Tutorial 5: Advanced Overview of Version 2.0 of the Open Archives Initiative Protocol for Metadata Harvesting

Michael L. Nelson
NASA Langley Research Center
MS 124
Hampton VA 23681 USA
+1 757 864 8511
m.l.nelson@larc.nasa.gov

Herbert Van de Sompel
Los Alamos National Laboratory
Research Library
Los Alamos NM 87545 USA
+1 505 667 4448
herbertv@lanl.gov

Simeon Warner
Cornell University
Department of Computer Science
Ithaca NY 14853 USA
+1 607 254 8605
simeon@cs.cornell.edu

Duration: Half-day
Level: Intermediate / Advanced

DESCRIPTION

This tutorial is a follow-on to "Introduction to the Open Archives Initiative Protocol for Metadata Harvesting" (OAI-PMH), given earlier the same day. 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, and then discuss possible migration strategies for 1.1 harvesters and repositories. Advanced topics and deployment scenarios will also be discussed, including: flow control, load balancing, error recovery, hierarchical harvesting, sets and alternate metadata formats.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *standards*.

C.2.2 [**Computer-Communication Networks**]: Internetworking – *standards*.

General Terms

Standardization, Languages

Keywords

OAI, OAI-PMH, Interoperability, Metadata, Harvesting

1. OUTLINE

1.1 Differences Between Version 2.0 and 1.1

Version 2.0 of the OAI-PMH [1] represents an evolution of the previous versions of the protocol and the collective experiences gained during the first 18 months since the protocol was made public [2]. In addition to an increased formalism in the definition of terms and concepts, the new version of the protocol has a number of immediately visible changes. Highlights include: all OAI-PMH responses are specified in a single XML Schema, as opposed to six separate schemas previously; and all error reporting is handled at the OAI-PMH level, instead of at the transport (http) level. No new verbs are introduced, but some of the existing verbs take new arguments and have slightly different response semantics. Flow control, through use of resumptionTokens, has been refined to offer greater precision and error recovery capabilities.

1.2 Migrations Strategies for Existing Implementations

We will cover experiences in porting existing OAI-PMH 1.1 repositories and harvesters to OAI-PMH 2.0. Examples will be drawn from real word experience involving: arXiv, NTRS, NACATRS, DP9, Arc and others.

1.3 Advanced Topics

The OAI-PMH document has a companion guidelines document [3], in which evolving best practices are recorded for the OAI community. This document includes possible implementations and optimizations of certain features specified in the OAI-PMH. We will present approaches and techniques for flow control, load balancing, error recovery, sets, and alternate metadata formats. We will focus on the motivation, features and implications of the protocol and how these might influence design choices.

2. REFERENCES

[1] Lagoze, C., Van de Sompel, H., Nelson, M., and Warner, S. The Open Archives Initiative Protocol for Metadata Harvesting, 2002. Available at: http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

[2] Lagoze, C., and Van de Sompel, H. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. in Proceedings of JCDL 2001 (Roanoke VA, June 2001), ACM Press, 54-62.

[3] Lagoze, C., Van de Sompel, H., Nelson, M., and Warner, S. The Open Archives Initiative Protocol for Metadata Harvesting Implementation Guidelines, 2002. Available at: http://www.openarchives.org/OAI/2.0/guidelines.htm

JCDL'02, JULY 13-17, 2002, PORTLAND, OREGON, USA.
ACM 1-58113-513-0/02/0007.