UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

System evaluation of archival description and access

Zhang, J.

Publication date
2011

Link to publication

Citation for published version (APA):
Zhang, J. (2011). *System evaluation of archival description and access*. [Thesis, fully internal, Universiteit van Amsterdam]. IR Publications.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:18 Nov 2026

Abstract

How do archives provide access to their records and let users search? The answer is archival description. Encoded Archival Description (EAD) in Extensible Markup Language (XML) is the *de facto* technical standard for 'electronic' archival descriptions. It is now used to bridge the gulf between tangible records in archives and digital objects on the World Wide Web. These descriptions are finding aids, which are tools to search and find information about, or references to, archival records. The archival finding aids in EAD are left to searchers (out of sight and contact) to explore in unknown ways: how do searchers interact with these finding aids, and what type of retrieval system is needed to support them? The approach of this dissertation is to apply XML retrieval techniques to the EAD finding aids, develop system evaluation of EAD retrieval, and study information seeking behavior of archival search.

The first study involves the design and implementation of the archival search engine README. The README system attempts to incorporate current technologies with the archival structure in finding aids—such as XML retrieval—and simultaneously to uphold the archival principles where this structure is based upon. The system is the proof of concept.

Having established this baseline, the next study explores and tests the construction of an information retrieval (IR) test collection. A test collection is a key component in IR evaluation. The basis of this test collection are the queries and clicks on archival descriptions that can be found in the search log files of the *Nationaal Archief* . There is no readily-available test collection for evaluating the accuracy of the retrieval of archival descriptions of records by an archival search engine. Manually creating such a collection is expensive. The study shows that automatically creating a test collection seems a viable alternative.

Archival principles—such as provenance and original order—are deeply rooted in the arrangement and subsequent description of archival records. The investigation continues by shedding new light on them in a system evaluation. Additionally, the experiments probe XML retrieval-specific issues, such as the retrieval of certain elements. The study concludes by reflecting on the README archival search engine, which is the baseline of the probes in this

dissertation. How effective are certain archival principles for archival access in this digital age?

Using the archival search log files, the research focus shifts to the arrangement of records in EAD and user search behaviors using this arrangement. The sub-document clicks within the finding aids point to the online interaction of users within 'electronic' archival descriptions of records. The analysis of the interactions comprises of quantifying the search behavior. This results in a state diagram that captures different information search behaviors of different people. By analyzing real-world interaction, the discussion on the use of the finding aid in this digital age as access tool becomes more complete. The result is more understanding of online archival search behavior within EAD finding aids, which can be used to improve a search system adapted to existing 'electronic' archival descriptions.

Finally, the system evaluation deals with tailoring a search engine to the different user stereotypes, namely 'expert' and 'novice' groups based on the number of times that a user re-uses the system. The results show that although there are significant differences in terms of search behavior, this does not necessarily mean that for more effective retrieval of archival descriptions, the system needs to be adapted to improve access for these different user groups.