NSF Post-DL Futures Workshop Short Paper, by Hsinchun Chen, Professor, University of Arizona

"Towards Building Digital Library as an Institution of Knowledge"

I. What are we trying to do? What is the problem we're trying to solve?

Digital library research aims to develop the engineering and science for generating, capturing, accessing, and utilizing data, information, and knowledge in various digital formats, for a variety of applications, and in a global, collaborative human and system network.

II. How is it done today, and what are the limitations of current practice?

Data and information are captured and represented in various digital formats and are proliferating rapidly. However, the techniques for accessing data and information are rudimentary and imprecise, mostly based on simple keyword indexes, relational queries, and/or low-level image or audio features (i.e., research results of the 70s and 80s).

Digital library should become an institution for capturing and transferring human knowledge, instead of simply for accessing data and information.

III. What is new in our approach/technology, and why do we think it will be successful? What gives evidence that it will work?

A knowledge discovery approach, based on both top-down knowledge creation (e.g., ontologies, subject headings, user modeling) and bottom-up automated knowledge extraction (e.g., data mining, text mining, web mining), promises to help transfer digital library from an institution of data and information to an institution of knowledge.

The rapid proliferation of digital data and information, the increased sophistication and robustness in machine learning and data/text/web mining techniques, and the ever-increasing power of modern networks and computers will help contribute to the new science of digital library.

IV. Assuming we are successful, what difference does it make?

Digital libraries will soon become a ubiquitous, global knowledge resource for education, training, and (international) collaboration. Its impacts will be felt in all aspects of human activities, from industries to governments, and from education to research.

V. How long will it take, how much will it cost, and what are the (measurable) milestones, mid-term, and final exams?

Many fundamental technical, social, and policy issues remain to be developed and researched. It calls for a long-term (ten-year), systematic research agenda to develop digital libraries into an institution of human knowledge. A long-term NSF research program with an annual funding level of $10M-$30M is needed to promote the field and develop the science.

In 3-5 years (mid-term), we expect to see significant advancement in knowledge creation for various digital library applications and domains based on more advanced algorithmic techniques and proper human-computer interaction principles. Instead of accessing low-level, fragmented data and information pieces, high-level, abstract and decision-relevant knowledge will be accessed in a seamless manner.

In 8-10 years (final exams), we expect the systematic science and theory for digital library knowledge creation to be developed and validated. We will be able to access human knowledge in a multi-lingual, multi-media, mobile, and semantics-based digital library knowledge network.