

ISDB: Interaction Sentence Database

Michael Bauer, University of Arkansas Little Rock, Little Rock, Arkansas

Background and Objectives

There continues to be rapid growth in the amount of scientific literature available online. Consequently, it is becoming increasingly difficult to do comprehensive searches for specific biological information such as biological entity interactions. There is a need to shift aspects of the task of analyzing data from the researcher or curator to the computer. A greater knowledge of the sentence characteristics that may indicate an interaction between two biological entities is needed to aid in the creation of better-performing, information extraction tools.

Methods

Sentences in a previously-created database are from abstracts in MEDLINE and refer to at least two biomolecules. MySQL was used as the database management system and Perl code was written to parse and manipulate the data and populate interaction tables. The sentences in the database all contain at least two biomolecule terms and one interaction-indicating term.

Results

The Interaction Sentence Database (ISDB) which allows researchers to retrieve a set of sentences fitting specific characteristics. The web interface to the database allows a user to query for sentences based on an interaction-indicating term, a single biomolecule name, or two biomolecule names as well as for sentences with indirect connections through an intermediate biomolecule name.

Discussion and Conclusions

The retrieved sentences can be used to further characterize sentences that describe an interaction and can be expected to improve precision and recall of text mining and information retrieval tools.