Index Object Schema and Replication Infrastructure

LDAP Deployment BoF
Amsterdam 12.5.2000

Peter Gietz
Peter.gietz@directory.dfn.de
Agenda

- Index Object Schema
  - The Problem
  - Crawler Policy
  - TIO schema
  - Proposal
- Replication infrastructure
  - Registration of server
  - TIO distribution
Index Object Schema: The problem

- Plan to set up an index based European White Pages service
- How to define which data to crawl
- How to tokenize the data
Crawler Policy

• Crawler visits only registered server
• Crawler binds to server to authenticate
• Server admin can define crawlable data:
  • via access control mechanisms
  • via a special entry that acts like robots.txt
• Crawler can be configured to only crawl specific entries (defined by Objectclass) and/or attributes
TIO Schema

version: x-tagged-index-1
updatetype: total
thisupdate: 958081068
BEGIN IO-Schema
createtime: FULL
labeleduri: FULL
sn: FULL
postaladdress: FULL
cn: TOKEN
title: FULL
facsimiletelephonenumber: FULL
dn: FULL
mail: FULL
telephonenumber: FULL
objectclass: FULL
personaltitle: FULL
version: FULL
street: FULL
o: FULL
END IO-Schema
TIO schema (contd.)

- Can be defined in LDIF output
- Can be defined by TIO conversion programm ("TIO Collapser")
Proposal for a European White and Yellow pages service

- Search only for objectclass=person, organization or organizationalUnit
- Include only the following attributes into the TIO
  - sn  FULL
  - cn  TOKEN
  - o   TOKEN (person: only in InetOrgPerson)
  - ou  TOKEN (person: only in InetOrgPerson)
  - mail RFC822
  - c   FULL
  - l   TOKEN
Replication Infrastructure