Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

Contextual Hierarchies in Classification Schemes

Susanne M. Humphrey
National Library of Medicine
Bethesda, Maryland 20894, USA
humphrey@lhc.nlm.nih.gov

1. INTRODUCTION

This paper is concerned with the encoding of contextual hierarchies. In particular, such hierarchies make it possible to create a single, complete classified display of very large thesauri. This classification may use the same descriptor with different views, as evidenced by the same descriptor as more than one node in the classification, where the nodes have different sets of children. This sort of knowledge organization requires special computer representation techniques using contextual indicators for encoding the parent-child relationship. One "solution" is to avoid having a unified classification in favor of many hierarchical families as used by the INSPEC® and ERIC® thesauri. However, this author considers it a particular strength to have a unified classification, of which the Medical Subject Headings (MeSH®) tree structures is a primary example.

This paper describes the traditional method of using *tree numbers* as contextual indicators. We then propose a new experimental method of *semantic labels*, developed for the MedIndEx™ prototype, as possibly having certain advantages over tree numbers. We conclude with the hope that this workshop will provide feedback regarding the significance of the problem and substance of our proposal.

2. USE OF TREE-NUMBER-BASED CLASSIFICATION

MeSH, NLM's thesaurus for indexing, cataloging, and retrieval from MEDLINE® and other NLM databases, is one of the most popular thesauri around, in large part due to its classification. MeSH is unique among large thesauri in providing a *unified* classification scheme, published as *MeSH Tree Structures*, a single top-to-bottom display of all MeSH. (Since "trees" may be used for referring to the entire tree or specific sections, e.g., the DIGESTIVE SYSTEM tree, to avoid confusion, we will use "nodes" for referring to sections of the tree.)

In MeSH, representing different contextual views is made possible by tree numbers. For example, the descriptor BONE AND BONES may be viewed, as the duality of its form indicates, as a tissue (BONE) or as parts of the skeleton (BONES). The MeSH trees do this easily, by labeling different views with different tree numbers as in the following display. The A2 tree number expresses the musculoskeletal context; the A10 number, the tissue context.

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60.
doi:10.7152/acro.v3i1.12595

MUSCULOSKELETAL SYSTEM	**A2**
...	...
SKELETON	A2.835
BONE AND BONES	**A2.835.232**
ARM BONES	A2.835.232.87
...	...
LEG BONES	A2.835.232.484
...	...
...	...
SKULL	A2.835.232.781
...	...
FACIAL BONES	A2.835.232.781.324
JAW	A2.835.232.781.324.502

TISSUE TYPES	**A10**
CONNECTIVE TISSUE	A10.165
...	...
BONE AND BONES	**A10.165.265**
BONE MATRIX	A10.165.265.166
... ...	
PERIOSTEUM	A10.165.265.746

Since 1966, this tree number scheme has traditionally been the basis for display and use of the MeSH classification. Each MeSH descriptor (excluding a few special terms known as check tags) has mapped to it one or more tree numbers. Since these are used for sorting the published *MeSH Tree Structures*, they mark the location of the descriptors in the tree, and are therefore used as indexes to the tree from the *MeSH Annotated Alphabetic List*. This is illustrated by the following BONE AND BONES entry as pointers to locations in the above tree display ("+" indicates that a node has children):

BONE AND BONES
A2.835.232+ A10.165.264+

Tree numbers also have an extremely important function in MEDLINE retrieval. Searchers use them in "explode" expressions as shorthand for the union of descriptors in a tree node. Furthermore, for efficiency, tree numbers are actual indexes to MEDLINE citations. This may be seen in the following display of the index to MEDLINE containing MeSH terms and tree number indexes (in the TERM column) and the count of MEDLINE citations for each index (in the POSTINGS column). The entries BONE AND BONES, A2.835.232., and A10.165.265. are indexes to the same set of 4,501 citations.

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

SELECT #	POSTINGS	TERM
1	554	BOMBESIN
2	116	BONDING, HUMAN-PET
3	4501	BONE AND BONES
4	437	BONE CEMENTS
5	121	BONE CONDUCTION

SELECT #	POSTINGS	TERM
6	12	A2.633.893.
7	11	A2.835.
8	4501	A2.835.232.
9	208	A2.835.232.251.
10	449	A2.835.232.251.352.

SELECT #	POSTINGS	TERM
11	3545	A10.165.114.
12	498	A10.165.114.322.
13	4501	A10.165.265.
14	303	A10.165.265.166.
15	131	A10.165.265.276.

Thus, when a searcher on NLM's retrieval system enters the expression EXPLODE A2.835.232, the search programs consider this to be equivalent to the expression ALL A2.835.232:(MN), which means truncate the A2.835.232 MeSH tree number as a search term. This then becomes a way of specifying the union of all MN indexes beginning with A2.835.232, which is the set of indexes corresponding to the union of descriptors in the BONE AND BONES node in the MUSCULOSKELETAL SYSTEM (or A2) subcategory (i.e., BONE AND BONES OR ARM BONES OR etc.). This use of tree number indexes is immensely more efficient than if the system were to look up the tree number in the MeSH file and form the union of actual descriptors, and then return to the MEDLINE database and access the descriptors as indexes. In the NLM retrieval system, searchers have the option of using the descriptor in the explode expression, e.g., EXPLODE BONE AN# BONES (wildcard character "#" is used here because AND would be processed as the Boolean intersection operator) is equivalent to the expression EXPLODE A2.835.232 OR EXPLODE A10.165.265.

It is important for searchers to understand that the descriptor, not the tree number, is the unit concept in MeSH. That is, each MEDLINE citation indexed by BONE AND BONES is indexed by both tree numbers (as illustrated by equivalent postings in the above display). Therefore, when a searcher uses this descriptor, resulting citations may discuss bone as tissue or bone as skeleton or both. Even when EXPLODE A10.165.265 is the search expression, the retrieval indexed by BONE AND BONES may well include citations concerned with bones as a skeletal component. This is not a serious drawback in retrieval since descriptors are seldom searched in isolation, and their intersection with other descriptors in a search strategy often results in the desired narrower context.

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-50. doi:10.7152/acro.v3i1.12595

3. SEMANTIC LABELS FOR REPRESENTING CONTEXTUAL HIERARCHIES IN THE MEDINDEX PROTOTYPE

The MedIndEx system, a prototype indexing expert system reported at the last year's SIG/CR Workshop, uses MeSH as its authority on indexing concepts. The prototype requires the use of MeSH in a classified form. Specifically, an important help to MedIndEx users is the on-demand request for classified displays of MeSH terms as permissible fillers of indexing frames.

As important as tree numbers are for representing and displaying the MeSH classification, there were some disadvantages when it came to using them for MedIndEx. In MedIndEx we wanted to be able to display the entire tree interactively (as is possible in its published form). However the computerized MeSH file does not contain a root node. We also might want to display a category comprised of more than one subcategory. But MeSH has gaps when it comes to tree numbers for these categories. For example, there is no MeSH descriptor "Organisms", which would be the parental node of MeSH subcategories B1 - B6 (INVERTEBRATES, VERTEBRATES, BACTERIA, VIRUSES, ALGAE AND FUNGI, and PLANTS, respectively). In other words, there is no MeSH number "B" as a node. MedIndEx being an experimental system, it would not be a problem to fill in these gaps, creating a root node and, optionally, the missing category nodes.

But there were more fundamental problems. We wanted flexibility in being able to "play" with various arrangements. Tree numbers do not lend themselves to this type of flexibility. For example, the descriptor MYCOBACTERIUM, ATYPICAL has three locations in Subcategory B3 (Bacteria), denoted by tree numbers B3.100.67.595.552.552.350, B3.510.460.400.410.552, and B3.510.595.552.552.350. Thus, using MeSH tree numbers would mean carrying over and maintaining these quite unwieldy codes. Moving or copying a high-level node would entail transactions of adding and deleting many codes like this.

In addition, in this author's opinion, users should never have to deal with such numbers. They should not have to enter such a number, nor even read it. This we felt safe in ensuring, since MedIndEx employs a graphical interface. Since users would not need to know of tree numbers, we also began to consider, why should the MedIndEx knowledge engineer (who makes the classification) have to deal with these numbers either. MedIndEx, using a frame data structure, could very well use a slot to link the terms into a MeSH-like classification, and this is how we implemented it. The following display shows the *Skeleton* frame with its CHILDREN slot having as values the terms *Bone and Bones, Joints*, and *Muscles*, and the *Connective Tissue* frame with its four children, also including *Bone and Bones*.

```
(|Skeleton|
   (CHILDREN
      (VALUE
         (... |Bone and Bones| |Joints| |Muscles|)))

(|Connective Tissue|
   (CHILDREN
      (VALUE
         (... |Adipose Tissue| |Bone and Bones| |Cartilage| |Elastic Tissue|)))
```


ISSN: 2324-

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

In order to display our classification, we wrote programs that generated it in "dot notation", in essence, tree numbers without the numbers part, where number of leading dots was an indicator of depth. The resulting display for the above frames, including parental nodes and the first child of *Bone and Bones* (which is different for each context) would be as follows:

KB-ROOT
. Anatomical Structures
.. Musculoskeletal System
... Skeleton
.... **Bone and Bones**
..... **Arm Bones**
.... Joints
.... Muscles
.. Tissue Types
... Connective Tissue
.... Adipose Tissue
.... **Bone and Bones**
..... **Bone Matrix**
.... Cartilage
.... Elastic Tissue

However, slots for linking descriptors in a binary relationship posed a problem. It became clear we would have to do something special with the CHILDREN slot in order to implement contextual hierarchies made possible by tree numbers traditionally. Otherwise, any child of *Bone and Bones* would appear in both locations in the above display, i.e., *Arm Bones* and *Bone Matrix* could not help but be children of *Bone and Bones* both in the *Skeleton* node and the *Connective Tissue* node.

We therefore implemented a system of semantic labels. Rather than have the value of the CHILDREN slot be simply a list of child-terms, we encoded the value as a list of terms where the first term in each list would be this label representing a context. For example, the frame for *Bone and Bones* with its CHILDREN slot may be encoded using the label *Musculoskeletal System* for the skeletal children and the label *Tissues Types* for the tissue children, as follows:

```
(|Bone and Bones|
   (CHILDREN
      (VALUE
         (|Musculoskeletal System| |Arm Bones| ...))
         (|Tissues Types| |Bone Matrix| ...))))
```

An obvious advantage of semantic labels is that they mean something. Compare asking anybody not especially familiar with the MeSH tree number system to distinguish between *Bone and Bones* in the context of A2 versus A10, in contrast to distinguishing *Bone and Bones* in terms of Musculoskeletal System versus Tissues Types.

Where would these labels come from? Since labels represent larger contexts for a particular term, they should be the same as terms higher in the classification. It would seem impractical to use the

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60.
doi:10.7152/acro.v3i1.12595

entire string of a descriptor's ancestral nodes to represent the context for that descriptor. We propose using only one of these nodes as the label. But which label to choose? For example, o earlier *Anatomical Structures* display leading to *Bone and Bones* provided several choices of lab namely, *KB-ROOT, Anatomical Structures, Musculoskeletal System, Skeleton, Tissues Types,* a *Connective Tissue.*

To switch to a less technical domain in MeSH, the *Pennsylvania* node in the *Geographicals* no also requires contextual labeling because of its classification in the *United States by Individual State* node and in the *Appalachian Region* node. *Philadelphia*, generally not considered part of Appalachia, is a child of *Pennsylvania* only in the former node.

KB-ROOT
. Geographicals
.. America
... North America
.... United States
..... **United States by Individual State**
..... **Pennsylvania**
..... **Philadelphia**
..... **United States by Region**
..... **Appalachian Region**
..... **Pennsylvania**

We have developed an algorithm for computerizing the representation of classifications using semantic labels. The input is a file comprised of a hierarchy in dotted notation, as above. The outp is corresponding frames with their CHILDREN slots including system-determined semantic label Heuristics in this algorithm are based entirely on factors such as depth and size of nodes and whether a node is a so-called "trouble node" (*Pennsylvania* in this example). These labels appe as names of paths in the following display:

"In |**Americ**a| path: "
(KB-ROOT
(|Geographicals|
(|America|
(|North America|
(|United States|
(|**United States by Region**|
(|**Appalachian Region**| ...
(|**Pennsylvania**|)
...)))))))

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595


```
"In |United States by Individual State| path: "
(KB-ROOT
    (|Geographicals|
        (|America|
            (|North America|
                (|United States|
                    (|United States by Individual State|...
                                        (|Pennsylvania| (|Philadelphia|))
                                        ... ))))))
```

Each frame corresponding to a term in the above *America* path used the term *America* as a semantic label for its children in that path. Each frame in the *United States by Individual State* path used this term as a semantic label for its children in that path. For example, the *United States* frame has both labels, as follows:

```
(|United States|
    (CHILDREN
        (VALUE
            (|America| |United States by Region|)
            (|United States by Individual State| |United States by Individual State|))))
```

The *Pennsylvania* frame, as follows, uses only the *United States by Individual State* label as a special context in which it is appropriate to have *Philadelphia* as a child of *Pennsylvania*:

```
(|Pennsylvania|
    (CHILDREN
        (VALUE
            (|United States by Individual State| |Philadelphia|))))
```

This labeling was determined computationally. The system identified the trouble node *Pennsylvania* (i.e., where one context demanded *Philadelphia* as an immediate child and another did not). It used the heuristic of third-level terms (i.e., *America*) as default labels. It ruled out *Pennsylvania* itself as a label using the rule to avoid trouble nodes as labels if at all possible. That is, *Pennsylvania* would not be useful in distinguishing between the contexts. That left *North America, United States, United States by Region, Appalachian Region,* and *United States by Individual State* as candidates.

Another piece of information known to the program is that the *North America* node contains trouble nodes only with respect to U.S. by region versus individual State. Based on this fact, the algorithm can determine that *North America* and *United States* are too high for making this distinction. There also are trouble nodes with respect to other regions, such as *Great Lakes Region,* where *Pennsylvania* is repeated without a child, and *Illinois* and *New York* are trouble nodes because of major cities under these nodes in the *United States by Individual State* context, as follows:

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60.
doi:10.7152/acro.v3i1.12595

"In |**America**| path: "
(KB-ROOT
(|Geographicals|
(|America|
(|North America|
(|United States|
(|**United States by Region**|
(|**Appalachian Region**| (|Kentucky|)
(|**Pennsylvania**|)
(|Tennessee|)
(|West Virginia|))
(|**Great Lakes Region**| (|**Illinois**|)
(|Indiana|)
(|Michigan|)
(|Minnesota|)
(|**New York**|)
(|Ohio|)
(|**Pennsylvania**|)
(|Wisconsin|))))))))

"In |**United States by Individual State**| path: "
(KB-ROOT
(|Geographicals|
(|America|
(|North America|
(|United States|
(|**United States by Individual State**| ...
(|**Illinois**| (|**Chicago**|))
...
(|**New York**| (|**New York City**|))
...
(|**Pennsylvania**| (|**Philadelphia**|)
...))))))

(Apparently a decision was made not to include cities in the regional node even when the city is i
that region, e.g., *Chicago* in the *Great Lakes Region.*)

Thus, choosing regions as semantic labels would result in not only *Appalachian Region* as a label
but *Great Lakes Region* as well. This brings into play another heuristic which is for trouble node
to share labels where possible. In the current example, if *Appalachian Region* were used for the
Pennsylvania problem, then *Great Lakes Region* would also be used for the same problem. Since
the contextual problem is shared by all instances, it seemed it would be better to use a higher ter
as a single label (*United States by Individual State* or *United States by Region*) instead of having
to use several lower ones (*Appalachian Region, Great Lakes Region*, and so forth).

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

The computation has now narrowed the choice to be between *United States by Individual State* and *United States by Region* as the special label. Either would work, as shown by the following, which uses the latter:

"In |**America**| path: "
(KB-ROOT
(|Geographicals|
(|America|
(|North America|
(|United States|
(|**United States by Individual State**| ...
(|**Illinois**| (|**Chicago**|))
...
(|**New York**| (|**New York City**|))
...
(|**Pennsylvania**| (|**Philadelphia**|))
...))))))

"In |**United States by Region**| path: "
(KB-ROOT
(|Geographicals|
(|America|
(|North America|
(|United States|
(|**United States by Region**|
(|**Appalachian Region**| (|Kentucky|)
(|**Pennsylvania**|)
(|Tennessee|)
(|West Virginia|))
(|**Great Lakes Region**| (|**Illinois**|)
(|Indiana|)
(|Michigan|)
(|Minnesota|)
(|**New York**|)
(|Ohio|)
(|**Pennsylvania**|)
(|Wisconsin|))))))))

But consider the editing situation where the thesaurus specialist is using a Thesaurus Management System (TMS) tool to update the thesaurus. The task would be to add *Detroit* as a child of *Michigan*. The TMS interface would require the user to select an existing label for *Detroit*. If *United States by Individual State* were the label (rather than *United States by Region*), the selection menu would display:

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

America
United States by Individual State

and the correct choice for the user to make would be *United States by Individual State*. If *United States by Region* were the label, the selection menu would display:

America
United States by Region

where the correct choice would be *America*. Again, the menu resulting in selecting the more specific label — *United States by Individual State* rather than *America* — seems preferable, being closer to the level of the trouble node itself. This became a heuristic for analogous situations of not needing a special label where there are no children in one of the contexts of a trouble node.

As an additional point of information, the second level term, *Geographicals*, is used as a label for immediate children that have no further breakdown, as follows:

"In |Geographicals| path: "
(KB-ROOT
(|Geographicals| (|Antarctic Regions|) (|Arctic Regions|) (|New Zealand|)))

That is, it does not seem worthwhile to create labels *Antarctic Regions, Arctic Regions*, and *New Zealand* in this situation.

4. ISSUES IN USING SEMANTIC LABELS IN RETRIEVAL

The purpose of this section is merely to suggest that a system of semantic labels may potentially assume the same functions that tree numbers are noted for. Our experimental system of semantic labels is still undergoing refinement in the MedIndEx project, and has not been subjected to scrutiny outside the project. Discussion in this section is strictly hypothetical and in a research context.

It goes without saying that the advantage of a system of semantic labels would be lost if appropriate labels could not be generated automatically, as addressed in the previous section. But in addition, the system would have to ensure the important functions that tree numbers serve, namely, displaying and finding terms in trees, and efficient explosion in retrieval.

Finding terms interactively, where the user enters the term being sought, would seem not to miss tree numbers. However, tree numbers are useful for looking up terms in the published trees. For example, the following entries in the alphabetic MeSH are used for locating UNITED STATES and PENNSYLVANIA in the trees:

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

PENNSYLVANIA
Z1.107.567.875.100.708+
Z1.107.567.875.600.313.708
Z1.107.567.875.600.66.708
Z1.107.567.875.600.513.708

UNITED STATES
Z1.107.567.875.100+

If terms as labels were used instead of tree numbers, if would be necessary to have some other index to the published trees. A possibility would be to use line numbers in the trees, and page:line numbers (e.g., 791:20,30,38) as indexes to trees in the alphabetic display. As currently, "+" would indicate that a node has children. For example:

PENNSYLVANIA
UNITED STATES BY INDIVIDUAL STATE 791:3+
AMERICA 791:20,30,38

UNITED STATES
AMERICA 790:20+

This may make for quicker look-ups than with tree numbers.

Labels and other locations might be displayed in the trees themselves, along with line numbers. The following display shows the current tree number display for pages 790 and 791, followed by a proposed display based on semantic labels. For ease of reading, we have omitted "..." indicators for missing terms.

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop. 47-60.
doi:10.7152/acro.v3i1.12595

TERMS	TREE NUMBERS	OTHER TREE NUMBERS (ABBREV. TO 3 NODES)
U.S.	Z1.107.567.875	
U.S. BY INDIV. STATE	Z1.107.567.875.100	

790

TERMS	TREE NUMBERS	OTHER TREE NUMBERS (ABBREV. TO 3 NODES)
PENNSYLVANIA	Z1.107.567.875.100.708	Z1.107.567. Z1.107.567. Z1.107.567.
PHILADELPHIA	Z1.107.567.875.100.708.820	Z1.433.820
U.S. BY REGION	Z1.107.567.875.600	
APPALACHIAN REGION	Z1.107.567.875.600.66	
PENNSYLVANIA	Z1.107.567.875.600.66.708	Z1.107.567. Z1.107.567. Z1.107.567.
GREAT LAKES REGION	Z1.107.567.875.600.313	
PENNSYLVANIA	Z1.107.567.875.600.313.708	Z1.107.567. Z1.107.567. Z1.107.567.
MID-ATLANTIC REGION	Z1.107.567.875.600.513	
PENNSYLVANIA	Z1.107.567.875.600.513.629	Z1.107.567. Z1.107.567. Z1.107.567.

791

TERMS	LABELS	OTHER LOCATIONS (PAGE:LINES)	CURRENT LINE NO.
U.S.	AMER. U.S. BY INDIV. STATE		19
U.S. BY INDIV. STATE	U.S. BY INDIV. STATE		20

790

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60. doi:10.7152/acro.v3i1.12595

TERMS	LABELS	OTHER LOCATIONS (PAGE:LINES)	CURRENT LINE NO.
PENNSYLVANIA	U.S. BY INDIV. STATE	791:20,30,38	3
PHILADELPHIA	U.S. BY INDIV. STATE	794:13	4
U.S. BY REGION	AMER.		17
APPALACHIAN REG.	AMER.		18
PENNSYLVANIA	AMER.	791:3,30,38	20
GREAT LAKES REG.	AMER.		23
PENNSYLVANIA	AMER.	791:3,20,38	30
MID-ATLANTIC REG.	AMER.		32
PENNSYLVANIA	AMER.	791:3,20,30	38

791

How would a system of semantic labels serve to retain the efficiency of tree numbers as truncated search terms (the explosion capability explained earlier)? This is a very important consideration. A possible solution would be to still use tree numbers internally. A system using semantic labels can automatically generate tree numbers suitable for internal computer processing of explosions. For example, suppose a searcher entered EXPLODE UNITED STATES BY INDIVIDUAL STATE. The system would locate the corresponding system-generated tree number. This number would be processed during retrieval the same way it is now.

If a term had multiple contexts, the retrieval language syntax would still have to provide a way for the user to specify a particular context over-riding the default of all contexts, but this would be done by specifying the semantic label rather than a tree number (e.g., exploding *Bone and Bones* in context of *Musculoskeletal System*, rather than exploding A2.835.232).

5. CONCLUSIONS

This paper has contended that a unified classification is a distinct advantage in systems using large thesauri for indexing and retrieval. We have proposed a new methodology, semantic labels rather than the traditional tree numbers, for representing contextual hierarchies inherent in such a classification. And we have described some heuristics for computing what these labels should be, which is the essence of the problem of implementing this methodology.

We have not found discussion of this problem in the information science literature. Perhaps this is because it also is a computer science problem. But the computer science literature is not replete with papers on encoding thesauri either. Disregarding the assumption in this work that terms have been created through a separate, established process involving expert analysis and consensus, one may well suggest that for the *Bone and Bones* example this be two different concepts, e.g., *Bones* and *Bone Tissue*. Referring to the *Pennsylvania* example, it may be more obvious that in most systems it would not be practical to create different indexing terms for each State in order to reflect the regional versus State context. Nevertheless, the proposed solution to split *Bone and Bones* and the absence of unified classified displays in other major thesauri suggest a view that contextual

ISSN: 2324-9773

Humphrey, S. (1992). Contextual Hierarchies in Classification Schemes. 3rd ASIS SIG/CR Classification Research Workshop, 47-60.
doi:10.7152/acro.v3i1.12595

hierarchies are to be avoided because they are messy to represent. But we contend they are ofte justified, and that methods should be proposed and developed to more easily manage and use the

We particular welcome this workshop as an opportunity to sound out the classification research community on this problem as to its significance and the solutions we have proposed.

ACKNOWLEDGMENT

Thanks to De-Chih Chien of Management Systems Designers, Vienna, Virginia, for his work in programming the algorithms developed for our system of semantic labels for representing contextual hierarchies.