is not usually appropriate for a back-of-the-book index, but is useful for displays of document surrogates, such as citations and abstracts; then the NEPHIS entries refer to such surrogates rather than to page or entry numbers in a published book or other document. Records may be sorted by up to eight fields, such as author, title, date, and/or various subject fields in any order. If classified sorting is not desired, 'input order' should be selected.

The IOTA indexing program creates the NEPHIS entries

```
rec\1
st1\@Ambiguous <Concepts>
st2\Information? and <Behavior? and>
st3\Capitalism? compared to <Communism? compared to>
rec\2
st1\History? of <Human <Communication>>
st2\@Role? of <Imagination? and <Communication? and>>? in <Human
<Consciousness>>
st3\@Role? of <Story-telling>? in <History? of <Human <Communica-
tion>>>
rec\3
st1\@Reference? to <Elliott (J. H.)>? on <@Role? of <Manuscripts>? in
<Human <Communication>>>
st2\Attorneys? See <@Lawyers>
st3\Library science? See also <Information science? See also>
```

Figure 2. A sample datafile as input via the IOTA input program. Data may be input also via any standard wordprocessing program that will produce an ASCII file without invisible wordprocessing codes. WordStar is an example. Each field consists of a single line, concluding with a carriage return, but long lines are 'wrapped around' here for display purposes.

as one of its indexing options. Processing is done in several stages. First NEPHIS strings are combined with entry numbers, which for a typical back-of-the-book index will be page numbers (figure 4). Next, these NEPHIS strings are converted into multiple entries (figure 5). A sort key is attached to the front of these entries to insure proper computer sorting (figure 6). Finally, the sorted entries are formatted (figure 7). During the formatting stage, questions relating to a thesaurus should be ignored unless a separate IOTA thesaurus was created for the index, something which is not generally done for a back-of-the-book index. Lead entries may be displayed in 'all-caps' or in upper and lower case as originally entered. Entry numbers referring 'to the classified file' should be included, since for a back-of-the-book index, these entry numbers will be the page numbers of the book.

Each line in the formatted index concludes with a line-feed character rather than a full 'carriage return'. This is the ideal form for transfer to an Apple Macintosh computer for typesetting and desktop publishing, but it is not helpful for the final editing of the index with a standard wordprocessing program. The index file can be

```
Record #   : 1
Statement 1 : @Ambiguous <Concepts>
Statement 2 : Information? and <Behavior? and>
Statement 3 : Capitalism? compared to <Communism? compared to>
Record #   : 2
Statement 1 : History? of <Human <Communication>>
Statement 2 : @Role? of <Imagination? and <Communication? and>>? in
              <Human <Consciousness>>
Statement 3 : @Role? of <Story-telling>? in <History? of <Human
              <Communication>>>
Record #   : 3
Statement 1 : @Reference? to <Elliott (J. H.)>? on <@Role? of <Manu-
              scripts>? in <Human <Communication>>>
Statement 2 : Attorneys? See <@Lawyers>
Statement 3 : Library science? See also <Information science? See also>
```

Figure 3. An IOTA proof file for checking data input.

```
@Ambiguous <Concepts>: \#0001
Information? and <Behavior? and>: \#0001
Capitalism? compared to <Communism? compared to>: \#0001
History? of <Human <Communication>>: \#0002
@Role? of <Imagination? and <Communication? and>>? in <Human
<Consciousness>>: \#0002
@Role? of <Story-telling>? in <History? of <Human <Communica-
tion>>>: \#0002
@Reference? to <Elliott (J. H.)>? on <@Role? of <Manuscripts>? in
<Human <Communication>>>: \#0003
Attorneys? See <@Lawyers>: \#0003
Library science? See also <Information science? See also>: \#0003
```

Figure 4. NEPHIS strings combined with entry numbers (i.e., page numbers). Page numbers attached to 'see' and 'see also' references are later removed.

```
Concepts \ Ambiguous—: \#0001
Information \ and Behavior: \#0001
Behavior \ and Information: \#0001
Capitalism \ compared to Communism: \#0001
Communism \ compared to Capitalism: \#0001
History \ of Human Communication: \#0002
Human Communication \ History: \#0002
Communication \ Human —. \ History: \#0002
Imagination \ and Communication. \ Role in Human Consciousness: \#0002
Communication \ and Imagination. \ Role in Human Consciousness: \#0002
Human Consciousness \ Role of Imagination and Communication: \#0002
Consciousness \ Human —. \ Role of Imagination and Communication: \#0002
Story-telling \ Role in History of Human Communication: \#0002
History \ of Human Communication. \ Role of Story-telling: \#0002
Human Communication \ History. \ Role of Story-telling: \#0002
Communication \ Human —. \ History. Role of Story-telling: \#0002
Elliott (J. H.) \ Reference on Role of Manuscripts in Human Communication:
\#0003
Manuscripts \ Role in Human Communication. \ Reference to Elliott (J. H.):
\#0003
Human Communication \ Role of Manuscripts. \ Reference to Elliott (J. H.):
\#0003
Communication \ Human —. \ Role of Manuscripts. Reference to Elliott (J. H.):
\#0003
Attorneys \ See Lawyers
Library science \ See also Information science
Information science \ See also Library science
```

Figure 5. Individual NEPHIS entries, not yet sorted or formatted.

```
CONCEPTS AMBIGUOUS— #0001\Concepts \ Ambiguous—: \#0001
INFORMATION AND BEVAVIOR #0001\Information \ and Behavior:
\#0001
BEHAVIOR AND INFORMATION #0001\Behavior \ and Information:
\#0001
CAPITALISM COMPARED TO COMMUNISM \Capitalism \ compared to
Communism: \#0001
COMMUNISM COMPARED TO CAPITALISM \Communism \ compared to
Capitalism: \#0001
HISTORY OF HUMAN COMMUNICATION #\History \ of Human Com-
munication: \#0002
HUMAN COMMUNICATION HISTORY #000\Human Communication \
History: \#0002
COMMUNICATION HUMAN — HISTORY \Communication \ Human —. \
History: \#0002
IMAGINATION AND COMMUNICATION ROLE IN \Imagination \ and
Communication. \ Role in Human Consciousness: \#0002
COMMUNICATION AND IMAGINATION ROLE IN \Communication \
and Imagination. \ Role in Human Consciousness: \#0002
HUMAN CONSCIOUSNESS ROLE OF IMAGINATION AND CO\Human
Consciousness \ Role of Imagination and Communication: \#0002
CONSCIOUSNESS HUMAN — ROLE OF IMAGINAT\Consciousness \
Human —. \ Role of Imagination and Communication: \#0002
STORY-TELLING ROLE IN HISTORY OF HUMAN COMMUNI\Story-
telling \ Role in History of Human Communication: \#0002
HISTORY OF HUMAN COMMUNICATION ROLE OF\History \ of
Human Communication. \ Role of Story-telling: \#0002
HUMAN COMMUNICATION HISTORY ROLE OF ST\Human Communi-
cation \ History. \ Role of Story-telling: \#0002
COMMUNICATION HUMAN — HISTORY ROLE OF\Communication \
Human —. \ History. Role of Story-telling: \#0002
ELLIOTT (J H) REFERENCE ON ROLE OF MANUSCRIP\Elliott (J. H.) \
Reference on Role of Manuscripts in Human Communication: \#0003
MANUSCRIPTS ROLE IN HUMAN COMMUNICATION \Manuscripts \
Role in Human Communication. \ Reference to Elliott (J. H.): \#0003
HUMAN COMMUNICATION ROLE OF MANUSCRIPTS \Human Com-
munication \ Role of Manuscripts. \ Reference to Elliott (J. H.): \#0003
COMMUNICATION HUMAN — ROLE OF MANUSCRI\Communication \
Human —. \ Role of Manuscripts. Reference to Elliott (J. H.): \#0003
ATTORNEYS SEE LAWYERS \Attorneys \ See Lawyers \
LIBRARY SCIENCE SEE ALSO INFORMATION \Library science \ See also
Information science \
INFORMATION SCIENCE SEE ALSO LIBRARY \Information science \ See
also Library science \
```

Figure 6. NEPHIS entries with sort-key prefix, ready for sorting. Each entry consists of a single line, but long lines have been reformatted to fit on this page.

Figure 7. Final sorted and formatted NEPHIS index, using 3-level format.

converted to a standard ASCII file with standard carriage returns by using the IOTA 'view' program to copy the original index file to a new file with carriage returns.

Conclusions

After using NEPHIS for several back-of-the-book indexes and training novice indexers in its application, we are convinced that it is a useful tool for both professional and novice indexers. Readers who are interested in the Rutgers IOTA software should write to James D. Anderson, PO Box 38, New Brunswick, NJ 08903, USA. We are happy to share it, asking only for the cost of disks, documentation and postage. It must be understood, however, that the IOTA software is essentially exploratory in nature. It is written in Microsoft BASIC, which will run on IBM, IBM compatible, and CP/M computers, and it suffers from the limitations of the BASIC programming language, such as the limit of 255 characters to variable length. A professional programmer (in contrast to Anderson who programmed the current version of IOTA) is now transforming IOTA into a professionally produced software package. Nevertheless, IOTA has been used successfully for several years in classes in information organization and database design at Rutgers and, as described here, for the production of back-of-the-book indexes. We invite readers to try it themselves.

Notes and references

1. Anderson, James D. Essential decisions in indexing systems design. In *Indexing Specialized Formats and Subjects*, ed. by Hilda Feinberg. Metuchen, NJ: Scarecrow Press, 1983: 1–21.
2. Milstead, Jessica L. *Subject access systems: alternatives in design*. Orlando, FL: Academic Press, 1984. 212 p.
3. Grunberger, Michael. Textual analysis and the assignment of index entries for social science and humanities monographs. Ph.D. Dissertation, Rutgers, the State University of New Jersey, New Brunswick, NJ, 1985. 127 p.
4. Weinberg, Bella. Word frequency and automatic indexing. D.L.S. Dissertation, Columbia University, New York, 1981. 221 p.
5. Computer programs for back-of-the-book indexing are regularly described in the Society of Indexers' *MicroIndexer* and in 'The Electronic Shoebox' column in the *Newsletter* of the American Society of Indexers. A compilation of software reviews appearing in the ASI *Newsletter* has been published, with plans for regular updates. See *A guide to indexing software*, compiled by Linda K. Fetters. Washington, DC: ASI, 1987. ($7.50 for ASI members, $10 for others).
6. Timothy Craven has published many articles on NEPHIS. A good overview of the system in the context of string indexing in general is provided in Craven, Timothy C. *String indexing*. Orlando, FL: Academic Press, 1986, 246 p. See especially Appendix B, A case study of string indexing, p. 192–212, and Appendix C, A brief manual for composition of NEPHIS input strings, p. 213–221.
7. So far, we have used NEPHIS to create back-of-the-book indexes for the following books: *Information and behavior*, vol. 2, 1988 (Brent D. Ruben, ed., New Brunswick, NJ: Transaction); Brent D. Ruben, *Communication and human behavior*, 2nd ed. (New York: MacMillan, 1988); Dennis K. Mumby, *Communication and power in organizations: discourse, ideology and domination* (Norwood, NJ: Ablex, 1988). In addition, Professor Andrew D. Abbott of Rutgers' Department of Sociology used NEPHIS to index his book *The system of professions* (Chicago: University of Chicago Press, 1988). All examples used in this paper are taken from the index for *Information and behavior*.
8. Anderson, James D. Information Organization based on Textual Analysis (IOTA): instructional programs for database design. In *The library microcomputer environment: management issues*, edited by Sheila S. Intner and Jane Anne Hannigan. Phoenix: Oryx Press, 1988: 145–168. 'IOTA2: Two-Disk Version Documentation.' New Brunswick, NJ: School of Communication, Information and Library Studies, 1987. 47 p.

The authors can be contacted at the School of Communication, Information and Library Studies, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903, USA.