

Assignment 3b
CS 410/510: Information Retrieval on the Internet
Due Thursday February 1, 2007

Goal: The goals of this assignment are to (a) get some practical experience with manual indexing, applying some of the principles and experiencing some of the challenges first hand, and (b) exploring a large thesaurus that is in regular use around the world. Note: there is not one “right answer” or ideal indexing for a particular document. Just give it a try, and have fun.

Details: Below are URLs for three web pages from the Oregon State University Extension Service. They are all about gardening and don’t require any specialized knowledge to understand. You are to manually index the first two using indexing terms from the Agrovoc Thesaurus, which is available from the Food and Agriculture Organization (FAO) of the United Nations (URL below). You can find terms in Agrovoc either by entering a search string or by browsing an alphabetical listing (searching will likely be much faster as there are nearly 40,000 terms in Agrovoc). Searches will often give you a short list of possible matches. Clicking on one of the terms will give you more information about the term, including its term code, its name in several languages, and possibly synonyms and related terms. For each indexing term you choose, record the term code (number) and the English preferred term. (If, on the right, you see USE: followed by a term name, you’ve found a lead-in term, or synonym, and the term name following the USE: is the preferred term). For the third page, assign “free terms,” that is, assign whatever keywords make sense to you, instead of using Agrovoc. In other words, you are not constrained by any pre-existing list of keywords.

For each page, you should consider the principles (and trade-offs) of exhaustivity and specificity that we discussed in class. You will probably want to assign about 3 – 10 terms for each page.

Agrovoc:

http://www.fao.org/aims/ag_alpha.htm?myLetter=all&myLangTerms=EN&qrylimit=100&myOrderBy=agrovocterm.termcode

Here are the articles to be indexed. All the articles come from *Extension Service Garden Hints*

(a) To index with Agrovoc:

1. Title: *Does your garden soil have enough organic matter?*
http://extension.oregonstate.edu/news/story.php?S_No=965&storyType=garden
2. Title: *Native shrubs can be leafy, evergreen and drought tolerant*
http://extension.oregonstate.edu/news/story.php?S_No=963&storyType=garden

(b) To index with “free” terms

3. Title: *Give houseplants TLC in winter*
http://extension.oregonstate.edu/news/story.php?S_No=958&storyType=garden

Deliverable:

For each of the three documents you have indexed, include the title of the document and then your list of keywords. Turn in your answers as hard copy at the beginning of class on February 1, 2007. Make sure your answers are legible. Word processed is best.