

A Service-Oriented Architecture for Digital Libraries

Yves Petinot^{1,2}, C. Lee Giles^{1,2,3}, Vivek Bhatnagar^{2,3}, Pradeep B. Teregowda²,
Hui Han^{1,3}, Isaac Council³

¹Department of Computer
Science and Engineering
The Pennsylvania State
University
111, IST Building
University Park, PA 16802
{petinot,hhan}@cse.psu.edu

²eBusiness Research Center
The Pennsylvania State
University
401 Business Administration
Building
University Park, PA 16802
{vivekb,pbt105,igc2}@psu.edu

³School of Information
Sciences and Technology
The Pennsylvania State
University
332, IST Building
University Park, PA 16802
{giles}@ist.psu.edu

ABSTRACT

CiteSeer is currently a very large source of meta-data information on the World Wide Web (WWW). This meta-data is the key material for the Semantic Web. Still, CiteSeer is not yet a Semantic-enabled service and therefore its meta-data, although potentially usable by Semantic Web agents, is not yet reachable using the Semantic Web mechanisms. The complexity of CiteSeer, that is the range of tasks it supports, make the transition to a Semantic-enabled service a non-trivial task. While human users tend to perceive CiteSeer as a single well-integrated service, we believe it is best seen – from a machine perspective – as a collection of services, each service performing a specific task. In this paper we show our approach to enable CiteSeer on the Semantic Web in order to allow the use of its meta-data through the Semantic Web. We first introduce an intuitive Application Programming Interface (API) to the CiteSeer software, then show that an efficient integration of CiteSeer in the Semantic Web can be best achieved by independently integrating the services that comprise it. We believe the effort presented here towards the Semantic-integration of a complex Information Retrieval system could be used as an integration model for arbitrary systems.