National Agricultural Library

Agricultural Research Service
United Department of Agriculture

Informatics and Indexing Branch Manual:
Change to Automatic Indexing (AI)

September, 2012 draft

Table of Contents

1. Introduction

The National Agricultural Library (NAL) produces the AGRICOLA database, (http://agricola.nal.usda.gov/), which is an online bibliographic database that contains over 5 million entries on the topic of agriculture.   AGRICOLA is composed of the AGRICOLA Catalog and the AGRICOLA Index.   The AGRICOLA Index is an abstract and indexing service for agricultural literature and primarily contains technical journal articles and research reports. The controlled vocabulary used for subject indexing of the AGRICOLA Index is the NAL Agricultural Thesaurus (NALT) (http://agclass.nal.usda.gov/agt.shtml).

The assignment of subject terms in the AGRICOLA Index (aka "subject indexing") has been performed entirely by human indexers since 1985. In 2009, annual production of the AGRICOLA Index is limited to 75,000 articles each year with a staff of six indexers and an indexing contract.

1.1 Change to Automatic Indexing

In February, 2010, Dr. Simon Liu became the new NAL Director and developed new strategic goals for NAL. One of these strategic goals is **to significantly increase (400%) the number of articles indexed each year while achieving a baseline quality standard**. To accomplish this goal with in-house indexers, the increase in publisher-supplied data and the acquisition of automated indexing software was sought as the solution. The ultimate goal is to implement the automated indexing software as an automatic indexing system with human quality review, and not just as a machine-aided indexing system for human indexers.

The reasons for this change is many-fold:

a) To increase the number of indexed articles in AGRICOLA from 75,000/year to 500,000/year so the database content is relevant and up-to-date for users.
b) To ensure that all articles in AGRICOLA are indexed with a controlled vocabulary.
c) To ensure that all articles in AGRICOLA have been subjected to a quality assurance procedure.

The Automatic Indexing Project began in June 2010 and manual indexing was shut down in July 2010. Since then, staff has been working over many months to shift to this new process. Over this time, NAL staff has acquired software, learned about natural language processing, learned about linguistics and computational linguistic methods, established workflows, and developed procedures. We have accumulated enough information now that a new Manual is now practical and necessary.

In addition, the new Informatics and Indexing Branch established in October 2012 is not the "Indexing Branch of old". That is, the old Branch was concerned with the selection criteria of journals, the selection of articles within those journals, the creation of citations in the library management system, the subject analysis of those articles and the application of controlled vocabulary terms and category codes to records. The selection of journals/articles and descriptive indexing are no longer functions within the Branch but reside elsewhere in the organization.

The mission of the new Informatics and Indexing Branch:

> The Informatics and Indexing Branch in the Data Production Division is responsible for the administering of programs for the semantic interpretation, subject description, classification and automatic processing of scientific research materials originating from worldwide sources in the subject area of agriculture and its related sciences. The work of the Branch results in the accessibility and retrievability of information in the AGRICOLA bibliographic database, NAL Digital Collections and other NAL information products. The Branch develops terminology tools, controlled vocabularies, thesauri, and other knowledge organizing systems that support semantic disambiguation and automation of text analysis.

Despite all of these changes in the indexing approach, workflows and delivery, the theory and purpose of subject indexing has not changed. Therefore, much information in Section 2 of this document is derived from previous NAL guidelines. However, most of the information regarding workflows and procedures is brand new, and is based on new and anticipated electronic data streams and implementation of the Luxid Software to support the goals and mission of the Branch.

This document is to establish "how" we do subject analysis (index process) rather than giving guidance on "what subjects" we cover in our index. The selection of journals and articles is done prior to the automatic indexing procedure and the policy governing that procedure and the descriptive indexing of those materials is covered elsewhere. Even though automatic indexing will most likely be applied to other than AGRICOLA articles, the indexing of scholarly literature in agriculture is our emphasis at this time.

Despite the change in how we will create our index, the purpose of indexing has not changed in an automated indexing scenario. We still have high expectations for the production of a useable and "value-added" index for users that goes beyond the generation of a "computer-generated by every word present" index.   The next section emphasizes the purpose of indexing and reiterates its significance to the user.

I.   The Purpose of Indexing

The purpose of indexing is to use terms to express the subject matter of items so that users can search with these terms to locate items of interest in an information storage and retrieval system.

Indexing is traditionally seen as a two step-process : 1) determine the subject matter of documents, and 2) translate the subject matter into the terms of the controlled vocabulary. Indexing guidelines usually spend much time on the second step as these are mechanical and are easier to express (e.g., use the most specific term in the hierarchy).   It is this first step, or the cognitive process of determining the "aboutness" of an item, that is usually vague and inadequate in indexing guidelines.

6.4 Retrieval performance of a database

The retrieval performance of a database and its index language is judged by recall and precision measurements.

Recall = Number of relevant records retrieved/Number of relevant records in database X 100%

Recall measures the percentage of relevant records in the database that are retrieved.

Precision = Number of relevant records retrieved/Total number of records retrieved X 100%

Precision measures how many of the records retrieved are relevant.

For example,
You have 10,000 records that are about kudzu in your database.
Search is conducted using descriptors "kudzu" and "weed control"
Result of search retrieves 8,000 records with 5,000 that are relevant.
The recall is 5,000/10,000 X 100% = 50%
The precision is 5,000/8,000 x 100% = 62.5 %

The indexer and the searcher both aim to improve precision or recall or both. However, recall and precision are somewhat inversely related. That is, effort to increase precision can decrease recall and effort to increase recall can decrease precision. Therefore, there is a trade-off between the two measurements while indexing.

The ability to find relevant articles is greatly influenced by the size and scope of the database (or corpus) in which information is being sought.   Agriculture and the AGRICOLA database encompasses a wide-range of subjects, including but not limited to:

- Animal and crop husbandry
- Biofuels and alternative energy sources
- Taxonomy and Biology of organisms
- Economics
- Agricultural Engineering and Design
- Climate change, pollution and the environment
- Hydrology and water management
- Pesticides
- Laws and regulations

Indexing rules and the consistency in their application are essential to ease the burden of finding relevant information for the searcher. Indexers need to be aware of the expectations of searchers. The indexer is to consider "value-added" terms to the index that are not present in the original text that will be of interest to users.

6.5 Document-oriented indexing versus User-oriented indexing

Most books on indexing practice center on the "document-oriented" or "document-centered" approach. This approach focuses on the strict adherence and faithful description of the text at hand. The subject analysis is solely based on the item itself and the indexer is not to add viewpoint to the item. In "User-oriented" or "User-centered" indexing, the indexer considers what questions the user may ask to retrieve the item at hand and considers which terms to describe the item based on the user's perceived viewpoint. Both of these approaches to indexing were used in the previous indexing guidelines for

AGRICOLA. **As we move to automatic indexing, both of these indexing approaches still apply**. Even though the AI software only has the author's text to process, it is our hope to embed our knowledge into the automatic indexing software tools so that context is taken into account and terminology that is applied will align with the perceived needs of users in that knowledge domain.

The NAL indexer has not been traditionally given data about users and their queries; however, the indexer has been trained to think of the user. One advantage that NAL indexers have is that they are professionally trained and usually hold an advanced degree in some subject domain of agriculture. Indexers usually index in an area in which they are familiar, but this is not always the case. An indexer indexing outside of their knowledge domain can still index the item, but it usually means that the text present in the document drives the index term selection. When indexers with subject expertise embed their knowledge of that domain into the automated indexing tools, then these are helpful to achieve a better index in that domain and serve to help novice or non-subject domain indexers to facilitate the indexing process with better results.

6.6 Context-oriented and domain-oriented indexing

Indexers draw on their education and experience to understand the "aboutness" and context of the text. Why is kudzu an important topic to agriculture? Kudzu is a vicious weed that is a threat to agricultural crops and an invasive species that can choke out native plants. Kudzu is a plant, but what role is this plant in the context of agriculture? Why do we care about kudzu? Even if the text is simply a description of the biology of the kudzu plant, should the term "invasive species" be assigned? Should the term "weed biology" be assigned? The indexer has been trained to consider the user and ask the question for decision making: "Will the user be delighted or disappointed if they retrieved this item when searching on 'invasive species' or 'weed biology'?" With human indexing, humans will differ in how they index this item. Inter-indexer and intra-indexer variation in the application of index terms is an accepted byproduct of human indexing. Some of this variation is inherent in the fact of being human. The selection of index terms relies on the Indexer's knowledge of the subject as well as how they were trained to index that scenario. If the indexing guidelines are general or vague, the guidelines will fail to be interpreted consistently by indexers and trainers. Such inconsistencies would not exist in automatic indexing if a rule to always index "weeds" when "kudzu" is present. "But beware Will Robinson!" If this is a strict relationship, ("always true"), then this would be present in the thesaurus hierarchical relationship. When it is not present this way, then we can assume that it is not always true. Kudzu is always a plant, but is not in all situations an invasive species. If that relationship was always true, then kudzu would be a narrower term than invasive species in the thesaurus. When developing automatic indexing rules, caution needs to be taken to not add information that is extraneous or implying a specific role that would not be true in all situations.

Context is also interpreted not only by the text in the item being indexed, but by the source journal for that article. If one is indexing about fish feeding, does one automatically assign "aquaculture" if it appears in the "Aquaculture" journal? Is this indexing for implication or is this simply providing context for the user? How often is this type of "general term" assigned by indexers? A study of the indexing terms for specific journals and the typical articles contained in journals could support our development of mechanized indexing rules. How valuable are these terms to users? If "fish culture" is assigned, which is a narrower concept of "aquaculture" in the thesaurus, there is not much value in assigning the broader term in addition to the specific. In a system where the thesaurus is integrated into search, a query on "aquaculture" will retrieve this article because of the relationship between "fish culture" and

"aquaculture" even though "aquaculture" is not explicitly assigned to this article. In the automatic indexing scenario, we can embed rules so that the most specific term is always applied.

 With automatic indexing, we can codify the knowledge of an educated indexer considering the user, and this knowledge can be incorporated into the AI system. Automatic indexing allows us to explicitly state our indexing practice for each scenario, to codify it in rules and have this request performed consistently as possible by the AI software.

II.   Patterns in Journals – "Journal Fingerprinting"

 As experienced indexers, we know there are patterns in journals.   We know that there are some journals where we use the same category codes over and over. In some journals, we find ourselves repeatedly using the same NALT terms and branches of the thesaurus. Can we somehow find a way to "characterize" each journal? Is there a unique "fingerprint" for a journal and can we use this fingerprint to our advantage when it comes to automatic indexing?

See separate document "Journal Fingerprinting for Automatic Indexing" for details on how we can discover the patterns of content within journals and systematically characterize them so that we can better devise strategies to index them.

III.   Some assumptions/decisions that influence "how we index":

1. We will not be implementing up-posting in the AGRICOLA database.
2. We will be implementing hierarchy expansion in search in AGRICOLA.
3. We will be implementing automatic inclusion of synonym rings for search.
4. We will no longer assign the alphanumeric category codes. (There is a separate document which discusses a classification system based on language, if so desired for searchers).
5. The order of descriptors is not important.
6. We will no longer assign genre terms, such as "literature reviews" and "bibliographies" and mix these with subject-oriented terms. These genre terms were deleted in 2011 and are not present in the 2012 thesaurus).
7. Our old indexing rule: Do not apply geographics in certain situations (e.g., do not apply for greenhouse grown crops). This is so subjective that it will be hard to write a rule for this. Also, it is unknown how stringent human indexers applied this rule in manual indexing practice.
8. We will continue to use NALT:
     a. Topical NALT terms (will appear in the 650 MARC field of Voyager).
     b. Geographical NALT terms (will appear in the 651 MARC field of Voyager).
     c. Non-descriptors will not appear in the record, but will be "translated" to their NALT descriptor (and this term shall appear in the 650 or 651 fields of Voyager).
     d. We will no longer continue to use the 653 for non-NALT terms, (unless a method is devised to derive highly ranking terms from Temis-provided skill cartridges).

IV.   **Automatic Indexing guidelines can be organized into three main sections**:
1) Configuration of the AI software

2) Subject-specific guidelines
3) General Guidelines for all subject areas

Configurations of AI software. For those requiring actual numerical values, the actual numerical values are not given. These starting values will be determined with the aid of the vendor during the NAL scenario analysis.

a. Max / Min # of terms assigned - minimum of 3, maximum of 15
b. Numerical rank, minimum value
c. Weighting of terms by position in text (e.g. title)
d. Weighting of terms by subject
- More weight to:
    1) Foods and products
    2) Organism names – scientific and common
    3) Organizations (?)
    4) NALT descriptors and non-descriptors
- Less weight to:
    1) Methodology terms
    2) Chemical terms not in title (?)
    3) (Are there other interdisciplinary terms we want to give less weight to?)
e. Stop word list – need to see vendor stop list as well as supply our stop list on s: drive.
f. Hierarchical specificity – For now we are assigning most specific in hierarchy with the argument that a user's search on the broader term will retrieve the more specific in future AGRICOLA when the thesaurus is integrated into search.
g. Use of cross references in NALT to generate NALT term assignment
- USE/UF
- USE "and type"
- Hidden labels
- Typographical errors mapped to correct spellings, e.g. broccoli.
h. Assignment of NALT terms
i. Assignment of relevant terms not in NALT but are given a high rank by the software (such as that which is assigned by the Luxid Relevance Finder Skill Cartridge). Assuming that higher rank is going to give higher importance of that subject to the searcher of documents. Assess whether these are significant and if we want them added to MARC 653.

Subject indexing guidelines for specific topics

j. Select a few journals in one topic (see results from Journal Fingerprinting…journal "clusters". Thinking we want to keep these topics small, but need to experiment.
k. Select a group of indexers that are knowledgeable about indexing of that topic.
l. Indexers to review any relevant data from the Journal Fingerprinting idea.

m. Indexer writes down the conventions and patterns in indexing practice
- E.g., Assignment of role terms
- E.g., disease and causal agent

n. Vet these conventions in the wider indexer audience.

o. Consider the creation of knowledge bases, cross references in thesaurus, or rules that can enforce these rules.

General indexing guidelines

General rule-making guidelines:

1. Indexers will write rules for *NALT* descriptors in a manner that is consistent with their broader terms (BT).
2. Indexers will write rules for NALT descriptors in a manner that is consistent with their scope notes (SN), if applicable.
3. Indexers will use a pre-coordinated descriptor for a concept in lieu of using two or more descriptors.
4. For "role" type terms, Indexers will use caution to write rules that are consistent with an "always true" relationship. Role term rules should be researched extensively and discussed with all indexers to ensure conformation to the always true relationship.
5. In the case where a thesaurus entry can be made, use this option over creating a rule. For example, for the text of "red meat quality", add cross reference to the thesaurus: USE "red meat" AND "meat quality".
6. If the meaning of an NALT term is unclear or too restrictive, please bring these to the attention of the thesaurus group.
7. Do not write a rule for a "stop word"…that is, a term that is "tertiary" and not likely to be assigned consistently by all indexers and is on the stop word list. If you have a candidate term for adding to this list, send it forward.

These were the general guidelines given when indexing/editing the 5K acceptance test articles

What types of changes to make:

1) Add terms that need to be added to convey major points of the article.

a) Add complex concept terms that are missing, e.g. bacterial diseases of plants, fire ecology, weed biology, crop weed competition.

b) Add geographics if they are missing and significant

c) Add role terms if they are clearly a focus of the article

d) Add missing major concept terms.

2) Delete terms that are

a) Delete "stop words" e.g., duration, research, diameter, measurements

b) Delete role terms that are extrapolated.

c) Delete methodology that is not the subject of investigation.

d) Delete terms that are used outside the term meaning (misapplied terms)

e) Delete "tertiary" terms that are very minor in use in the article.

f) Delete terms where you do not see any evidence for them in the title and abstract.

Checklist of Indexing:

WHO - the "players", such as organisms, products, organizations, social groups, communities, cultivars, hybrids, genes. Other agents, such as chemicals, pesticides and drugs. "Roles of the who", when the role is significant, such as "invasive species" or "plant parasitic nematodes"

WHAT - actions such as treatments, agricultural practices, and phenomenon - breeding, genetic transformation, brewing

PROPERTIES – acidity, disease resistance

RESULTS or EFFECT - weed control, pest control, fire scars

WHEN - time of year, preharvest treatment

WHERE – geographical terms are indicated in thesaurus and placed in 651

HOW - methodology - only index when subject of investigation is the methodology, not mere mention.

SPECIFICITY – index to the most specific. Ignore the "rule of 3".

EXHAUSTIVITY – Index what you see in the title and abstract.

SECTION 3 – Subject Analysis

This section describes the two steps of subject analysis:

1) identifying concepts for indexing

2) selecting concepts for indexing.

3.1 - Identifying concepts for indexing

The following is adapted from: *Guide to indexing for AGRIS and CARIS,* by Monique Bonnichon. It is one example of an aid that can help indexers conduct a thorough analysis of a document. <u>Not every document will have each type of concept nor will every identified concept be selected for indexing</u> (Section 3.3).

The principle underlying the analysis of content is identification of <u>concepts</u> rather than <u>words</u>. The concepts represent the knowledge conveyed by the document.

In order to analyze a document completely and precisely, it is advisable to identify its various concepts based on their <u>function</u> in the logical structure of the document:

- object

- action

- point of view

- agent

- means

- location/environment

- other results

In practice, this examination becomes a series of questions the indexer asks about the document, until all possibilities of interrogation are exhausted. The entire document must be examined during the analysis process.

Except for the object, none of the functions must be included in the document. The subject does not have to include an action, an agent or the employment of the means, etc.

Identifying the function of a concept will sometimes seem arbitrary or difficult. However, it's not necessary to distinguish an action from a process, or a point of view from an environment. An error in the choice of function has no effect on the quality of the analysis. What is important is to identify <u>all</u> the concepts.

<u>OBJECT or CENTRAL SUBJECT</u>

It is necessary first of all to determine the things or phenomena, i.e. the **object(s)** on which the ideas or knowledge of the document are centered. Every document has an **object**. The **object** of a document may relate to one or more of the following groups of concepts:

1) Those corresponding to the concrete entities that are produced, used or processed in the course of the various agricultural activities, e.g. plants, animals, agricultural products or foods, fertilizers, pesticides, fuels, equipment, materials, food additives, drugs for animals, etc.

2) Those corresponding to abstract entities: phenomena, conditions and/or activities affecting or involved in agriculture, e.g. an agricultural activity; a method, operation or process used in agricultural activities; a situation; economic, social, legal, environment or climatic conditions; a disease; etc.

These document titles are given as examples:

– A description of **lychee cultivars**

– Observations of **automatic teat cup attachment** in an automatic milking system

– Work performance of the **camel** as a draught and riding animal

– Sizing **horizontal silos** for safety and strength

– The **cleaning process**: some definitions of terms

– The **international commission on microbiological specifications for foods** (ICMSF)

– **Food security**: an ODA view

– **Agriculture** in Britain

– **University education** in agroforestry

– **Remote sensing applications** to human and animal health

One must then make sure that these concepts have been completely identified by checking whether the object relates to:

      • all or part (components or elements) of these concepts (Section 4.1.1.c, item I)

      • particular characteristics or attributes of these concepts or their components

– **Work performance** of the **camel** as a draught and riding animal

– Infection of **aerial parts** of **potato plants** by Colletotrichum coccodes and its effects on **premature vine death** and **yield**

– Comparison of the **immunotoxicity** of **propanil** and its **metabolite, 3,4-dichloraniline**, in C57B1/6 mice

– Comparative **physicochemical and functional properties** of **cassava starches** obtained by conventional and enzyme integrated conventional techniques

– Comparative study on **fat globule membrane** of **buffalo, cow, goat** and **sheep raw milk**

– Leguminous browse supplementation effect on the **agronomic value** of **sheep and goat manure**

– Effects of urea fertigation of apple trees on **soil pH, exchangeable cation** and **extractable manganese** in **sandy loam soil** in New Zealand

– **Cooking quality and nutritional attributes** of some **newly developed cultivars** of **chickpea (Cicer arietinum)**

ACTION

The object may be partially or totally subject to or affected by an **action** (activity, operation, phenomenon, etc.)

– **Filtration** of soy sauce by ceramic membrane

– Effects of **urea fertigation** of apple trees on soil pH, exchangeable cation and extractable manganese in sandy loam soil in New Zealand

– Inhibition of degreening in the peel of bananas **ripened** at tropical temperatures

– Sperm precedence in Helicoverpa zea (Lepidoptera: Noctuidae): response to substerilizing dose of **radiation**

-(can add verbs and other parts of speech as well as examples, The crop was harvested early in the morning. – harvesting.)

If a concept or concepts are recognized as corresponding to **actions**, it is then essential to make sure that they have been completely identified by checking whether the action is considered:

• in its entirety or in its parts (components or elements)

    • in relation to specific parameters or conditions.

– Inhibition of degreening in the peel of bananas **ripened** at **tropical temperatures**

– Sperm precedence in Helicoverpa zea (Lepidoptera: Noctuidae): response to **substerilizing dose** of **radiation**

POINT OF VIEW

It is essential to ascertain whether the object of the document is considered from a particular **viewpoint**, in relation to a particular objective or a particular context. (See the section immediately following this excerpt.)

– Work performance of the camel as a **draught** and **riding animal**

– Sizing horizontal silos for **safety** and **strength**

– **Food security**: an ODA view

– **Systematic** and **palynological study** of the genus Whitfieldia (Acanthaceae - Whitfieldieae) in Central Africa

– Animal surgery: **veterinary-ethical considerations**

AGENT

An **agent** or **agents** can affect the object or intervene during an action. It must also be ascertained whether the document is concerned with components or elements of these **agents** or with certain of their characteristics.

– Infection of aerial parts of potato plants by **Colletotrichum coccodes** and its effects on premature vine death and yield

– Deacidification by distillation using **nitrogen** as stripper. Possible application to the refining of edible fats

– Collective human food poisoning by **clenbuterol residues** in veal liver

– Cation exchange capacity, induced in calcareous soils by fertilization with **manure**

MEANS

Special **means** may be used to study the object or perform actions:

- materials, machines or equipment

- processes, methods, or techniques

Just as in the preceding groups, it must be ascertained whether particular aspects (elements, components or characteristics) of these means are considered. In addition, careful consideration must be given to **means** during the selection of concepts for indexing according to the guidelines for **Methodology** (Section 3.3.2).

– Filtration of soy sauce by **ceramic membrane**

– Nursery rearing of Penaeus monodon (Fabricius) using **suspended (hapa) net enclosures** installed in a pond

– Rapid Salmonella detection by a combination of the **conductance** and **immunological techniques**

– Determination of chemical constituents in processed [Japanese] green tea **near infrared analysis**

– Evaluation of bull semen fertility by **homologous in vitro fertilization tests**

– Portable-chain saws, Automatic chain break and cutting equipment. **Operator's safety test**

– A systematic study of Machaeranthera (Asteraceae) and related groups using **restriction site analysis of chloroplast DNA**

ENVIRONMENT/LOCATION

It must be ascertained whether the object, action, agent or means are considered as part of a particular location and/or environment, and whether factors or conditions in this environment or location are studied

in a specific way. In addition, careful consideration must be given to **environment/location** during the underline selection of concepts for indexing according to the guidelines for **geographical terms**. (Section 3.3.1, item H)

– Effects of urea fertigation of apple trees on soil pH, exchangeable cation and extractable manganese in sandy loam soil in **New Zealand**

– Nitrogen mineralization after leys ploughed in **early or late autumn**

– Agriculture in **Britain**

– Brewed in the Pacific Northwest: a history of beer-making in **Oregon** and **Washington**

Concepts relating to the environment and location are especially important for agricultural databases. It must be borne in mind that a spatial location may also imply particular physical, socioeconomic or cultural conditions that must be identified in the analysis:

– Level of penetration of artificial insemination in some rural areas in the **Assiul Governorate (Egypt)** (The study relates to the attitude of the peasants of the Assiul Governorate with regard to the acceptance of this new technique, so particular sociocultural conditions are also involved.)

– Drought-resistant trees and shrubs for dry plantations in **Bahrain (**Here Bahrain also has the meaning of climatic zone.)

OTHER RESULTS

It is also essential to check whether results other than those noted in the preceding evaluation have been obtained.

Consider concepts that are important, based on experience and judgment, even if they are not specifically stated by the author:

To identify these concepts it may be helpful to determine why a document was selected for the AGRICOLA database.

Another approach is to determine the role or context of organisms, products, etc.

Example:

> A document discussing pathogenicity of a parasite for a pest insect should be indexed with the concept "biological control agents" even if the author does not use this term.

Examples of context or roles include:

– Plants: "crops", "dye plants", "host plants", "pasture plants", "weeds"

– Animals: "food animals", "game animals", "working animals", "pets"

– Insects: "pollinating insects", "insect pests", "aquatic insects", "storage insects"

– Foods: "chilled foods", "dietetic foods", "infant foods", "snack foods", "staple foods"

– Health: "animal health", "human health", "plant health"

– Equipment: "cleaning equipment", "heating equipment", "application equipment"

(Section 4.1.1.d provides additional information for locating these terms in the *NALT*.)

This level of analysis, based upon indexers' education, training and experience, provides important and extremely valuable information to AGRICOLA users as to the "agricultural significance" of documents. It is not, however, the responsibility of indexers to attempt to determine every possible agricultural use or application for the information they are analyzing. *Role/context concepts must apply to the research, discussion and/or circumstances of the document at hand* (see Section 3.3.1, item G).

Example:

> A rabbit may be a pest, a pet, a source of food, a source of fur, etc. However, in a discussion of electrified fencing to deter rabbits from eating crops, "rabbits" would only be "vertebrate pests."

> In an article on "Evolution of plasma parameters in rabbit does in relation to their physiological condition and with restriction of the quantity of feed provided," an indexer would be in error to assign any of the roles sited in the previous example. The concepts of "animal nutrition" and "animal physiology" however are applicable.

3.3 - Selecting concepts for indexing

All of the concepts deemed necessary to adequately represent the document's significant aspects should be selected for indexing (exhaustivity). These concepts should be represented at the level of detail consistent with the treatment of the subject in the document (specificity).

It is vitally important to keep in mind that indexers must select concepts not words. Always remember that the same concept can be expressed by authors using many different words or combinations of words. (see Section 2.4.1 - *The NAL Thesaurus*)

3.3.1 - General considerations

A) Give high priority to concepts the author considers important as evidenced by the manner and frequency of their treatment in the document. Choose concepts which are substantively discussed, not just mentioned in passing.

Consider the treatment of the subject in the document:

>   Data: substantial, cursory, detailed, and/or summary
>
>   Discussion: specific, general discussion, lengthy, and/or brief
>
>   Review: extensive or cursory

B) Carefully select concepts that are important, based on experience and judgment, even if they are not specifically stated by the author.

C) Select only information which warrants the time and (possible) expense of retrieval. Consider whether or not the inclusion of a concept will satisfy or mislead an AGRICOLA user. Place yourself in the position of a reference librarian or end user searching the AGRICOLA database and ask: Would you want to retrieve this document when searching this topic? (see item F below)

D) Select significant negative results.

Example:

> In a study of the effect of iron deficiency on peroxidase isoforms in sunflowers, the author(s) found "...while ascorbate peroxidase activity significantly decreased, <u>no change was observed</u> in either soluble guaiacol or syringaldazine-dependent unspecific (POD) peroxidases activity following iron starvation.

The finding that there was no change in these two enzymes is a significant result and these concepts should be indexed.

E) Treat equally concepts which are treated equally in the document.

Example:

> In a project listing four different dinitroaniline herbicides, represent each of them or use a concept that represents the group, such as "dinitroaniline herbicides". Do not select only one or two of the four.

F) It is important for indexers to focus on choosing concepts that convey the subject(s) or objective(s) of selected documents. It is not necessary to index the specific, routine measurements of these subjects or objectives.

Example:

> If a study examines the impact of a specific feed supplement on hog growth rate and lean meat, the concepts selected for indexing must include: "feed supplements", "animal nutrition", "swine", "growth", "pork". It is not necessary, to represent the specific measurements of growth which may include: "weight gain", "length", "feed conversion", etc.

These specific measurements must be indexed, however, when the techniques of their measurement, i.e. methodology is, a major focus of the document. (see Section 3.3.2 - <u>Methodology and techniques</u> below)

G) Do not select concepts which are present only through implication or speculation. Documents usually contain references in the text to related work. In particular, the Introduction may contain a brief literature

review as a background or rationale. The Summary or Discussion may refer to the work of others, possible applications of the current research or plans for additional investigations. Do not be misled into indexing concepts that represent work not conducted as part of the research described <u>in the document at hand</u>.

Examples:

> Indexing for an article discussing a clinical trial on the effect of folic acid supplementation during pregnancy on the incidence of neural tube defects should include "maternal nutrition," even if it is not mentioned specifically by the author(s).

> However, for an article on the folic acid content of breakfast cereals, the concept of "maternal nutrition" would not be selected although it might be discussed extensively in the Introduction and/or Discussion. This would be implication. The concepts selected for this article would include "food composition," "vitamin content" and "nutrient content".

H) Situations in which geographical concepts should be considered include, but are not limited to, the following:

 - there is a geographical term in the title

 - there is a map in the article

 - a survey has been carried out

 - field research has been carried out

 - flora or fauna of a region has been studied

 - new taxa or other taxonomic information are a topic, i.e. geographical distribution

 - organizations or institutions are a significant topic

 - the article is biographical in slant

 - legislation, patents

Consider the suitability of geographical concepts when any of the following are subjects of a document:

 - ecology, including pollution

- economics

- education/training

- epidemiology

- history

- human ecology

- land and/or water resources

- legislation, jurisprudence, or politics

- meteorology/climatology

- sociology

3.3.2 - Methodology and techniques

<u>When the methodology or technique concepts **should** be selected for indexing</u>

I) If a methodology concept is a primary focus of the research project, index it. This includes:

– New methodology

– Focus of the research is analytical methods

– Research includes improving on a known methodology

– New use of a known methodology

– Evaluation of known methodology, i.e., evaluation of a known methodology for accuracy or reliability

– Comparison of two or more methods

J) Consider indexing methodology if: 1) it provides a unique or characteristic mechanism for studying or evaluating the primary topic, or 2) helps to describe the nature of the results.

Examples:

If a document is about the isolation, purification or fractionation of a mixture of compounds, "high performance liquid chromatography" might describe more precisely what kind of "isolation" is involved.

An article about determining of the structure of a chemical compound by nuclear magnetic resonance spectroscopy needs both "chemical structure" and the technique as indexed concepts because a user trying to determine the structure of a new compound by nuclear magnetic resonance would want to compare the spectral data of the new compound to the one in the article.

An article showing determination of the ultrastructure of a biochemical compound by X-ray diffraction gives different data than one using scanning electron microscopy. Users might need to know the technique used to determine the ultrastructure.

An article about the effect of nutrient intake on some biochemical parameter or disease condition is best described by adding the method of data collection, e.g. epidemiological survey, dietary survey or clinical trial. While the article does not focus on the survey technique, including these concepts may help to describe what the article is about and effects the interpretation of the results.

When the methodology or technique concepts **should not** be selected for indexing

K) Do not select the methodology or technique routinely. There needs to be a specific reason that the methodology or technique is included in the description of a document. Before using a technique descriptor, consider whether or not a user would benefit from being able to retrieve that article using that descriptor. If every article using "polymerase chain reaction" is assigned the descriptor, there is no distinction between articles based on human judgment of "aboutness".

L) Some research articles give the details of the stepwise procedure of the overall analysis, such as, collection of leaves, drying of leaves, milling of leaves, acid digestion of leaves, etc. Do not index the individual stepwise procedures.

Subject analysis is, in many respects, the most demanding aspect of subject indexing. If a concept is missed or is misunderstood, the value of a document to the database is diminished; that information is lost. Thorough, accurate and consistent selection of concepts is the key for improving AGRICOLA users' access to relevant information.

Subject analysis – Summary

1) <u>All</u> of the concepts necessary to adequately represent the document's significant aspects should be selected for indexing (exhaustivity).

2) Selected concepts should be represented at the level of detail consistent with the treatment of the subject in the document (specificity).

3) Indexers will index significant negative results.

4) The concept(s) that represent the AGRICOLA selection criteria of the document will be selected.

5) Equal concepts will be treated equally.

6) Concepts that were not studied in the document at hand will not be included.

7) Methodology concepts will be selected in accordance with the following specific items from Section 3.3.2 - <u>Methodology and Techniques</u>:

  A) Methodology <u>should</u> be indexed if a methodology concept is a primary focus of the research project.

  B) Routine techniques or individual stepwise procedures <u>should not</u> be indexed.

8) Indexers will focus on the concepts that convey the subject(s) or objective(s) of selected documents, rather than the specific measurements of those concepts as described in item F, Section 3.3.1

SECTION 4 – "Translation" of concepts into indexing languages

The second step of subject indexing is "translation". This is the process of communicating the concepts that were selected for indexing using:

  1) descriptors from the *NALT*

  2) identifiers (uncontrolled vocabulary)

  3) AGRICOLA Subject Category Codes

As described earlier, the information in AGRICOLA bibliographic records is organized using the MARC format:

1) MARC tag 072 codes (*AGRICOLA Subject Category Codes*): used to classify each document into one or more broad areas of agriculture

2) MARC tag 650 descriptors (*NALT*): express the subject concepts in a document that have been identified for indexing

3) MARC tag 651 descriptors (*NALT*): describe geographical locations

4) MARC tag 653: as the need arises, indexers can use natural language to further describe concepts; these terms are called "identifiers" or "nondescriptors."

The following sections include rules and guidelines to assist indexers in effectively using these "language" tools.

4.1 - Assignment of *NALT* terms

4.1.1 - MARC tag 650 descriptors

4.1.1.a - Word blocks - descriptor relationships and general information

Accurate and consistent application of controlled vocabulary terms strengthens precision (relevance of the information retrieved). To preserve the value of subject indexing to AGRICOLA, it is critical for indexers to examine each word block carefully, and assign descriptors by the rules and guidelines described in this section.

Each descriptor (word or phase) in the *NALT* has a word block, which is graphically displayed as follows:

1) hierarchical relationships: Broader terms (BT)

                                        Narrower terms (NT)

The hierarchies in the thesaurus may contain up to seven levels of increasingly broader (BT1, BT2, BT3, etc.) or increasingly narrower (NT1, NT2, NT3, etc.) descriptors.

2) equivalence relationships: Use/Used for (USE/UF, displayed in italics)

Use AND/Used for AND (USA, displayed in italics)

3) associative relationships: Related terms (RT)

4) Scope notes (SN)

The *NALT* web site (http://agclass.nal.usda.gov/) offers a very thorough discussion of the relationships between terms and notes (see http://agclass.nal.usda.gov/structure.shtml). For proper and effective use of the *NALT*, it is critical that indexers read and completely understand this material.

All terms must be used with all capitalization and punctuation appearing exactly as it does in the *NALT*.

Word block examples:

agroforestry

Used for:

agri-silvicultural systems

agrisilvicultural systems

agro-silvicultural systems

agrosilvicultural systems

Broader Terms:

farming systems

forestry

Narrower Terms:

alley cropping

silvopastoral systems

Related Terms:

plant strata

sustainable agriculture

virus curing

Definition:

Causing the loss of a dormant virus which has inserted itself into a bacterial genome (a lysogenized virus)

Used for:

curing, viruses

Broader Terms:

genetic techniques and protocols

amino acid substitution

Scope Note:

Use for substitution of one amino acid for another; for amino acid substitution as experimentally induced by man USE amino acid substitution AND protein engineering.

Definition:

The replacement of one or more amino acids in a protein with another. If a functionally equivalent amino acid is substituted, the protein may retain wild-type activity. Substitution may also diminish or eliminate protein function.

Broader Terms:

mutation

Related Terms:

missense mutation

protein engineering

protein structure

structure-activity relationships

egg clutches

Scope Note:

Use for birds; For insects and other arthropods USE egg masses

Broader Terms:

oviposition

Related Terms:

clutch size

fruits (plant anatomy)

Definition:

The matured ovary of a flower

Used for:

fruit (plant anatomy)

Broader Terms:

plant anatomy

Narrower Terms:

bolls

pericarp

pods

U.S. Soil Taxonomy

Used for:

Soil Taxonomy (United States)

United States Soil Taxonomy

Broader Terms:

>   soil classification

Related Terms:

>   genetic soil types

A) Indexers must use only *NALT* descriptors in field 650. Indexers may not create their own terms to enter in this field.

B) Nondescriptors, indicated by the USE, UF, and USA designations cannot not be used by indexers as descriptors or identifiers (Section 4.2.2, item F). *NALT* descriptors cannot be used as identifiers (Section 4.2.2, item E). Furthermore, "USE" instructions, definitions and scope notes are not optional and must be followed. Nondescriptors are displayed in italics in the *NALT*.

Examples:

>   ### *DNA fingerprints*

>   **Use:**

>>   DNA profiling

>   The phrase "DNA fingerprints" may not be used as a descriptor or an identifier. "DNA profiling" must be used.

>   ### *postemergence herbicides*

>   **Use AND type:**

herbicides

postemergent weed control

The phrase "postemergence herbicides" may not be used as a descriptor <u>or</u> an identifier. Indexers who want to convey this concept must assign the descriptors "herbicides" <u>and</u> "postemergent weed control".

C) Do not use a descriptor from the *NALT* unless the broader terms (BT), definitions, and scope notes are consistent with the meaning needed for the article you are indexing. If an indexer does not follow this instruction, he/she runs the risk of seriously misleading AGRICOLA users and negating the value of using a controlled vocabulary.

Example:

compliance

Broader Terms:

human behavior

Related Terms:

dropouts

laws and regulations

The term "compliance" may not be used to describe animal behavior, because its broader term (BT) is "human behavior".

This instruction applies even if some of the related terms (RT) <u>do</u> have meanings consistent with the meaning conveyed by the document you are indexing.

Similarly, an indexer should not feel prohibited from using a descriptor based solely upon its related terms. Related terms are provided in word blocks to suggest to thesaurus users concepts that might be of additional interest, but are not considered part of the BT/NT (hierarchical) term relationship.

D) In the case of polyhierarchies, where one descriptor has more than one BT, the descriptor can be used in the sense of any of its broader terms.

Examples:

oilseed protein

Broader Terms:

oilseed products

plant source protein

protein products

Narrower Terms:

coconut protein

cottonseed protein

rapeseed protein

sunflower seed protein

Related Terms:

oilseeds

This descriptor may be assigned in the context of an "oilseed product" OR "protein product" OR "plant source protein"

irrigation

Scope Note:

Application of water to soil for the purpose of plant production

Used for:

herbigation

watering

Broader Terms:

irrigation and drainage

soil management

Narrower Terms:

irrigation canals

irrigation management

irrigation systems

irrigation water

Related Terms:

chemigation

fertigation

irrigated conditions

irrigated farming

The term "irrigation" may be assigned as a more specific concept of "irrigation and drainage" or "soil management".

E) Examine the complete word block of a descriptor before assigning it to a document. Never assign a descriptor on the basis of a "USE" or "USA" reference without consulting their hierarchies:

1) You may discover that the descriptor to which you are being referred has the wrong hierarchical context for the document at hand.

2) You may discover that an NT of the descriptor to which you are being referred is more appropriate and/or specific.

4.1.1.b - Specificity

F) For each concept, indexers will assign the <u>most specific descriptor(s)</u> in the *NALT*. The level of specificity will reflect the level of specificity used by the author(s).

Example:

> In a discussion of "...clear-cutting in the boreal forest of Alaska..." the most specific descriptor ("clearcutting") must be used rather than the broader term "silvicultural practices".

G) Indexers may also use the broader term (BT) of any assigned descriptor, if, in their opinion, a document should also be retrievable as part of a broader concept. This may be true, for example, if the author uses one or more objects to represent a larger group.

Example:

> In a study of analgesics and appetite in animals, the authors use acetaminophen and aspirin in the study. The descriptor "analgesics" may also be applied along with "acetaminophen" and "aspirin" because they were used to represent the larger concept of "analgesics".

The broader term of any descriptor may also be used if the document addresses a concept at both the broader and narrower levels.

Example:

> In a special issue on Soil Moisture Theories, the author might address a variety of specific issues ("soil structure", "soil temperature", "soil texture") <u>and</u> their overall contribution to "soil physical properties".

H) If more than three narrower terms of the same hierarchy are at the same level, the indexer may choose to use their BT to represent this concept.

Example:

> If an article discusses "barley", "millets", "buckwheat", "oats", "rice", "rye", and "wheat", an indexer may choose to use the descriptor "small grains" instead of these narrower terms (NT).

4.1.1.c - General considerations

I) Many concepts consist of more than one component. Index each concept completely, i.e., for all its components. It will often be necessary to assign more than one descriptor to index a single concept.

Examples:

> The concept of <u>spleen weight</u> consists of two components, represented by the *NALT* terms "spleen" and "tissue weight."
>
> The concept of <u>computer monitors</u> would be indexed using the *NALT* terms "computer hardware" and "monitors (image display)".

(Note: This is also addressed throughout Section 3.2, beginning with item 2 under <u>OBJECT or CENTRAL SUBJECT</u>.)

J) However, do not use two descriptors to represent a single concept when a precoordinated descriptor for the concept exists.

Example:

> For the concept of "soil temperature" do not use "soil" and "temperature" individually.

K) The order of descriptors is not critical to retrieval of information from AGRICOLA. However:

> 1) AGRICOLA users may perceive the first descriptors to be the most important.
>
> 2) Proximity searching (searching for the occurrence of one word or term near another) is available in many search formats.

Therefore, indexers <u>may</u> wish to arrange their descriptors in such a way that:

> 1) the main concepts of a document appear as the first descriptors
>
> 2) related concepts are in proximity to one another

L) There may be situations where descriptors in the *NALT* are not adequate to convey the desired level of specificity. There may also be instances when hierarchies, definitions, or scope notes prohibit the desired use of a descriptor. Occasionally a concept may not be represented in the *NALT* in any way. The goal of indexers should be to bring AGRICOLA users "as close as possible" to the concept using the *NALT* descriptors in an accurate and meaningful manner. Next, <u>after carefully checking in the *NALT*</u>, assign identifiers as described in Section 4.2.

4.1.1.d - Use of Taxonomic Names (see Supplement C for a summary of this section)

<u>General Considerations</u>

M) For scientific names, use the highest level of specificity available in the *NALT* in field 650. This might be order, family, genus, or species. If necessary, reflect a greater level of specificity as an identifier (in field 653). Standards and guidance are provided below by general organism type.

N) In cases where the use of a scientific name is mandated by this *Guide* but is not available in the document at hand, it is the responsibility of the indexer to seek it out. Often, a quick way to do this is to

search the AGRICOLA database or the Internet. However, common names are highly variable and often very localized. Therefore, if an indexer cannot trace the scientific name at any level of the taxonomic hierarchy 1) with a reasonable degree of certainty and 2) within a reasonable period of time (a few minutes), it may be necessary to assign a context or role descriptor to convey the intended concept. The scientific name provided by the author(s) may be added as an identifier.

O) In the *NALT* the "common" or "group name" of an organism may be a nondescriptor for the taxonomic name. The reverse may also be true; the taxonomic name may be a nondescriptor for the "common" or "group name." Adhere to the rules for nondescriptors in these instances (Section 4.1.1.a, item B and Section 4.2.2, item F).

P) If a role or context descriptor is the BT of one or more organisms or products, it may be assigned (see Section 4.1.1.b, item H).

Q)   Plants

1) Index the taxonomic name for the plant specified in the document.

2) Add the commodity or product name where available. Check the RT relationship in the *NALT* for the commodity name.

Example:

> Weed management systems for Georgia cotton (Gossypium hirsutum L.) is indexed as "Gossypium hirsutum" and "cotton."

3) Based upon the circumstances of the research or discussion at hand, consider the role of the plant, e.g., "weeds," "medicinal plants," "poisonous plants," "conservation plants," etc. Perform a string search

in the thesaurus on "plants" or consult the "plants" hierarchy to find many of these role or context descriptors.

Example:

An article with the title "Growing dandelions for food and profit" would be indexed with the terms "Taraxacum officiale" and "food crops"; but "Control of dandelions in turfgrass" would be assigned the terms "Taraxacum officiale" and "weeds".

R)   Plant products

1) Use the commodity/product name of the plant if it is available as a descriptor.

2) If the commodity name is not available, use the taxonomic name and add the commodity/product name to the identifier field.

Example:

Texture of cooked carrots after frozen storage is indexed as "carrots." Do not assign "Daucus carota."

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of the plant product(s), e.g. "chilled foods," "convenience foods," "textile fibers," or "stored products." Consult the "agricultural products" hierarchies such as "foods," "nonfood products," or "plant products." Perform a string search in the *NALT* on "products" to locate many of these role or context descriptors.

Example:

Fiber strength of cotton polyester blends is indexed as "cotton" and "textile fibers."

S)   Fungi

1) Index the taxonomic name for the fungus specified in the document, including the subspecific designation where it is available as a descriptor.

2) Based upon the circumstances of the research or discussion at hand, consider the role or context of the fungus, e.g., "mycorrhizal fungi," "plant pathogenic fungi," "entomopathogenic fungi," "thermophilic fungi," etc. Perform a string search in the *NALT* on "fungi" or consult the "fungi" hierarchy to find many of these role or context descriptors.

Example:

>An article on "Schizophyllum commune and sapstain development on jack pine logs" would be indexed using "sapstain fungi" and "Schizophyllum commune."

T)   Bacteria

1) Index the taxonomic name for the bacterium specified in the document, including the subspecies designation where available as a descriptor.

2) Based upon the circumstances of the research or discussion in hand, consider the role or context of bacteria, e.g., "nitrogen fixing bacteria," "rumen bacteria," "lactic acid bacteria," "cellulolytic microorganisms," etc. Perform a string search in the *NALT* on "bacteria" and/or "microorganisms" or consult the "bacteria" hierarchy to find many of these role or context descriptors.

Example:

>An article on "cloning a beta-glucosidase from the cellulolytic actinomycete Thermobifida fusca," which specifically discusses cellulose degradation, would be assigned the descriptors: "Thermobifida fusca," and "cellulolytic microorganisms." If, for example the article only discussed the molecular genetics of cloning, it would be fine not to use "cellulolytic microorganisms" as a role term.

U)   Protozoa

1) Index the taxonomic name for the protozoan specified in the document.

2) Based upon the circumstances of the research or discussion at hand, consider the role or context of protozoa, e.g., "entomopathogenic protozoa," "rumen protozoa," "soil fauna," etc. Perform a string search in the *NALT* on "protozoa" and/or "microorganisms" or consult the "microorganisms" hierarchy to locate many of these descriptors.

Example:

An investigation of "a spinated type of Diplodinium in the rumen of two Japanese shorthorn calves" would be indexed with the terms: "Diplodinium" and "rumen protozoa."

V)   Viruses

1) Index the taxonomic name for the virus specified in the document.

2) Based upon the circumstances of the research or discussion at hand, consider the role or context of viruses, e.g., "plant viruses," "arboviruses," "invertebrate viruses," etc. Perform a string search in the *NALT* on "viruses" and/or "microorganisms" or consult the "viruses" hierarchy to find many of these role or context descriptors.

Example:

An article examining a "reverse transcriptase PCR assay detection of three fish viruses" would be indexed with the names of the three viruses ("infectious hematopoietic necrosis virus," "infectious pancreatic necrosis virus" and "viral hemorrhagic septicemia virus") and the descriptor "vertebrate viruses."

W)   Arthropoda (all members)

1) Index the taxonomic name for the arthropod specified in the document.

2) Add the "common group name" when it is appropriate, e.g. "mites", "crabs", "lobsters", "insects", etc.

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of arthropods, e.g., "soil arthropods," "hematophagous arthropods," "predatory mites," "insect pests," etc. Perform a string search in the *NALT* on "arthropods," "mites," and/or "ticks" or consult the "arthropods" hierarchy to find many of these role or context descriptors.

Example:

> An article on the "efficacy of a diatomaceous earth against Acarus siro in small bins of wheat under conditions of low temperature and high humidity" would be indexed with "Acarus siro" and "storage mites".

X) <u>Invertebrates other than arthropods</u>

1) Index the taxonomic name for the invertebrate specified in the document.

2) Add the "common group name" when it is appropriate, e.g. "snails," "oysters," etc.

3) Based upon the circumstances of the research or discussion at hand, consider the role of invertebrates, e.g., "soil invertebrates," "earthworms," "plant parasitic nematodes," etc. Perform a string search in the *NALT* on "invertebrates," "nematodes," and/or "worms" or consult the "invertebrates" hierarchy to find many of these role terms.

Example:

> "The prevalence of feline heartworm disease in Italy" would be indexed with the following descriptors: "Dirofilaria immitis" (the taxonomic name of the helminth) and "heartworms."

Y) <u>Fish</u>

1) Index the taxonomic name for the fish specified in the document.

2) Add the common group name of the fish when it is appropriate, e.g., "trout," "catfish," etc.

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of fish, e.g., "marine fishes," "aquarium fish," "farmed fish," etc. Perform a string search in the *NALT* on "fish," or consult the "fish" hierarchy to find many of these role or context descriptors.

Example:

> "Seasonal population changes of parasitic copepods on species of farmed carp in China" would be assigned the descriptors "Ctenopharyngodon idella," "Hypophthalmichthys nobilis," "carp," and "farmed fish." (It would also be assigned "Copepoda" and "ectoparasites" – see Arthropoda above.)

Z)   Amphibians

1) Index the taxonomic name for the amphibian specified in the document.

2) Add the group name "frogs," "salamanders," or "toads" when it is appropriate.

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of the amphibians.

Example:

> The article "Non-invasive treatments inducing spermiation in Bufo americanus toads" would be indexed with the descriptors "Bufo americanus", "toads," and possibly "beneficial organisms."

AA)   Reptiles

1) Index the taxonomic name for the reptile specified in the document.

2) Add the group name for the reptile where appropriate, e.g., "snakes," "alligators," etc.

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of the reptiles.

Example:

An article appearing in a veterinary journal investigating "Intraocular pressure of mature loggerhead sea turtles (Caretta caretta) held in various positions" would be indexed with "Caretta caretta" and "turtles". The indexer would probably not convey the "role" of the turtle because it's not a key concept in this instance, but "animal physiology" or one of it's narrower terms is a possibility.

However, an article with the title "South East Asia's vanishing turtles" would be indexed with the taxonomic names as needed, "turtles," and "endangered species".

BB)   Mammals and Birds (**not** as livestock, experimental animals, or pets)

1) Index the taxonomic name for the mammal specified in the document.

2) Add the "common group name" (often an RT relationship in the *NALT*)

3) Based upon the circumstances of the research or discussion at hand, consider the role or context of the animals, e.g.,"wild animals," "game animals," "zoo animals," "game birds," "old animals", "wild birds," etc. Perform a string search in the *NALT* on "birds," "animals," "mammals," or consult the "animals" hierarchy to find many of these role or context descriptors.

Example:

"Health protocol for rescued hawks (Accipitridae)" would be indexed using the descriptors "Accipitridae", "hawks," and "wild birds."

CC)   Mammals and Birds (as livestock, experimental animals, or pets)

1) Index the "common group name" of the mammal.

2) Based upon the circumstances of the research or discussion at hand, consider the role or context of the animals, e.g. "racing animals," "pets," "old animals," etc. Perform a string search in the *NALT* on "animals" or "mammals" or consult the "animals" hierarchy to find many of these role or context descriptors.

3) Add the breed name of the animal when it is significant to the research.

Examples:

In a discussion of "facility decontamination techniques used to improve the health and welfare of transgenic mice," there is data presented on the condition of "old mice." The descriptors for this article should include: "mice," "old animals," and "transgenic animals" and "laboratory animals" (see item DD below).

For "A cross-sectional study of paratuberculosis in 1155 Danish Holstein-Friesian, Ayrshire and Jersey dairy cows," the breed names "Holstein," "Ayrshire (cattle breed)," and "Jersey" would be assigned, along with "dairy cows."

NOTE: Useful and appropriate role or context terms for any of the organisms listed in items Q through CC may be located by a string search on "species" or "organism."

DD)   Laboratory Animals

1) Do not supply any taxonomic indexing for laboratory animals; use only the "common group name" as given in the document.

2) If, however, there is no descriptor for the common or group name, index the taxonomic designation for the subject as instructed under item BB [Mammals and Birds (**not** as livestock, experimental animals, or pets)] above.

3) For experimental animals, either as the subject of research or as an experimental tool, add the descriptor "animal models."

Example:

> For a human nutrition research project studying the effect of beta-carotene intake on antioxidant status, where rats are used as experimental animals, assign the terms "rats" and "animal models," not "Rattus rattus."

4.1.1.e - Use of descriptors for diseases and experimental animal models

A primary consideration in this section are indexing practices that will recognize the relationship(s) between the host organism, the disease organism and the resulting disease. Unless noted, in the circumstances below, the specificity of descriptors will be assigned as described under Section 4.1.1.b – Specificity.

EE)   Animal models

1) When animals are used as models for human disease, add the term "human diseases."

2) In documents of this sort, index the animal as with laboratory animals above and add also the descriptor "animal models."

Example:

"Spontaneous porphyria of the Long-Evans Cinnamon rat: an animal model of Wilson's disease."
Index with "porphyria," "animal models," "human diseases," and "rats."

FF)   Diseases/infections in animals

1) Index with the most specific descriptor available for the disease/infection.

2) Add the taxonomic designation for the infecting organism, if known. Consult the *NALT* related term for links between disease names and causal agents, e.g., "Aujeszky disease," RT: "Suid herpesvirus 1."

3) Consider the role or context of the infecting organism, such as "entomopathogenic nematodes" or "vertebrate viruses."

Examples:

Index Taenia ovis taeniasis as "taeniasis" and "Taenia ovis."

Index a Hymenolepis nana parasitosis as "cestode infections" and "Hymenolepis nana."

4) Index diseases/infections in animals with the appropriate descriptor for the disease and the pre-coordinated "animal diseases" heading.

Examples:

Index "cystic ovaries in cattle" with the terms "ovarian cysts" and "cattle diseases."

Index "congestive heart failure in Great Danes" with the terms "congestive heart failure" and "dog diseases."

Index "Brucella abortus infection in cattle" with the terms "bovine brucellosis" and "Brucella abortus, but not with the term "cattle diseases," since "bovine brucellosis" is a narrower term under "cattle diseases" in the hierarchy.

GG)   <u>Plant diseases</u>

1) Index diseases in plants with the appropriate descriptor for the disease from the "plant diseases and disorders" hierarchy. Note that several terms may be applied for one article, e.g., "tree diseases" and "root diseases."

2) Add the taxonomic designation for the infecting organism, if known.

3) Based upon the circumstances of the research or discussion at hand, consider the role of the infecting organism, such as "plant pathogenic nematodes" or "plant viruses."

Examples:

Index "Phytophthora cinnamomi root rot as a cause of oak mortality" with the terms "Phytophthora cinnamomi," "tree diseases," and "root rot."

Index "A Ustilago segetum infection of oats" with the terms "smut diseases" and "Ustilago segetum."

Index "A Meloidogyne incognita infection of alfalfa" with the terms "root-knot nematodes" and "Meloidogyne incognita."

4.1.1.f - Publications by type

Consult the facet "(publications by type)." Use these *NALT* terms when the document **is** one of these types. The most commonly used publication type is "literature reviews." This term should be assigned regardless of "publication type" coding that might be assigned elsewhere in the bibliographic record.

Assignment of descriptors - Summary

1) Indexers will only use *NALT* descriptors in field 650.

2) Indexers will assign *NALT* descriptors in a manner that is consistent with their broader terms (BT).

3) "Use" instructions, definitions, and scope notes in the *NALT* will be strictly observed.

4) Concepts will be expressed using *NALT* descriptors at the most specific level of specificity used by the author of the document at hand. If, however, an indexer is using more than three narrower terms (NT) in the same hierarchy at the same level, he/she may choose to use the broader term (BT) to represent this concept.

5) Two descriptors may not be used to represent a single concept when a precoordinated descriptor for the concept exists in the *NALT*.

6) In circumstances where a) concepts are not represented at the required level of specificity, b) concepts are not covered or covered adequately in the *NALT*, or c) the *NALT* word block erroneously restricts the use of one or more terms, indexers will use guidelines A through C in Section 4.2 - Assignment of MARC tag 653 identifiers to assign terms.

7) Indexers will assign taxonomic and common names of organisms and products according to Use of Taxonomic Names and in a manner that is consistent with all other standards in Section 4.1.

8) Indexers will assign descriptors for diseases (human, animal and plant) and experimental animal models according to the sections on Use of descriptors for diseases and experimental animal models and in a manner that is consistent with all other standards in Section 4.1.

9) Indexers should examine word blocks carefully for each descriptor considered for assignment.

10) Indexers may use the broader term (BT) of any assigned descriptors if they feel the document at hand should be retrieved on this broader concept.

11) A descriptor and its broader term (BT) may be used if the concepts are addressed in the article at both levels.

12) It may be necessary to use two descriptors to represent a single concept.

13) The "Indexing Selection Scope and Coverage Guidelines", Supplement B, can help indexers assign terms to convey the relevance of a document to AGRICOLA.

14) The order of descriptors is not critical to information retrieval; however, indexers may choose to consider proximity and order when arranging descriptors.

15) Consult the facet "(publications by type)." Use these *NALT* terms when the document is one of these types.

16) Indexers may use as many descriptors as necessary to convey all concepts identified for indexing. However, when the number of descriptors for one document exceeds 20, indexers are encouraged to re-evaluate their choices. Subsuming some narrower terms (NT) under their broader terms (BT) may provide AGRICOLA users with a more meaningful record.

17) Problems with the *NALT*, including missing terminology, restrictive hierarchies, definitions and scope notes, or other problems, must be brought to the attention of the NAL Thesaurus staff.

4.1.2 - MARC tag 651 geographical descriptors

A) Use the instructions and guidance provided in sections 4.1.1.a through 4.1.1.c when assigning geographical descriptors.

B) Generally, if the work is conducted in a laboratory or a greenhouse, use of a geographical descriptor is not warranted, unless it is apparent where the analyzed material originated.

C) The geographical origin or provenance of a strain or line <u>may</u> be of significance in a way that warrants use of a geographical descriptor, even if the work is done in vitro. Note, however, that this does not apply to geographical words appearing in the nomenclature of strains of bacteria and other organisms.

D) For articles in which subjects (animals, plants, humans) originating from one country/area are studied in a country to which they have migrated or been imported, do not apply a geographical descriptor for the country of origin. Apply the geographical descriptor for the country in which the subjects were studied and indicate their geographical or ethnic origin in some other way.

Examples:

> A study about food habits of Salvadorans living in Manhattan would be assigned "New York" and "Hispanics" as descriptors. However, this practice is warranted only in the case of living subjects studied in a natural environment.

> An article about insect specimens collected in Ghana and studied in a laboratory at the Agricultural Research Service in Maryland would be assigned "Ghana" (not "Maryland") as its geographical descriptor.

E) The affiliation of the author(s) may sometimes be a valuable clue in deciding upon geographical descriptors, but it should never be used as the sole criterion for doing so.

F) There is no limit to the number of geographical descriptors which can be used for a given article. However, in most cases in which more than five or six are applicable, it will probably be more useful to apply terms for broader geographical units that subsume the others.

Assignment of geographical descriptors - Summary

1) All of the instructions provided in Sections 4.1.1.a through 4.1.1.c are to be used by indexers when applying geographical descriptors.

2) Geographical terms may not be assigned solely based upon the appearance of geographical terms in the nomenclature of strains of bacteria and other organisms**.**

3) For articles in which subjects (animals, plants, humans) originating from one country/area are studied in a country to which they have migrated or been imported, apply the geographical descriptor for the country in which the subjects were studied and indicate their geographical or ethnic origin in some other way. This applies only in the case of living subjects studied in a natural environment.

4) Geographical terms may not be assigned based solely upon the affiliation of the author(s).

5) Generally, if the work is conducted in a laboratory or a greenhouse, use of a geographical descriptor is not warranted.

6) The geographical origin of a strain or line may sometimes be of significance in a way that warrants use of a geographical descriptor even if the work is done in vitro.

7) There is no limit to the number of geographical descriptors that may be used for a given article. However, in most cases where more than five or six are applicable, it will probably be more useful to apply terms for broader geographical units that subsume the others

8) Problems with the *NALT*, including missing terminology, restrictive hierarchies, definitions and scope notes, or other problems, must be brought to the attention of the NAL Thesaurus staff.

4.2 - MARC tag 653 identifiers

The *NALT* should be used to index the article as completely as possible. If additional subject description is needed after indexing the article with the *NALT* descriptors, identifiers may be used. Identifiers are uncontrolled subject terms not found in the *NALT* either as descriptors or nondescriptors.

Identifiers are used to:

a) achieve a higher level of specificity

b) enhance retrieval further

4.2.1 - When to use identifiers

A) Concepts not represented at the desired level of specificity in the *NALT*.

If a concept that should be indexed cannot be found after searching the *NALT*, use the best available broader descriptor(s) to convey the concept(s).

Example:

>   If the article is about 'sacroglycans,' index the article with the broader term "polysaccharides" in the 650 field and use the term 'sacroglycans' in the 653 field.

When creating an identifier for a product or organism, consider the use of a role or context descriptor as a means of conveying the desired concept in field 650 (see Section 4.1.1.d).

Use identifiers for significant local geographical concepts. For instance, add 'Milwaukee, Wisconsin' to field 653 if the city itself is emphasized. ("Wisconsin" will, of course, be entered in field 651.)

Use identifiers for taxonomic names. For instance, the species 'Senecio arvensis' is not in the *NALT* and so must be entered in field 653. ("Senecio," which is in the *NALT*, will represent this species in field 650.)

Use identifiers for commodity names that are not in the *NALT*. Use the most specific taxonomic descriptor available in the *NALT* in field 650.

B) Subjects and concepts not covered well by the *NALT*.

To enhance retrieval, use identifiers to index new concepts and emerging technology not yet included in the *NALT*. Look in the *NALT* for a broader term (BT) to index the concept. If a broader term cannot be found, consider a related concept.

Example:

> The term "photosynthesis" might be used in the descriptor field to index the concept of electron energy levels in an article about the molecular events in a photosynthesis study with 'molecular energy levels' added to the identifier field.

Do not use a 650 descriptor that will misrepresent the subject content of the article. Use identifiers liberally to aid retrieval but not hurt consistency. It is important to add new concepts to the identifier field to enhance retrieval and to refer suggestions for new terms to the NAL Thesaurus staff.

C) *NALT* word block restriction

Use the identifier field to index concepts for which the descriptors in the *NALT* are in hierarchies that restrict their use according to section 4.1.1.a, item C. After carefully searching the NALT for the descriptor(s) that express the desired concept(s), assign one or more identifiers.

D) Special cases for the use of the identifier field

Use identifiers for clarification of some abbreviations, acronyms, initialisms, etc. in document titles. For instance, 'PTA' in a title is explained by adding 'Parent Teacher Association' in field 653.

Use identifiers for proper names not in the *NALT*. For instance, 'Disaster Assistance Act of 1988' or 'Alabama Home Economics Association' or 'Coalition for Peaceful Schools'.

When they are not descriptors, use the American spellings for British words as identifiers.

4.2.2 - Restrictions of Use of Identifiers

E) Do not use as identifiers terms which appear in the *NALT* as descriptors, even if the thesaurus descriptor represents a concept other than the concept associated with the term in the document at hand. For instance, the descriptor "decline" carries a BT "plant diseases and disorders." It is not permissible to

enter "decline" in field 653 for a document that discusses decline in some other context, such as human health.

F) Do not use as identifiers terms which appear in the *NALT* as nondescriptors. It is mandatory to follow the **"USE"** instructions in the *NALT*. For instance, if the *NALT* provides the instruction "sunlight USE solar radiation," it is not permissible to enter "sunlight" in field 653.

G) Do not use long phrases as identifiers; identifiers should be in the same general format as descriptors in the *NALT*. For instance, 'effect of rototilling on garden yields' is not an appropriate identifier.

H) Do not use identifiers as a short-cut to avoid searching the *NALT* for an appropriate descriptor.

Assignment of identifiers - Summary

1) Abbreviations, acronyms, initialisms, etc. will be spelled out in the 653 field when the appropriate term is not available as a *NALT* descriptor.

2) Indexers may not use descriptors which appear in the *NALT* as identifiers, even if the thesaurus descriptor represents a concept other than the concept associated with the term in the document at hand.

3) Indexers may not use as identifiers terms which appear in the *NALT* as nondescriptors.

4) Indexers should examine the *NALT* thoroughly before using the 653 field.

5) Problems with the *NALT*, including missing terminology, restrictive hierarchies, definitions and scope notes, or other problems must be brought to the attention of the NAL Thesaurus staff.

4.3 - AGRICOLA Subject Category Codes

The Subject Category Codes (or codes), found at http://agricola.nal.usda.gov/help/categorycodes.html, when assigned to a document, should represent that document as completely as possible and without any necessity to refer to the *NALT* descriptors assigned, and vice-versa. They are used separately, not in combination with one another, to represent the document, although, there may well be what appears to be redundancy in assignment of codes and descriptors.

Descriptors representing the same concepts represented by assigned codes should always be assigned, though most often the descriptors will be more specific than the codes to which they relate. To put this

another way, a user should be able to comprehend why each code was assigned by examining the assigned descriptors.

The reverse, however, is not necessarily true; descriptors may be used to bring out aspects of an article that are not represented by the assigned codes, because these aspects are not treated in sufficient depth or with sufficient emphasis to merit code assignment.

In expressing concepts in the indexing languages, aim for as much CONSISTENCY as possible in applying codes and descriptors to represent concepts. Use the same code(s) and the same descriptor(s) to represent the same concept each time you index that concept. Indexing will be valuable to AGRICOLA users only insofar as it is CONSISTENT.

The Subject Category Codes divide the field of agriculture into 21 broad subject areas. Each of these areas is further subdivided, but even the subdivisions are very broad. Each of the subdivisions carries an alphanumeric notation consisting of one letter (which remains the same within one of the 21 broad subject areas) and three numbers. There are some hierarchical relationships within the scheme, as in the following example:

L830 Animal Diseases - General

L831 Animal Diseases - Fungal

L832 Animal Diseases - Bacterial

L833 Animal Diseases - Viral

Each code also has a 'title,' known as a 'section heading' (e.g., notation: J800 = section heading: Soil Conservation).

A) Descriptors representing the same concepts represented by assigned codes should always be assigned.

B) An unlimited number of codes may be assigned per document.

C) Use relator codes (e.g., K001, L001, L002, L003, M001) to narrow the focus of large categories.

• Never place relator codes in first position.

• Never use relator codes without assigning at least one other code

D) Assign codes on the basis of the subject of the work actually conducted in the study at hand; do not assign them on the basis of what the subject can be used for (implication).

E) Follow the instructions provided in the codes' scope notes, which sometimes specify the order in which certain codes should be applied.

F) Never assign codes on the basis of the alphabetical index; always consult the scope note itself before assigning a code.

G) Use the most specific code(s) applicable.

Assignment of AGRICOLA Subject Category Codes - Summary

1) Indexers will assign one or more appropriate codes for each indexed document.

2) Descriptors representing the same concepts represented by assigned codes should always be assigned.

3) Use relator codes (K001, L001, L002, L003, M001) to narrow the focus of large categories.

    • Never place relator codes in first position.

    • Never use relator codes without assigning at least one other code

4) Assign Category Codes on the basis of the subject of the work actually conducted in the study at hand; do not assign them on the basis of what the subject can be used for.

5) Follow the instructions in the codes' scope notes, which sometimes specify the order in which certain codes should be applied.

6) Use the most specific code(s) applicable.

7) The code(s) assigned to a document should represent that document as completely as possible and without any necessity to refer to the *NALT* descriptors assigned, and vice-versa.

8) Assign the code(s) in descending order of importance.

9) Never assign codes on the basis of the alphabetical index; always consult the scope note itself before assigning a code.

SECTION 5 – Communicating with NAL Thesaurus Staff

Individuals with comments concerning the use of the *NALT* are welcome to contact the NAL Thesaurus Staff:

1) For general inquiries: thes@nal.usda.gov

2) To suggest changes to the *NALT*: http://agclass.nal.usda.gov/contact.shtml

3) Mailing address: National Agricultural Library, c/o Lori Finch, Room 012, 10301 Baltimore Avenue, Beltsville, MD 20705