SILS Bioinformatics Journal Club
INLS 310-109
Fall 2002

Date: December 10, 2002
Presenter: John MacMullen
Topic: Literature-based discovery in biomedicine
Slides: http://ils.unc.edu/bioinfo/docs/20021210-MacMullen.ppt
Reading(s):

- Lindsay, R.K. & Gordon, M.D. (1999). Literature-based discovery by lexical statistics. Journal of the American Society for Information Science, 50(7), 574-587. http://www3.interscience.wiley.com/cgi-bin/abstract/60502468/START

- (Optional) Gordon, M.D. & Lindsay, R.K. (1996). Toward discovery support systems: A replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. Journal of the American Society for Information Science, 47(2), 116-128. http://www3.interscience.wiley.com/cgi-bin/fulltext?ID=57701&PLACEBO=IE.pdf

Summary:
Lindsay and Gordon attempted to reproduce work done by Swanson & Smalheiser (see September 10, 2002 journal club) who have used semi-automated methods to discover implicit connections in "complementary but disjoint" literatures.

Lindsay and Gordon attempted to formalize and automate Swanson & Smalheiser's method by using an approach limited to lexical statistics; no syntactic or semantic evaluation was employed. Their method used full MEDLINE records instead of the title only, which Swanson & Smalheiser had used.

Questions raised in the seminar included:

- Is the use of term frequency statistics useful, given that it has not been definitively shown that relevance or aboutness are proportional to term frequency?
- Could this experiment be reproduced by others, given that they employed human judgments and external domain knowledge (a medical student's judgments)?
- Given the manual intervention, is this method really all that different from Swanson & Smalheiser's?
- In general, is it reasonable to expect that one intermediate literature in between source ('C') and target ('A') literatures is enough to find implicit connections? If more are required in practice, how does that impact system complexity?