

docoll enhancement: search multiple trees

docoll uses Xapian and Xapian Omega to provide a browser-based UI to text search a single file system tree. It is similar to the desktop search solution, recoll, except it searches a headless server (and is not as fully developed). Docoll's indexing component is a bash script using Xapian's omindex to do the clever stuff. Docoll's UI component is a lightly customised Xapian Omega and an Apache configuration file.

We want to enhance docoll's indexing and UI components to search multiple trees. There are several directories, say /foo/{A,B,C, ... Z}, of which only some should be indexed and searchable via omega. There are several user groups, each interested in a different selection of {A,B,C, ... Z}, so the solution needs to provide a URL for each group which invokes omega (by CGI) to search the relevant mix of {A,B,C, ... Z}. For efficiency we need a single Xapian database for each of {A,B,C, ... Z}. Each group's omega instance must be configured to search the appropriate databases.

There was some discussion of how to do this on the Xapian mailing list, starting at <http://lists.xapian.org/pipermail/xapian-discuss/2013-May/008974.html>.

The docoll project home page is <https://savannah.nongnu.org/projects/docoll>. The most specifically relevant documentation is [docoll collating and indexing developer's guide. 2apr12.odt](#). The source code is in [docoll_server-0.7.6.tgz](#).

docoll was written by a member of the Blue Light team.

Alternatively it may be possible to run recoll on the server and use recoll's web browser UI (<https://github.com/koniu/recoll-webui/>).