richtags
cross repository browsing...

richtags allows you to search across multiple repositories from numerous institutions covering hundreds of disciplines for research that is of interest to you!

1. **browse** — find research you never knew existed
2. **search** — search multiple repositories with one click
3. **interact** — tag & discuss articles

category inference

A Category can be inferred for each eprint by searching for the associated Journal or Conference in the DMOZ & Wikipedia databases.

This provides another new way to browse the information space, producing links between articles VIA inferred categories that have not been possible before

Additional columns can be dragged into the slice to allow exploration on different fields.

Columns already in the slice can also be re-arranged to change the focus of browsing.

social interaction

Richtags provides a social, community based tagging system.

Users can register for an account and apply their own tags to any article in the system. These tags are yet another way to explore the information available and allow users to catalog any items they wish to easily find again.

Future Work

When a social tag reaches a certain threshold of popularity for a particular item, it will be promoted to a keyword.

year/decade extraction

The year and decade of publication is extracted from the Date field provided by the OAI feed.

Using range based identifiers such as year/decade allow for temporal browsing of the information which is not so easily done with just a date field.

author ambiguity

Future Work

Author ambiguity is caused from different text formatting conventions from one repository to another. We aim to reduce this by linking each Author to their email using text extraction techniques from the full text PDFs.

text extraction

Most of the meta-data for an Eprint is available through the repositories OAI feed.

Some data however, such as Journal/Conference titles must be extracted directly from the Eprint webpage using screen-scraping techniques.

Eprints 3 makes it possible to extract this information using the inbuilt XML Export Plugin

whois lookup

The repository title can be retrieved by interrogating the OAI feed.

The institution to which the repository belongs is not available in from the Eprints software. Richtags makes use of the whois lookup to associate repositories with Institutions.

tf-idf keywords

Keywords are extracted from the title and abstract of each Eprint using the TF-IDF ratio.

A list of known Stopwords and Word stemming are used to reduce the amount of irrelevant or redundant words

http://beta.richtags.net