dissertations, making them highly accessible, useful and shareable — at absolutely no cost.

Patents Online is the company behind www.freepatentsonline.com, the #1 worldwide Website in terms of audience in the patent space. With its worldwide patent database, freepatentsonline.com averages 10 million unique monthly users, and 1 million users have registered (registration is free and gives additional functionality).

Having started in the patent space, Patents Online's mission has expanded to include offering free access to all types of technical and academic documents. Theses and dissertations form an important body of documents in this space. OpenThesis.org will allow authors to upload their theses/dissertations, and the documents will then be vaulted for posterity and made searchable anywhere in the world. If an author (or an author's family) has only a paper-copy of a thesis/dissertation, the site will give guidance on how to get it scanned into computer-readable form, for subsequent uploading. OpenThesis.org will also work directly with Universities to ensure as comprehensive a database as possible. The existence of a comprehensive ETD database will benefit Universities in licensing efforts, and authors via prospective job opportunities.

As is currently true for www.freepatentsonline.com, visitors to OpenThesis will be able to register in about 1 minute, for free, and gain access to special features such as the ability to organize documents into folders, annotate documents, share documents with other users, and set up alerts for automatic notification of new documents of interest.

A few other important aspects of OpenThesis.org:

- Authors will create accounts that will be e-mail-verified as part of uploading their document, and can choose on their very own author page (www.OpenThesis.org/their-name) to show their contact information and/or LinkedIn profile page as desired.
- Every College and University will have a free microsite, e.g., www.OpenThesis.org/Massachusetts-Institute-of-Technology, such that they can link to a repository of documents from their own authors, co-branded with their logo.
- The site will allow users to search the entire document collection, or to limit the search to a particular institution. The documents from a given institution also can be browsed with that institution's microsite.
- OpenThesis.org will offer, via partner companies, the ability to print and bind theses and dissertations faster and cheaper than through traditional sources.

Theses and Dissertations represent an important compendium of vetted research. OpenThesis.org will ensure that these works, which authors created with formidable investments of inspiration and time, and which universities have ratified as part of awarding graduate degrees, are instantaneously available anywhere in the world via the Web at no cost.

NDLTD Union Catalog / VTLS Visualizer

» Panelist: Vinod Chachra, VTLS

This presentation will discuss in depth the structure and use of the NDLTD Union Catalog. Information will be provided on how the data is harvested, indexed and made available for access. Statistical information will be provided on the size of the database (in excess of 750,000 ETD's) and its source broken by continent, language and country. Useage statistics will be provided showing the source, frequency and pages viewed. Finally pointers will be provided on what to do with the metadata to make dissertations from you institution more accessible to the world at large.

Topical Categorization of Large Collections of Electronic Theses and Dissertations

» Panelist: Edward Fox, NDLTD Director, Department of Computer Science, Virginia Tech

» Panelist: Venkat Srinivasan, Department of Computer Science, Virginia Tech

Objectives: The NDLTD Union Catalog has metadata for over 600,000 Electronic Theses and Dissertations (ETDs) in diverse languages from universities around the world. The users can access these ETDs through various search and browse Web interfaces reachable through the NDLTD Website (example, from Scirus and VTLS). We aim to improve those services in two ways. First, we develop approaches to build larger collections of ETDs, which consist of ETDs not only collected via NDLTD's Union Catalog, but also those collected through focused crawling of many universities' Webpages. Second, we develop approaches in order to make these large collections more amenable to being used by students and researchers.

Methods and Results: We have identified repositories for some universities that host ETDs but that are not yet part of NDLTD. We have developed custom crawlers in order to crawl some of these repositories as well as the NDLTD Union Catalog in order to harvest ETDs and their metadata (where permissible). Our current collection has about 40,000 ETDs from Union Catalog for our initial experimentation, and we actively continue to collect more ETDs.

We also have developed a categorization system, based on the Library of Congress categorization system and Wikipedia, that is more suitable for categorizing ETDs, and have categorized ETDs into the resulting category tree. Users can first browse this category tree based on their needs and then can either browse a particular node, or search it for items of interest.

Conclusions: Through focused crawling, we have been able to increase content available to users, and made it available at a single place. Categorization of ETDs has helped organize the ETDs semantically in order to make it easier to find relevant information. As part of future work, we will improve our methods to collect as many ETDs as possible from the NDLTD Union Catalog and from various universities around the world, categorize them, and provide a Web interface facilitating access.

NEW TRENDS PLENARY PANEL SESSION

» Moderator: Vinod Chachra
» Thursday, June 12 4:15 p.m.–5:45 p.m.

The Semantic Electronic Scientific Thesis

» Panelist: Peter Murray-Rust, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Lezan Hawizy, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Jim Downing, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Joe Townsend, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Nick Day, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Peter Sefton, University of Southern Queensland

We have developed a range of tools and protocols that allow the creation, validation, and re-use of ""born digital"" theses in scientific domains, especially disciplines reporting chemical