

Advanced Tutorial on the Open Archives Initiative Protocol for Metadata Harvesting

A proposal for a half day tutorial at the
2003 ACM/IEEE Joint Conference on Digital Libraries
<http://www.jcdl.org/jcdl03/>

Submitted by:

Michael L. Nelson
Old Dominion University
Norfolk VA 23529
mln@cs.odu.edu
<http://www.cs.odu.edu/~mln/>

Description

This tutorial is based on the successful tutorial of the same name given at JCDL 2002 and is a follow-on to "Introduction to the Open Archives Initiative Protocol for Metadata Harvesting" (OAI-PMH), to be given earlier the same day (the Introductory Tutorial is being submitted by Tim Cole (UIUC) and we are coordinating regarding the respective contents). It is appropriate for those who have completed the earlier tutorial or are already familiar with OAI-PMH. The tutorial will begin by highlighting the differences between versions 1.1 and 2.0 of the OAI-PMH. Possible migration strategies for 1.1 harvesters and repositories and techniques for mixed version harvesting will be discussed. Advanced topics and deployment scenarios will also be discussed, including: flow control, load balancing, error recovery, hierarchical harvesting, sets and alternate metadata formats.

Objectives

The tutorial attendees will meet the following learning objectives:

- an in-depth explanation of the features and nuances of OAI-PMH 2.0
- a migration strategy from previous versions of OAI-PMH to version 2.0
- a discussion of the advanced and optional features
- a review of current best-practices and implementation guidelines

Intended Audience

The intended audience for this tutorial includes technologists and managers with already familiar with the basic architecture and operation of the OAI-PMH but wish to discuss advanced concepts such as load balancing, error recovery, hierarchical harvesting, alternate interpretations of the OAI-PMH data model and related topics.

Outline

- How 2.0 evolved from SFC and 1.x
 - o people, processes, events
- What's new in 2.0
 - o comparison with 1.x
- Guidelines, recommendations, best practices for 2.0 implementations
 - o harvesters, repositories, aggregators, optional containers
- Novel applications of OAI-PMH

The slides for the tutorial given at JCDL 2002 are available at: <http://www.cs.odu.edu/~mln/jcdl02/>. The JCDL 2003 tutorial will cover similar topics but will draw on experience gained from working with OAI-PMH version 2.0 in the year since its release.

Materials

This tutorial will draw from the OAI-PMH document available from the OAI home page (<http://www.openarchives.org/>) as well as slides prepared by the individual presenters. All materials will be available for download before the conference dates.

Presenters

Michael L. Nelson <mln@cs.odu.edu>

<http://www.cs.odu.edu/~mln/>

Michael L. Nelson received his B.S. in computer science from Virginia Tech in 1991, and his M.S. and Ph.D. in computer science from Old Dominion University in 1997 and 2000. He worked at NASA Langley Research Center from 1991 – 2002, originally in distributed and parallel computing and then shifting to WWW and digital libraries in 1993. He was a visiting assistant professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill for the 2000-2001 academic year. He joined the department of computer science at Old Dominion University in July 2002. Michael is a member of the OAI technical committee.

Simeon Warner <simeon@cs.cornell.edu>

<http://www.cs.cornell.edu/people/simeon/>

Simeon Warner is one of the maintainers and developers of the arXiv e-print archive (<http://arXiv.org/>). He is a member of the Digital Libraries Group in Computing and Information Science at Cornell University. Before that he worked at Los Alamos National Laboratory and the Physics Department at Syracuse University. He has implemented and maintains an OAI interface for arXiv, along with test-suite of harvesting software. Simeon is a member of the OAI technical committee.

Herbert Van de Sompel <herbertv@lanl.gov>

<http://lib-www.lanl.gov/~herbertv/>

Herbert Van de Sompel graduated in mathematics and computer science at Ghent University, and in 2000, obtained a Ph.D. from Ghent University for his research on dynamic and context-sensitive reference linking, now commonly known as the OpenURL framework. From 1982 to 1998 he worked as Head of Library Automation at Ghent University. While at the Los Alamos National Laboratory in 1999, Herbert started the Open Archives Initiative with Paul Ginsparg and Rick Luce. With Carl Lagoze, Herbert forms the executive committee of the OAI; he also is on the technical and the steering committees. Herbert was a Visiting Professor in Computer Science at Cornell University in 2000 - 2001. Afterwards, he was the Director of e-Strategy and Programmes at the British Library. Currently, he is a digital library researcher at the Los Alamos National Laboratory.