PARALLEL PROCESSING  HORIZONTAL SCALING  DATA RESTRUCTURING

LIGERCAT

Ryan Schenk, Anthony Goddard, Holly Miller, Catherine N. Norton MBLWHOI Library, MBL

MEDLINE is a database of citations and abstracts in the diverse fields of medicine curated by the National Library of Medicine® (NLM). As part of the indexing process within MEDLINE, key words from a medical controlled vocabulary called MeSH descriptors are associated with each article. The web interface for accessing MEDLINE is PubMed (http://pubmed.org) developed by the National Center for Biotechnology Information (NCBI) at the NLM. In order to offer an intuitive way to explore the biomedical research in PubMed we created LigerCat (http://ligercat.ubio.org). This web application allows users to perform a keyword search for articles, then explore the resulting set via the a tag cloud of the key words (MeSH descriptors). The quantity of data required in this processing and the response speed required by a web based application lead us to the design of a distributed processing engine based around Ruby, Redis, RabbitMQ and MySQL.

This work has been funded by the Ellison Medical Foundation and National Library of Medicine (R01 LM009725)