
Austrian On-Line Archive: Current Status and Next Steps

Andreas Rauber¹, Andreas Aschenbrenner¹, Alfred Schmidt²

(¹)Department of Software Technology (ifs)
Vienna University of Technology
<http://www.ifs.tuwien.ac.at>

(²)Austrian National Library
<http://www.onb.ac.at>

IFS

TU

AOLA

Overview

- Project Overview
- Current Status
- Next Steps

AOLA

Austria On-line Archive

- URL: <http://www.ifs.tuwien.ac.at/~aola/>
- Cooperation between the Austrian National Library and the Department of Software Technology, Vienna Univ. of Technology
- Pilot study: preparations since 1999, 1. phase since March 2001
- Linux-System with 240 GB harddisk plus 6-fold tapedrive
- Open source approach to ensure independent access
- Initially: Nedlib harvester (incl. modifications and expansions)
- Goal: snapshot of the Austrian webspace

AOLA

Pilot Phase 1

- Nedlib
- between May 7 and May 16 2001 approx. 10 crawler parallel
- Download during pilot phase: approx. 1GB per day
- at-domain as well as selected subdomains, esp. **.cc*, **.com*, **.edu*, etc.
- Statistics May 7. - 16.:
 - about 666.000 unique URLs harvested
 - 1.210 sites accessed
 - total of 8.3 GB of data stored
 - some problems with Nedlib harvester encountered and partially fixed

AOLA

Pilot Phase 2

- Combine (adapted version - thanks to Kulturaw3-project)
- June 2001
- Download: approx. 7 GB per day
- at-domain as well as selected subdomains, esp. *.cc, *.com, *.edu, etc.
- Stopped on June 21, 2001 - first partial snapshot
- 150 GB, 21.000 sites, 2.7 mio pages
- Stored onto DDS3 tapes
- Basically successful run
- More HDD storage would be helpful for archive organization

AOLA

Results of Pilot Phase

- Basically, the setup works!
- Archiving system problems (XFS file system, Linux kernel 2.4)
- Definitely stick to open-source solutions
- All harvesters require constant monitoring
- HDD space for at least 1 complete run desirable

AOLA

Next Steps

- Creating a complete snapshot
- Storage system decision
 - hierarchical storage systems
 - harddisks to facilitate access
 - back-up copies on tapes
 - distribute back-ups
- Thinking about preservation and future access

AOLA

Conclusions

- **Goal:** Preservation of (modern) cultural heritage
- **Selection:** Combination of manual selection and free harvesting
- **Archiving:** Migration of (hierarchical) storage media
- **Preservation:** Emulation and conversion approaches
- **Access:** Interfaces and legal aspects
- **Urgency:** We have to start **NOW!**

AOLA

AOLA Project-Homepage:

<http://www.ifs.tuwien.ac.at/~aola>