

Keynote Address

Next Generation CiteSeer

Dr. C. Lee Giles
The Pennsylvania State University
University Park, PA
giles@ist.psu.edu

Abstract

CiteSeer, a computer and information science search engine and digital library, has been a radical departure for scientific document access and analysis. With nearly 700,000 documents, it has sometimes two million page views a day making it one of the most popular document access engines in science. CiteSeer is also portable, having been extended to ebusiness (eBizSearch) and more recently to academic business documents (SMEALSearch). CiteSeer is based on two features: actively acquiring new documents and automatic tagging and linking of metadata information inherent in an academic document's syntactic structure. Why is CiteSeer so popular? We discuss this and methods for providing new tagged metadata such as institutions and acknowledgements, new data resources and services and the issues in automation. We then discuss the next generation of CiteSeer.

Bio

Dr. C. Lee Giles is the David Reese Professor at the School of Information Sciences and Technology, Professor of Computer Science and Engineering, Professor of Supply Chain and Information Systems, and Associate Director of Research at the eBusiness Research Center at the Pennsylvania State University, University Park, PA. He has been associated with Princeton University, the University of Pennsylvania, Columbia University, the University of Pisa and the University of Maryland; and has taught at all of the above. His current research and consulting interests are in intelligent information processing systems, including: intelligent portals, novel web tools, search engines, web search and measurement; business models for search and search engines; Web search, knowledge extraction, information and data mining, digital libraries and web databases; computational issues in e-commerce, the e-world, markets and betting; and novel applications of neural and machine learning, agents and AI in web computing, databases, information retrieval, telecommunications, parallel and distributed computing, multi-media, computer systems, neuroscience, adaptive control, system identification, networking, pattern recognition and signal processing, language processing, time series and finance. He has published over 200 journal and conference papers, book chapters, edited books and proceedings in these areas. He has also been involved in the creation and development of various novel search engines and digital libraries. He was one of the creators of the popular computer science search engine, CiteSeer, an autonomous citation indexing search engine for computer science documents. Dr. Giles is a Fellow of the IEEE and of the International Neural Network Society, and a member of AAAI, ACM. He is also a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu. His previous positions include a Senior Research Scientist at NEC Research Institute (now NEC Labs), Princeton, NJ; a Program Manager at the Air Force Office of Scientific Research in Washington, D.C.; a research scientist at the Naval Research Laboratory, Washington, D.C.; and an Assistant Professor of Electrical and Computer Engineering at Clarkson University. During part of his graduate education he was a research engineer at Ford Motor Company's Scientific Research Laboratory. His graduate degrees are from the University of Michigan and the University of Arizona. His academic genealogy includes two Nobel laureates and prominent mathematicians.