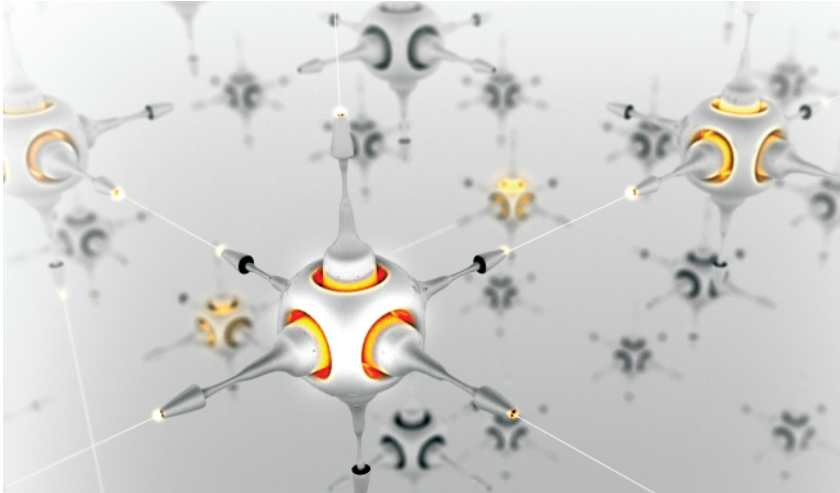


Cyberinfrastructure Laboratory

<http://ist.psu.edu>



“Essentially, we look at how to construct, manage and host digital libraries,” Giles said. “We design, research and implement intelligent tools that help people analyze data.”

Cyberinfrastructure, according to an Educause report, “consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.”

The explosion of big data in recent years has altered the academic research landscape, Giles said.

“In some ways, we’re enabling e-science,” he said.

In 2000, Giles left NEC Research Institute, now NEC Labs, to become the David Reese Professor at the College of IST. He is also an affiliate professor of computer science and engineering, affiliate professor of supply chain and information systems, and associate director of research at the eBusiness Research Center at Penn State. Giles’ current research and consulting interests are in intelligent information processing systems. He has been involved in the creation and development of various novel search engines and digital libraries. He

OVERVIEW

The Cyberinfrastructure Laboratory at the College of Information Sciences and Technology (IST) is directed by C. Lee Giles, David Reese Professor of Information Sciences and Technology, and Prasenjit Mitra, associate professor of IST. The lab, which is located in room 309 in the IST building, was established in 2010. The goal of the lab, Giles said, is to “explore and understand cyberinfrastructure and its impact on science.”

Dr. C. Lee Giles, Director
giles@ist.psu.edu

Dr. Prasenjit Mitra, Director
pmitra@ist.psu.edu

Cyberinfrastructure Laboratory
The Pennsylvania State University
College of Information Sciences
and Technology
319 IST Building
University Park, PA 16802-6823

was one of the creators of the novel metasearch engines, Inquirus and Inquirus2.

Beginning in 1995, Mitra worked for five years at Oracle Corporation in Redwood Shores, Calif. as a senior member of the technical staff at the Server Technologies division developing database software. He also worked part-time as a senior engineer at Narus and DBWizards. His research and teaching interests include database systems, semantic Web, digital libraries, data communications, and computer networks.

The two major projects of the Cyberinfrastructure Lab are CiteSeerx and ChemX Seer. CiteSeerx is a scientific literature digital library and search engine that focuses primarily on the literature in computer and information science. CiteSeerx aims to improve the dissemination of scientific literature and to provide improvements in functionality, usability, availability, cost, comprehensiveness, efficiency, and timeliness in the access of scientific and scholarly knowledge.

Rather than creating just another digital library, CiteSeerx attempts to provide resources such as algorithms, data, metadata, services, techniques, and software that can be used to promote other digital libraries. CiteSeerx has developed new methods and algorithms to index PostScript and PDF research articles on the web.

CiteSeerx, which contains about 3 million documents, gets 1 million hits a day and has about 800,000 unique users.

ChemXSeer, on the other hand, deals exclusively with environmental chemistry. It is an integrated digital library and database allowing for intelligent search of documents in the chemistry domain and data obtained from chemical kinetics.

CiteSeer X and ChemXSeer were both developed with the SeerSuite architecture. SeerSuite is a framework for scientific and academic digital libraries and search engines built by crawling scientific and academic documents from the web with a focus on providing reliable, robust services. In addition to full text indexing, SeerSuite supports autonomous citation indexing and automatically links references in research articles to facilitate navigation, analysis, and evaluation. SeerSuite enables access to extensive document, citation, and author metadata by automatically extracting, storing, and indexing metadata. SeerSuite also supports MyCiteSeer, a personal portal that allows users to monitor documents, store user queries, build document portfolios, and interact with the document metadata.

SeerSuite provides open-source tools that can be used to expand the framework into numerous areas. Other digital libraries and search engines developed using the SeerSuite source code include RefSeer, which recommends documents from the CiteSeerX repository as citations; and ArchSeer, a search engine for archaeological documents.

The Cyberinfrastructure Lab receives funding primarily from the National Science Foundation, and has also received grants from Dow Chemical, Microsoft, and NASA.

Cyberinfrastructure

Cyberinfrastructure Laboratory

Dr. C. Lee Giles, Director
giles@ist.psu.edu

Dr. Prasenjit Mitra, Director
pmitra@ist.psu.edu

The Pennsylvania State University
College of Information Sciences
and Technology
319 IST Building
University Park, PA 16802-6823

PENNSTATE



Penn State is committed to affirmative action, equal opportunity, and the diversity of its workforce. U.ED. IST 13-45