

Harvest: A Digital Object Search and Discovery System for Distributed Collections with Different File Types and Structures

Webb, Frances, Paulson, Joy

The Harvest site, <http://harvest.mannlib.cornell.edu> is implemented using Fedora for data management, SOLR/Lucene for search, and Drupal for the user interface. Its goals are to provide an integrated search interface in which differences in format, structure and location are disguised in favor of treating objects that are conceptually alike as like, parallel objects. This is done by building Fedora content models that keep track of the complexity while providing services normalized to the objects' conceptual types; Lucene search documents that are fully normalized to hide implementation differences; and a Drupal front end that can treat all of the objects as generic objects until and unless specialized front-end services are built.