

Searching a Music Database with Semantically Organized Vocabulary Sets Compiled from a Thesaurus of Library of Congress Subject Headings

Fred Rowley

ARIS, Anderson Rowley Information Systems

James D. Anderson

**School of Communication, Information and Library Studies
Rutgers, The State University of New Jersey**

Harriette Hemmasi

**Laurie Music Library
Rutgers, The State University of New Jersey**

1. SUMMARY

Vocabulary sets of semantically related words and phrases were assembled and applied to searching bibliographic records in a music database. The sets were created from a comprehensive selection of Library of Congress Subject Headings concerning music, which were converted to an online thesaurus. The paper describes techniques and rationale used create, name and organize the vocabulary sets, which may be permanently stored as pre-packaged queries for later reuse by various users searching different music databases.

2. OBJECTIVES

As a result of experiments described here, we hope to be able to construct a prototype classification structure—a kind of semantically organized computer index—from which end users may select a variety of pre-designed queries used for focusing their searches in a music database. Each query will correspond to one or more semantic premises or arguments contained in the classification structure, which will appear on the computer screen like a menu displayed in a succession of linked windows. End users will select queries with the objective of “modeling” a portion of the music database—an area where they expect to find their ultimate search targets.

Database queries themselves, will be sets containing multiple instances, even hundreds or thousands of them, which correspond to the semantic premises under which the sets are organized. Because multiple queries (and thus multiple premises) will be available, queries may be “played” against each other using “and”/”or” set operations similar to those available in relational database technology. Users will be able to model areas in the music database corresponding to compound, multi-faceted concepts.

Our current experiments utilize Library of Congress Subject Headings of relevance to music. Subject headings appear as instances in query sets. The query sets containing subject headings may be used to search appropriate topical fields in bibliographic databases where LC subject headings also appear.

We are aware of studies [1] which indicate that subject headings, tendered on an "exact match" basis are not particularly effective for conducting conventional database searches. We have not yet attempted to measure the effectiveness of our system which is not exactly conventional, but we realize that something beyond "exact match" searches is desirable and also possible. One opportunity exists because LC subject headings are composed using a small, controlled vocabulary. We have made it possible to include appropriate instances of the modular LC vocabulary—subterms and words found within LC subject headings—as members of the query sets, in addition to full-length LC subject headings which they comprise.

3. BACKGROUND

James D. Anderson and Fred Rowley have been working on developing and applying end-user thesauri for searching in bibliographic databases [2]. Their experiments have employed techniques which generate end-user queries from online thesauri of their own creation. In this case however, the "thesaurus" is not of their own design, it is what might be called a "pseudo thesaurus", because it is exclusively composed of Library of Congress Subject Headings arranged in hierarchies according to the Library of Congress. The headings were selected for their relevance to a single subject domain, music.

These experiments began after Harriette Hemmasi, who is head cataloger at the Laurie Music Library at Rutgers received a grant from the Council on Library Resources to explore possibilities for building a Music Thesaurus. Part of her project converted a substantial number of LC subject heading records appearing in the Soldier Creek Press publication "Music Subject Headings" [3] into machine readable form. These data were combined with updated LC Music Subject Headings obtained from recent MARC records and they form the basis for our "pseudo thesaurus" used for these experiments.

The word "pseudo" is used here because the Library of Congress has converted its Subject Headings using a database structure which is consistent with traditional thesaurus design. Only a few of the LC headings however, are appropriately organized into traditional thesaurus hierarchies, and these hierarchies are often partial and incomplete. Despite these shortcomings, the hierarchical arrangement of the LCSH database nevertheless enabled us to compile the selected music headings into an online thesaurus using ARIS (Anderson Rowley Information Systems) software.

As a result, we produced an LCSH "thesaurus" having about 10,000 music descriptors fully cross referenced with scope notes, examples, related, equivalent, broader and narrower terms. There are also built-in cross references between many subject headings and their LC Call Numbers. We wish to reemphasize here, that the thesaurus produced is far from ideal. Part of the reason for this is that LC headings are pre-coordinated terms and are likely to represent compound concepts rather than the single concepts usually described by thesaurus descriptors.

We found many of the music subject headings to be complex combinations of multiple concepts indicating, for example, the several instruments for which a musical score is arranged. It was not impossible to find five or more instruments listed in such headings together with additional descriptive information (see Figure 1).

Figure 1

Examples of Individual Pre-coordinated LC Subject Headings
Pertaining to Music:

Concertos (Jazz Trio)
Harpsichord, guitars (2), harp, double bass with orchestra
Sextets (Piano, horn, trombone, trumpets(2), tuba), Arranged
String Orchestra Music
String Orchestra Music, Arranged
Symphony Orchestras [May Subd Geog]
Wiener Konzerthaus (Vienna, Austria)
Wit and humor, Musical

Their complexity makes LC subject headings notoriously "hard to use" and "hard to find". Locating exactly the "right" LC Heading from among thousands of complex descriptors is one of the daunting tasks facing library catalogers. The difficulty of finding LC subject headings was one of the reasons we compiled our LCSH thesaurus. We hoped to develop software making it easier to retrieve pertinent subject headings.

We felt that a hypertext-like display might be useful. The ARIS program provides two onscreen fully-linked thesaurus projections which act something like hypertext. One is a hierarchical display and the other accesses thesaurus records from an alphabetical index. As a consequence of modifying the alphabetical display, our semantic set experiments began in earnest.

4. THE SEMANTIC SETS

a. Organizing Semantic Sets of Subject Headings

The reason we modified our online thesaurus display was to enable catalogers to conveniently gather together various subject headings of their choice. It became immediately clear however, that it was also easy for us to organize subject headings under various themes or "semantic premises". Some themes were traditional "lexical" ones like you might find in a thesaurus, but others were not.

For example, we could create an exhaustive set of headings having a "semantic theme" like SACRED MUSIC. The sets could be fashioned to include all the possible music headings we could find which related to the theme. But we could also organize headings using personal or subjective arguments. Groups like LOUD MUSICAL INSTRUMENTS or FAVORITE INSTRUMENTATIONS could also be formed. This is not to imply that these sets should be formed, but merely that it was possible to organize sets under a wide variety of themes and purposes, whether in good or questionable taste.

We found we could create "general sets", much like broad categories in a thesaurus. These were often large groups containing all the instances we could find in our collection of 10,000 subject headings which corresponded to a premise like MUSICIANS, for example.

Once a general set was organized, we could then break it down into several smaller sets each having narrower, more specific meanings than the parent. Also it was possible to combine terminology from various parent groups so that multi-faceted concepts and topics were represented.

MUSICIANS: PLAYERS OF MUSICAL INSTRUMENTS; MUSICIANS: COMPOSERS; or MUSICIANS: PLAYERS OF POPULAR WESTERN MUSICAL INSTRUMENTS, could all be composed from the parent set MUSICIANS, for example. While the subsets could parallel conventional thesaurus relationships based on lexical considerations, it was also possible for them to represent unconventional arguments too, even though their parent sets reflected conventional ones.

We programmed our software so that there was no inconvenient limit placed on the number of sets which could be produced, nor on the number of headings which occupied a set. We soon found ourselves producing hundreds of sets for hundreds of different premises.

b. The Sets Applied as Queries

Once sets were organized, we could apply them for searching through the LC Subject Headings Thesaurus. The sets could be used to identify subgroups within subgroups of headings, thus multiplying the search result potential while maintaining the focus on or near the original semantic premise. For instance, we had grouped together all subject headings which corresponded to the premise "MUSICIANS: INDIVIDUAL PLAYERS OF MUSICAL INSTRUMENTS". We had also created a set of subject headings for "MUSICAL INSTRUMENTS: PLECTRAL" which contained the names (which were also subject headings) of all the stringed instruments we could find which are strummed—banjos, guitars, harps, lutes, zithers, etc.

We searched "MUSICIANS: INDIVIDUAL PLAYERS OF MUSICAL INSTRUMENTS" using the semantically organized heading set "MUSICAL INSTRUMENTS: PLECTRAL". The computer looked for any instance of a plectral instrument occurring in a subject heading from the "players" set. The search revealed many instances: banjo players, guitarists, lute players, harpists, and the like.

In an obvious extension of using heading sets for searching the thesaurus, we programmed additional software which employs the sets as queries for searching a music database. The database contains bibliographic records that include topical fields where LC subject headings for music also appear. More about this appears below.

c. Naming the Sets

The importance of naming the semantic sets soon became obvious, since it was very easy for us to produce so many of them. In our first software version, set names were lumped together in a single

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

onscreen list. To select a set, users highlighted its name and “clicked” on it. Without careful naming we found it was difficult to retrieve the “right one” from a large group of names.

As time went on, our set name list grew bigger. We then discovered it was possible to inadvertently produce a number of sets for the same concept. Although we had safeguards against producing duplicate set names, we realized that the same semantic concept could often be described in a variety of ways, from different points of view. Subtle differences in word arrangements of set names sometimes lead us to create duplicate sets, which began spoil the crafted quality of our name list. We needed to devise a naming system which accounted for these different viewpoints, otherwise we would eventually be manufacturing redundancy and ambiguity on a grand and chaotic scale.

A desirable feature in a system for organizing semantic set names would be to emulate the same “focusing” qualities found in a thesaurus hierarchy—those which lead readers from general, broader points into progressively narrower, more specific instances of a theme or idea. A formal thesaurus hierarchy cannot be compiled from our sets however, because the subject heading sets we are producing are not constructed like a thesaurus to identify broader or narrower instances. Each query set we produce contains only peers which have no link with each other except for the argument under which they are organized. In our case, any hierarchical structure which organizes query set names, will necessarily remain independent of data contained in our semantic sets.

d. Semantic Set Names as a Computing Device

We reasoned that the way in which sets names were displayed, would strongly influence the effectiveness of any computer program where a large variety of semantic set names were presented. We preferred a display having “focusing” qualities similar to a thesaurus hierarchy. This would provide end users with clear and easy access to the query sets and also show how the sets fitted together into an overall structure. A primary consideration besides the human aspects of the system then, was the computer.

We desired to have semantic set names which also worked efficiently as computing devices. The design we settled on is a string, which the computer “sees” as containing a number of “compartments” representing different levels or stratifications. Compartments in the string contain various descriptors, which together describe the semantic premise for the set. The stratifications are somewhat akin to hierarchies, but not exactly the same in every case. Some examples of semantic set name strings are shown below. Keep in mind the actual names for the semantic sets are the entire string contained between quotes. Boundaries between different levels are indicated with the back slash (\) character.

"MUSIC\INSTRUMENTATION"
 "MUSIC\INSTRUMENTATION\CONCERTOS"
 "MUSIC\INSTRUMENTATION\CONCERTOS\CHIMES&\XYLOPHONE &\VIOLINS (2)
 &\VIOLA &\VIOLONCELLO &\STRING ORCHESTRA"
 "MUSIC\INSTRUMENTATION\CONCERTOS\PIANO &\ORCHESTRA"
 "MUSIC\INSTRUMENTATION\CONCERTOS\JAZZ ENSEMBLE &\ORCHESTRA"
 "MUSIC\INSTRUMENTATION\CONCERTOS\JAZZ ENSEMBLE &\CHAMBER
 ORCHESTRA"
 "MUSIC\INSTRUMENTATION\CONCERTOS\JAZZ ENSEMBLE &\STRING
 ORCHESTRA"

Computers can advantageously handle these compartmentalized string names. The series of such semantic names shown above will breakout into the following stratified structure:

```

MUSIC
  INSTRUMENTATION
    CONCERTOS      <--- 1.
      CHIMES &
        XYLOPHONE &
          VIOLINS (2) &
            VIOLA &
              VIOLONCELLO &
                STRING ORCHESTRA
    PIANO &
      ORCHESTRA    <--- 2.
    JAZZ ENSEMBLE &
      ORCHESTRA
    CHAMBER ORCHESTRA
    STRING ORCHESTRA
  
```

In a computing system, the stratified arrangement shown above will appear on the screen. Users move a cursor to select a query represented by the path leading to the descriptor beneath the cursor. Clicking on "CONCERTOS" (<--- 1.) would activate the query set containing all LC music subject headings representing "Music: instrumentation for concertos".

Moving further down the structure to click on "MUSIC\INSTRUMENTATION\CONCERTOS\PIANO &\ORCHESTRA" (<--- 2.), would retrieve a query set for finding "Music: instrumentation for concertos with piano and orchestra".

The sample structure shown above indicates how general sets and those representing more specific semantic arguments of the same "parent" premise, may appear closely together in the same area of the structure. The cursor, by moving up and down to different levels, also moves between "general" and "more specific" arguments.

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

e. A Stratified Index, not a Thesaurus

It is important to realize that a semantic set name classification structure built in the manner shown above will behave more like an index than an integrated system of definitions similar to a thesaurus or dictionary. A characteristic of this type of index is that its semantic structure can be deliberately fashioned through editorial means to "paint a picture" or "provide a map" which directs end users into areas of their subject. Ideally these will be areas best suited for users to find their ultimate search targets.

A semantic name structure like this then, is meant to provide meaningful, "focused" access to the query sets. There is also an option here to let the structure's design determine which query sets are to be created. And the structure may also be modified when an appropriate premise appears in the raw data— concepts implicit in the subject headings, for example, may be reflected in the structure.

We have been able to create semantic sets produced under a variety of different motives, lexical, syntactical, statistical and others. Our structure, unlike many thesauri, is able to accommodate multiple motives and viewpoints and reflect them in onscreen displays for end users.

Below, in Figure 2 (1..8) are listed some different motives under which subject heading sets might be organized. In each case, a viewpoint is expressed by a succession of descriptors which are listed in different levels. The succession, or "path" is also the semantic name for the set. In the first example of Figure 2, the following 5 set names appear:

MUSIC

MUSIC\MUSICAL INSTRUMENTS

MUSIC\MUSICAL INSTRUMENTS\WIND INSTRUMENTS

MUSIC\MUSICAL INSTRUMENTS\WIND INSTRUMENTS\BRASS INSTRUMENTS

MUSIC\MUSICAL INSTRUMENTS\WIND INSTRUMENTS\WOODWINDS

Unlike a thesaurus, it is possible to integrate all descriptors into the structure, regardless that some are duplicated and would logically conflict with each other in a thesaurus. Integration is assured because the overall semantic set name is a complete, multi-part "path" rather than a single concept.

The effect produced by combining many semantic set name "path" strings, resembles an outline, rather than a thesaurus. Ambiguities are kept separate from each other, locked in isolated compartments, never interfering with their contradictory counterparts. In this system (except for "top terms") there are no stand-alone, thesaurus-like descriptors. Every descriptor is tied to others in a complex statement, like a disjoined twig from a tree, but describing the branch and the trunk to which it may be reconnected. An index consolidated from all the various paths shown in Figure 2, follows in Figure 2 A.

FIGURE 2 Semantic Sets Created Under Various Motives

1. Traditional, lexically driven "thesaurus"-like sets, where relationships are defined by conventional language elements. Example--

```

Music
  Musical Instruments
    Wind Instruments
      Brass Instruments
      Woodwinds
    
```

2. Sets determined by relationships which are metaphorical or represent similes. Example--

```

Music
  Musical Instruments
    That Make Sounds Like
      Animals
      Machines
      Natural Phenomena
    
```

3. Sets which are "syntactical", representing a fixed result derived from a function common to all members in the sets. Example--

```

Music
  LC Subject Headings
    CONTAINING the names of Musical Instruments
      Wind Instruments (all)
      Woodwinds (all)
      Woodwinds (ethnic versions)
      Woodwinds (traditional ones)
      Brass Instruments
      Percussion Instruments
    EXCLUDING names of the following Musical Instruments
      Wind Instruments (all)
      Woodwinds (all)
      Woodwinds (ethnic versions)
      Woodwinds (traditional ones)
      Brass Instruments
      Percussion Instruments
    
```

4. Sets which express a subjective "viewpoint", a personal or corporate value judgment, even a frivolous one. Example--

```

Music
  World's Greatest Musical Instruments
  Best Loved Wind Instruments
  Top ten Woodwinds
  America's Favorite Brass Instruments
  Megahit Strings
    
```

5. Morphologically organized sets which may indicate relationships having to do with linguistic characteristics. Example--

```

Music
  Musical Instruments
    Linguistically Organized
      sharing the root
        "viol"
        "iano"
        "lute" or "leut" or "liut" or "laut"
        "ola"
    
```

6. Statistically or expertly determined or enumerated sets-- "shopping list" queries. Example--

```

Music
  Musical Instruments
    Stringed Instruments
      likely to be used in
        Early Christian Music
        20th Century English Music
        European Folk Music
    
```

7. Chronographically or geographically progressive sets listed in hierarchies that indicate evolution, change and movement through time or space. Example--

```

Music
  Musical Instruments of Ancient times
    Stringed Instruments
    Wind Instruments
  Musical Instruments of the 17th-19th Centuries
    Stringed Instruments
    Wind Instruments
    Mechanical Instruments
  Musical Instruments of the 20th Century
    Stringed Instruments
    Wind Instruments
    Percussion Instruments
    Electronic Instruments
    
```

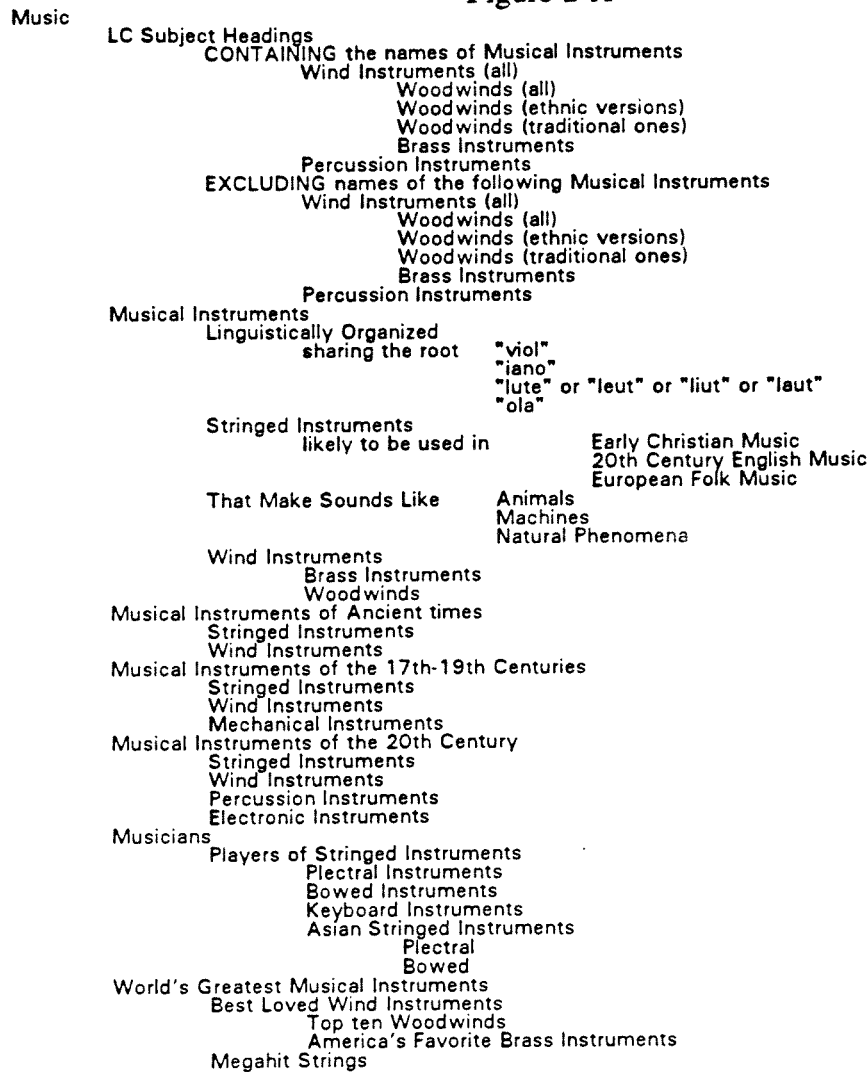
8. Sets determined by "Joining" multiple concepts. These are created by combing sets representing different semantic premises with "AND" or "OR" set operators. Example--

```

Music
  Musicians
    Players of Stringed Instruments
      Plectral Instruments
      Bowed Instruments
      Keyboard Instruments
      Asian Stringed Instruments
      Plectral
      Bowed
    
```


PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Figure 2 A



Of course there are good grounds to criticize the arrangement in Figure 2 A. Please keep in mind that because the structure is an editorial statement, it can be easily changed to please the most discriminating critics. It is the method we are playing with here, not its madness.

f. Focusing Searches by Choosing & Joining Different Sets

What we are really aiming towards from a computing viewpoint, is to affect a semantic search that models the "portion" of the music database which reflects the semantic premise of the query sets chosen. In addition to the search advantages, there would be distinct computing dividends in this approach. Semantic modeling would enable us to concentrate searches only to those records where our ultimate search targets were likely to occur, freeing up facilities.

There are a number of aspects to the "focusing" process which causes "modeling" in the database. One aspect is selecting an appropriate semantic set or group of sets. Another aspect concerns the "joining" and combining of sets into new ones representing new, hybrid arguments.

Semantic focus is sharpened during set selection, by choosing a semantic query set with a narrower premise. Focus is widened by choosing a more "general" set. A smaller portion of the database would be recovered for the semantic query set "MUSICAL INSTRUMENTS-- AFRICAN", for example, than its broader parent "MUSICAL INSTRUMENTS".

Focus may also be sharpened by "JOINING" sets for musical premises. If the ultimate search target was likely to be contained a portion of the database representing "Musical Compositions for African Musical Instruments" for example, a modeling search might be conducted using a semantic set created by joining together the two semantic sets "MUSICAL INSTRUMENTS--AFRICAN" and "MUSICAL COMPOSITIONS".

In this case, an operation similar to a relational technology INTERSECT would be performed to create the new search set. INTERSECT acts like a boolean ANDing of sets, modeling a database area that could be paraphrased by "Database Records for Musical Compositions Written for African Musical Instruments".

Focus can be shifted, spotlighting different, but closely related premises by using combinations of relational UNION and INTERSECT operations. UNION acts like a boolean "OR"ing of sets. A new set representing a different focus, "AFRICAN and ASIAN WOODWINDS", for example, could be modeled in two stages. First by a UNION of the sets "MUSICAL INSTRUMENTS--AFRICAN" and "MUSICAL INSTRUMENTS--ASIAN". Finally, by their INTERSECTION with "WOODWIND INSTRUMENTS".

g. The Role of Vocabulary

LC subject headings are part of a controlled language which might also be viewed as a kind of inventory of musical topics. LC Subject Headings are highly controlled descriptions of subject matter found in musical materials making up vast collections at the Library of Congress and affiliated libraries around the world. We made an informal analysis of the LC subject headings for Music and found they were composed from a core vocabulary of about 3,200 unique words.

We found ways to apply core words and combinations of them for searching through all the headings to find additional headings containing the same components. The results have proved to be effective. In the headings, only a small number of words and terms are available for a given concept. A term like "Sacred Songs" for example, is a modular element. It is used and reused again and again in many subject headings.

All headings containing "sacred songs" can easily be retrieved. A review of those captured can determine which headings which fit a semantic premise, but such an examination may also reveal additional modular terms of semantic value, "sacred part songs" for instance. The newly discovered modular term may be a clue to finding even more headings that match the original semantic concept.

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

h. Vocabulary Resource Sets

Because of the importance of vocabulary, our program allows users to create "Vocabulary Resource Lists" (VRLs) from subject heading sets. VRLs can also be generated from automatic searches through the LC Subject Headings thesaurus hierarchies. Like heading sets, VRLs may also be constructed to reflect a semantic premise.

The Vocabulary Resource Lists behave differently than heading sets. They are expressly designed to search the subject headings but may also be used as database queries. VRLs may contain additional vocabulary not found in the subject headings. This may include "natural language"; terms acquired from other thesauri; foreign language; variants; equivalents; and even morphological stems and substrings.

It is thought VRLs will prove to be important because they will permit use of vocabulary outside the limited words and terms of the LC Subject Headings, enabling searches in database fields other than topical "600"s found in MARC bibliographic records.

Vocabulary Resource Lists act like questions rather than statements. They produce a result, rather than being the result. Instead of representing instances of a semantic premise, they are designed to identify them when a search is conducted. In this respect, words and terms contained in a VRL are operands in an argument rather than the product of the argument.

Because VRL components behave like operands, and because the LC subject headings can be searched for substrings, VRLs may contain substrings designed to identify headings that exhibit certain patterns and characteristics--morphological, grammatical and syntactical ones for example.

1). Morphological Stems

The substring search utility produces matches equivalent to a "truncate right and left" search. This can help locate headings which share morphological stems indicating root semantic relationships. We have not run such searches in the music headings. The following example, discussed in a paper by Dagobert Soergel [4], would identify chemical terms having the root "solution" The common syllable substring for the search is "sol". A search (through chemistry headings, not music!) would produce headings with the words:

Solving	Solution
Dissolving	Solved
Solvent	Solvability
Solvabilize	Soluble
Solubilize	etc.

2). Syntactical and Grammatical Substrings

Whatever the function of a syllable, a substring search may be helpful in revealing groups of headings sharing a common semantic relationship—A search for "ISTS" uncovered many headings which were then classified as musicians:

FlutISTS
HymnISTS
ViolinISTS
GuitarISTS, etc.

A search for two hyphens "--" found headings to which LC subdivisions have been appended. Subdivisions are applied by LC Catalogers using a strict syntax. Subdivisions can be organized into vocabulary resource lists that reflect their syntactic purpose, stated as a semantic premise:

- "Geographical Subdivisions"
- "Solos Subdivisions"
- "History and Criticism Subdivisions"

i. Exploiting the Thesaurus Structure

Despite many criticisms, the LC Subject Headings contain a veritable wealth of semantic organizations and statements. We created a utility which can automatically retrieve selected headings in the thesaurus which themselves pose a semantic premise. The utility obediently follows along all the narrower hierarchical paths projecting from a chosen heading, storing terms found in narrower hierarchies in a VRL. Users have the option of designating the type of term to be gathered—whether narrower (NT), equivalent (UF) or related (RT) terms are to be saved.

Users also have the option to generate "core vocabulary"—which is a separate, but attached list of unique words comprising the terms found. The two lists are merged as one when the VRL is used for searching.

Figure 3a and 3b show how VRLs are automatically produced from hierarchical thesaurus relationships which are part of the LC Thesaurus. 3a shows the LC organization and 3b shows the terms collected as a result of an automatic hierarchy search.

5. DATABASE SEARCHING & MODELING

(A schematic summary of the semantic search and modeling process we have devised is shown in figure 4.)

a. Massive Database Queries

Database modeling would not be practical without the ability to test for hundreds, even thousands of database-query term matches. This must happen in a very short time frame. One-statement-at-a-time searching, the kind found in most "boolean" systems, would perform poorly.

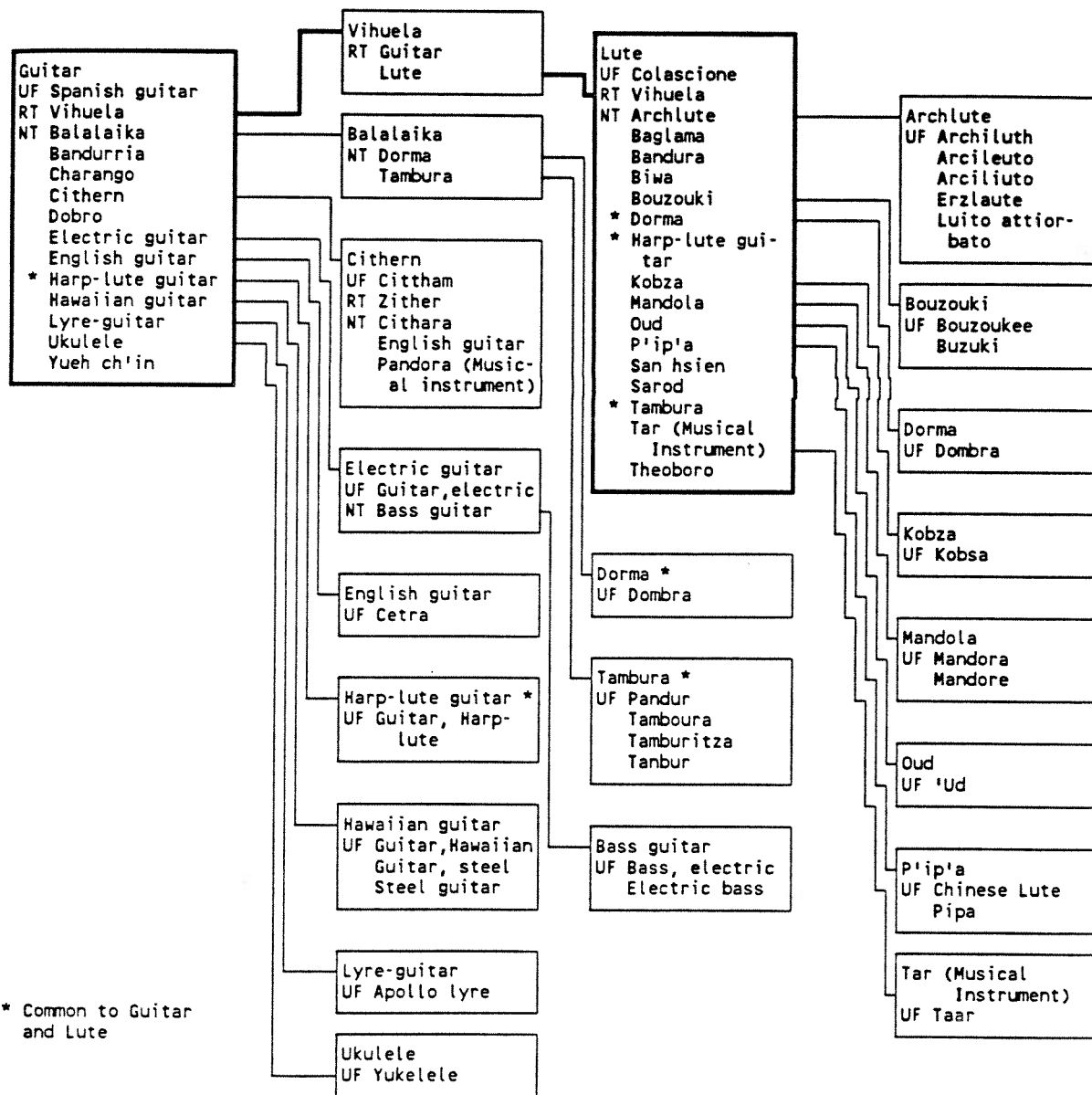
We developed a database search engine that handles massive queries. Searches are conducted against an index prepared for the entire database. As searching progresses, the index can be subdivided to reflect modeling. This means that progressively smaller sections of the index need to be searched as users close in on their targets.

There are no limits to the number of query terms staged. Queries larger and longer than the database contents are permissible. Queries having only a single term are also permitted.

FIGURE 3a.

Guitars And Lutes

Relationships Charted by the Library of Congress in LC Subject Headings



Thesaurus relationships established in the Library of Congress Music Subject Headings are shown above. The links shown can be used to automatically create Vocabulary Resource Lists. Users have the option to designate whether Related (RT), Equivalent (UF), and/or Narrower (NT) terms are acquired. A Vocabulary Resource List produced from the thesaurus is shown in Figure 3b.

FIGURE 3b.

**Guitars And Lutes--Semantically Organized Subject Headings Set
(Created automatically by searching through thesaurus hierarchies)**

<i>'Ud</i>	<i>Buzuki</i>	<i>Guitar</i>	<i>Oud</i>	<i>Tar (Musical instrument)</i>
<i>Apollo lyre</i>	<i>Cetra</i>	<i>Guitar, Harp-lute</i>	<i>P'ip'a</i>	<i>Theoboro</i>
<i>Archiluth</i>	<i>Charango</i>	<i>Guitar, Hawaiian</i>	<i>Pandora (Musical instrument)</i>	<i>Ukulele</i>
<i>Archlute</i>	<i>Chinese Lute</i>	<i>Guitar, electric</i>	<i>Pandur</i>	<i>Vihuela</i>
<i>Arcileuto</i>	<i>Cithara</i>	<i>Guitar, steel</i>	<i>Pipa</i>	<i>Yueh ch'in</i>
<i>Arciliuto</i>	<i>Cithern</i>	<i>Harp-lute guitar</i>	<i>San hsien</i>	<i>Yukelele</i>
<i>Baglama</i>	<i>Citham</i>	<i>Hawaiian guitar</i>	<i>Sarod</i>	<i>Ziher</i>
<i>Balalaika</i>	<i>Colascione</i>	<i>Kobza</i>	<i>Spanish guitar</i>	
<i>Bandura</i>	<i>Dobro</i>	<i>Kobza</i>	<i>Steel guitar</i>	
<i>Bandurria</i>	<i>Dombra</i>	<i>Luito attiorbaio</i>	<i>Taar</i>	
<i>Bass guitar</i>	<i>Dorma</i>	<i>Lute</i>	<i>Tamboura</i>	
<i>Bass, electric</i>	<i>Electric bass</i>	<i>Lyre-guitar</i>	<i>Tambura</i>	
<i>Biwa</i>	<i>Electric guitar</i>	<i>Mandola</i>	<i>Tamburiza</i>	
<i>Bouzoukee</i>	<i>English guitar</i>	<i>Mandora</i>	<i>Tamburiza</i>	
<i>Bouzouki</i>	<i>Er-laute</i>	<i>Mandore</i>	<i>Tanbur</i>	

b. Building a Semantic Search Machine

Our experiments suggest it may be possible to soon build a prototype semantic search machine using techniques discussed here. It would be used for modeling portions of a database where end users might expect to find their ultimate targets. The "Machine" would be distributed like a reference publication on disk.

The search machine would allow users to model areas in a bibliographic database that corresponded to semantic premises contained in a semantically organized index. Premises would be chosen from a menu, joined with each other as desired, then searched against a database, producing subsets in the database representing the selected premises.

Once portions of the database were modeled, users could then conduct careful, precise searches through the modeled areas—searches with greater specificity than the modeling search provided.

Such modeling will not be foolproof. Search exhaustivity will depend on a number of variables. But such searches, even the crudest of them are likely to generate far more exhaustive results than an individual could produce using conventional searching, even after applying a battery of boolean statements.

Since semantic queries would be built around the language of the subject domain, they would be applicable for use against different databases for the same domain. In addition, if semantic sets contained in the machine were prepared with care and by experts, these qualities would be reflected in the search results obtained even by novice users.

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

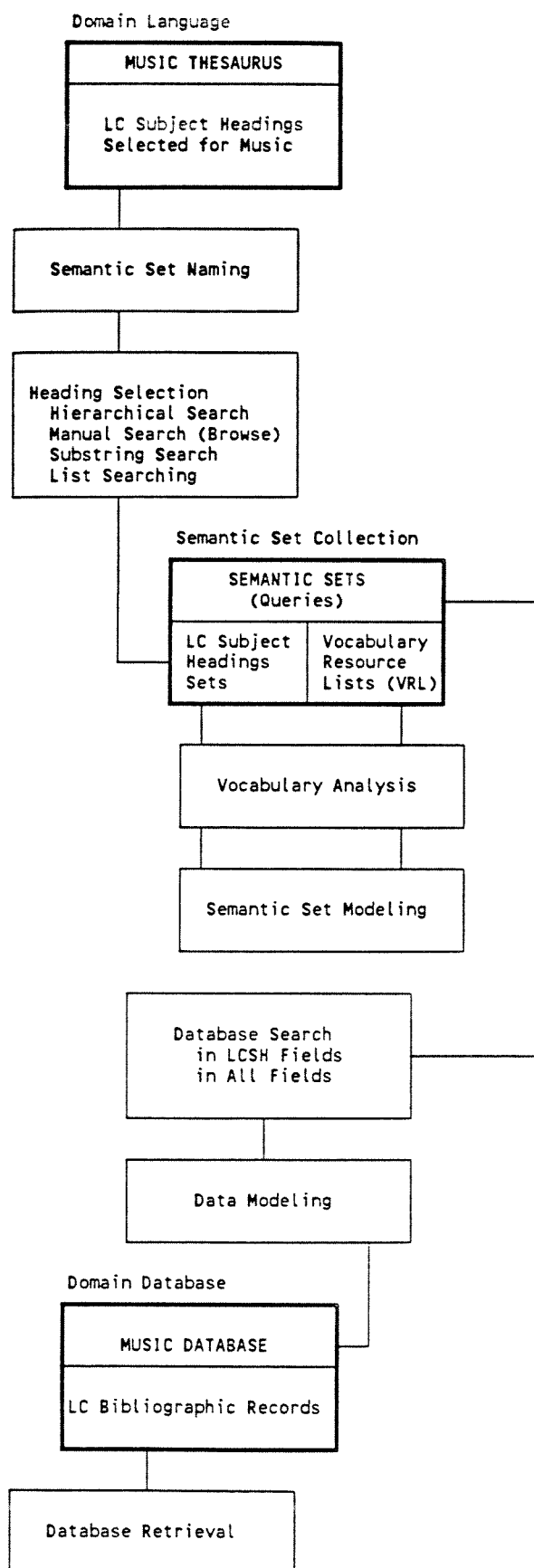


FIGURE 4

An online computer thesaurus consisting of Library of Congress Subject Headings selected for Music provides the raw material from which semantic sets are organized.

The sets are named for musical concepts and also the contexts in which they occur.

Subject Headings which illustrate the semantic premise, are selected for membership in the set. They are instances of the semantic concept for which the set is named.

A large number of semantic sets can be created. Sets are retrieved from an onscreen index, where their names are arranged hierarchically. Two types of sets, one containing subject headings and the other containing vocabulary derived from the headings, are stored on the disk.

By analyzing vocabulary contained in the LC headings and by adding terms obtained from other sources, semantic sets may be modeled which represent new and different concepts.

Semantic sets acting as queries may be searched against the Music Database. Sets consisting of subject headings may be compared with data in LCSH fields, while all fields may be searched using vocabulary sets.

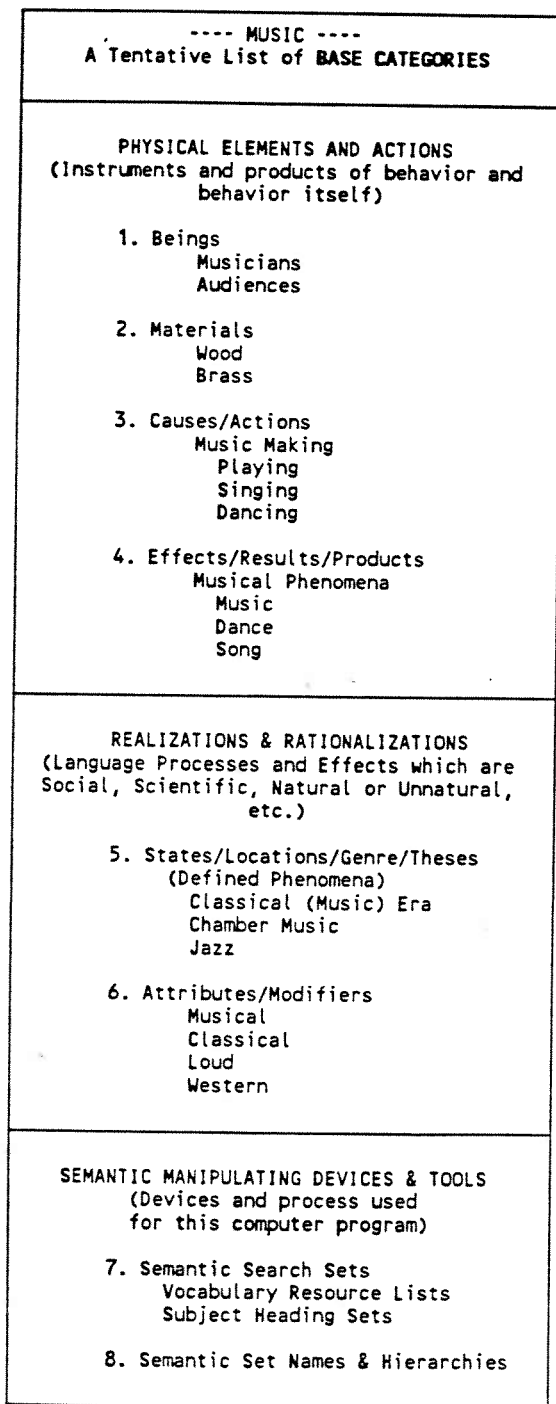
Semantic concepts can be "joined" with one another as in relational database technology. UNION and INTERSECT operations act like boolean OR and AND operators, but on a massive scale.

Once sets in the database are modeled, they may be retrieved for detailed searching.

Ranganathan's
PMEST

Personality
Material
Energy
Space
Time

FIGURE 5



If a semantic search machine is to be built, it may be practical to construct it using a small number of base categories representing contexts. They would indicate the broadest light in which a term, name, premise, or concept is to be considered or understood.

A desirable feature of a system featuring these categories would graphically lead users from the very general, schematic base contexts represented here, into narrower, more specific ideas.

Users would follow along hierarchical paths extending from the categories until they were satisfied that a context represented an environment in which their ultimate target was likely to appear.

If a user chose to search for a concerto for orchestra and piano, then the context would be the sum of a path emanating from one of the pivotal contexts, terminating in piano concertos for orchestra. Perhaps a path like this:

MUSIC
 ..Effects/Results/Products
Musical Compositions
Concertos
for Piano
and Orchestra

The base categories resemble "top terms" in a thesaurus, but when integrated into a semantic search machine they will act much like eight different thesauri, intertwined in a multi-dimensional network.

Base categories include semantic sets, their names, and the hierarchies in which they appear, because these are also active, indexable elements, describing meanings central to the operation of the search machine.

PROCEEDINGS OF THE 3rd ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

A semantic search machine cannot be built until some fundamental classification and editorial problems are solved. There are no tested rules for constructing the intricate semantic hierarchies needed to contain thousands of contexts implicit in a domain like music. Interfaces that handle tens of thousands of menu options need to be developed, and a system of priorities will need to be devised to ensure that concepts important to most users are incorporated as queries.

6. DEVELOPING RULES FOR SEMANTIC NAMING

We have used the word "semantic" in this paper in a very liberal way. Perhaps this is technically inappropriate, but semantic is a good word here because it implies "meaning" in its clinical sense as opposed to its real life role. Scientists who conduct truly semantic research seem to suggest there is a psychological, methodological process of meaning—that individuals carry in their minds a kind of hierarchical, multi-dimensional, but highly portable, memory-based semantic network, which is called upon like a dictionary, to interpret the gist of ongoing conversations.

If so, perhaps this project is leading to the building of a concrete, but artificial model of a semantic network—or should it be called a system of semantic webs? The big problem for us in spinning our webs is their multi-dimensionality. It is somewhat of a computing problem, but more of a classification problem which faces us. It is a problem of communication; of clarity and good taste.

We are looking into the following—Perhaps there are a small number of universal contexts from where searchers can begin their quests, branching from broad contexts into narrower, more specific descriptions that imply the "meaning" they seek. We are devising a skeleton for contexts that may apply in Music. The categories are loosely based on Ranganathan's PMEST model. (see figure 5).

Building semantic headings sets onto our framework is easy and even fun, but there is no assurance the resulting sets will be useful to searchers. Searchers themselves are another great variable. Their motives, intentions, goals and abilities will never be constant, but will always be a factor in the success of any semantic search machine. We must find out what users expect and desire. Search premises which meet their needs can then be created.

Regardless of all the unknowns it seems practical to proceed with our experiments because the language domain represented in the LC headings is well-established and notably stable. Because LCSH also contains a comprehensive inventory of concepts and contexts for musical media, it is a rich source for ideas and concepts which are socially, artistically and professionally pertinent. These same ideas and concepts can easily be transferred into semantically organized hierarchies, possibly of great value for searching.

7. REFERENCES

- [1] Computer and Manual Analysis of Subject Terms Entered by Online Catalog Users. Diane Vizine-Goetz & Karen Markey Drabenstott. p. 166, *Proceedings of the 54th Annual Meeting of ASIS*, 1991. Vol 28. ASIS

- [2] Building End-User Thesauri from Full-Text. James D. Anderson & Frederick A. Rowley. p. 1, *Proceedings of the 2nd ASIS SIG/CR Classification Workshop*, October 27, 1991. ASIS.
- [3] Music Subject Headings, Compiled From Library of Congress Subject Headings. Perry Bratcher & Jennifer Smith; Richard P. Smiraglia, General Editor. Soldier Creek Press, Lake Crystal, MN, 1988
- [4] Beyond Facets: Semantic Roots and Modifiers as Elements of a Conceptual Morphology. Dagobert Soergel. p. 149, *Proceedings of the 2nd ASIS SIG/CR Classification Workshop*, October 27, 1991. ASIS.