Background

A significant hurdle in biodiversity informatics is getting access to the wealth of information that has been published, but is not "linked in" to the internet age. While those of us at major universities do have access to a great deal of primary literature through our libraries, obtaining old articles is still a challenge. Furthermore, the fact that the content of these articles is not linked into major internet resources introduces a hurdle because researchers may not know that the article exists.

Biostor (http://biostor.org) is a new website written by Dr. Rod Page. The goal of the site is to resolve a citation for an article from the systematics literature to the URL of a page that displays a scanned version of the article. Most of the scanned pdfs of the articles in Biostor come from http://www.biodiversitylibrary.org; that site also produces plain text versions of the the article that are produced by "optical character reading" software (which recognizes letters by their shapes in a scanned document). While Biostor's primary goal is to allow you to find the article from a citation, it also tries to improve the potential for recognizing links between articles by "mining" the text of an article to identify taxonomic names, and using webservices to associate these names with useful links to further information. Biostor is *very* new, and so it would not be too surprising if some of the content of the pages was incomplete or incorrect (*eg.* broken or inappropriate links, or failure to correctly identify the taxonomic names in the text). In this exercise you'll get practice reading a species description, using internet resources to research a taxon, and we'll also provide some feedback to Dr. Page (I'll send him a summary of our experiences with Biostor).

Assignment 5: internet resources for biodiversity

In the "My Grades" section of blackboard you should see a column labelled "biostor" the content of the column should be a URL of a short article that is stored in BioStor.

1. Read the article carefully.
2. 6 pts List all of the taxonomic names in the article. Does your list agree with the one that BioStor produced?
3. 6 pts Is the taxonomy shown on Biostor correct? What reference did you use for finding a taxonomy to compare with Biostor?
4. 15 pts One goal of biodiversity informatics is improving the linking between websites that contain information for the same taxon so that scientists can quickly learn about more aspects of an organism. Choose a species that is mentioned in the paper that you were assigned (in some cases there may only be one species). Find at least **3 web resources** (other than Biostor) that contain scientific information about this taxon. For each web-resource:
   (a) provide the URL
   (b) describe what type of information the web site contains, and
   (c) discuss briefly whether you think that it would be possible for a programmer to write software that would have found the link that you found and recognized that it contained information about the same taxon – remember that computers are still not very good at processing human languages!

   You may want to consult the websites that I mentioned in the lecture on April 15th (links on blackboard)
5. 5 pts Did you find any errors data in Biostor's attempt to extract data from the text? (if "yes", please list the errors).
6. 7 pts The paper that you've been assigned involves the description of new species or subspecies. Can you tell what species concept, or species delimitation rules the author was using?
7. 5 pts Where is the type specimen for the species housed? (note: sometimes a museum collection is absorbed by a different museum, so the museum where the type was deposited may not exist anymore)
8. 7 pts What is the current valid name for the new species/subspecies described in the paper? Provide the reference for the answer (a web site is OK). If your paper has more than one species described, answer this question based on the first species mentioned.
9. 9 pts On the Wiki page `http://biostor.org/wiki/Main_Page` you will see that Dr. Page has added links to each of the references that we are using for this exercise. Find "your" assigned paper and click the `Reference:####` link. Near the top of the Wiki page you'll see the references and "Pages " followed by links to each page. Click on the page number that is assigned to you in the "OCR page" column of blackboard. When the page fully loads, you'll see both the scanned and OCR version of the page.
   Correct the OCR errors for this page. You can do this by:
   (a) creating an account on the Wiki and logging in to make the changes (if you do this send me your login name so that I can check your edits). Also make sure to put add a "Summary" comment in the Summary box in the bottom of the page to describe why you edited the page ("fixing OCR errors" is OK), Or
   (b) you can copy and paste the OCR text and email in the corrections (I'll later transfer your updates to BioStor).

   Obviously I would prefer if you did the edits yourself (it would be less work for me), but if you are having a hard time editing the Wiki or if you don't want to create an account on the Wiki you can use the second approach (I certainly won't take points off).