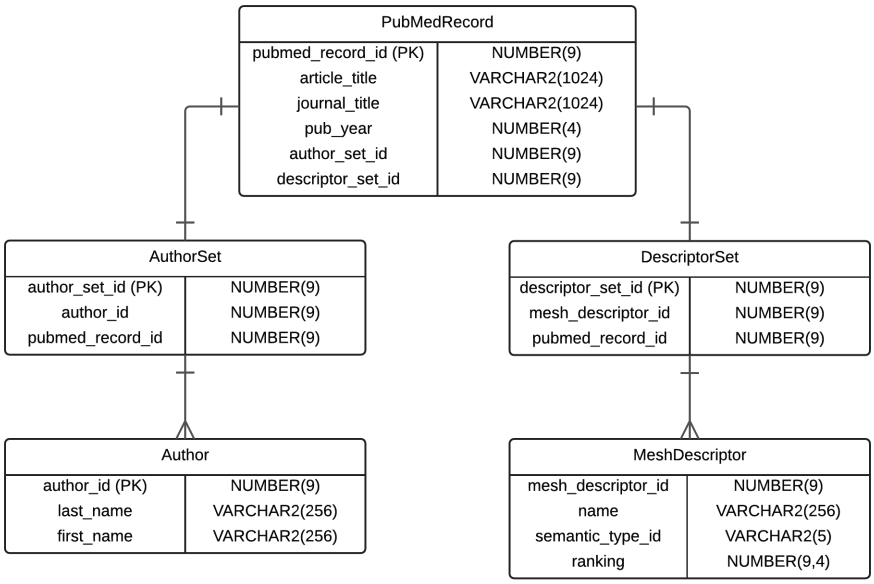
ABOUT THE SOFTWARE

The software was written in Java 1.7 and relies on JAXB for marshaling and unmarshaling results from Entrez. A pair of simple client classes (EntrezSearch and EntrezFetch) were created to do the eponymous work on Entrez; both of these classes extend from EntrezClient, which provides common functionality (mostly setting the WebEnv and providing a class that gives access to the query state.) EntrezFetch relies on XSLT to reformat the incoming XML into a form easily parseable by JAXB. These classes are used by the class ComorbidityRanking, whose main() method performs the work to load MeSH descriptors, perform the queries and emit raw data. The risk ratio for each descriptor was also calculated in this method (in Java, rather than in R or STATA, because, frankly, it's three divides and a pair of adds and who has time to load R libraries for that?)

MeSH descriptors were loaded from the MeSH 2015 XML vocabulary file using JAXB and an XSLT, which formatted it into a form easily digestible by the JAXB annotations. A utility class, MeSHPruner, pared the descriptors down to only the entries that were of semantic types T047 (Disease or Syndrome), T048 (Mental or Behavioral Dysfunction) or T191 (Neoplastic Process.) This process was performed once and its output loaded by the ComorbidityRank class before its query work started.

ABOUT THE SCHEMA

Below is a schema for the PubMed articles with a MeSH heading "obesity", selected in step number two and extracted in step number three. Note that journal information nor article information have not been factored out into their own entities, since the requirements did not specify more than what's shown in the table PubMedRecord. Selecting articles based on titles or publication year is straightforward; selecting by authors or MeSH descriptors requires a join via their respective 'set' tables.



Joseph Rose, Nov. 2015