A Clustering-Driven LDAP Framework

VASSILIKI KOUTSONIKOLA and ATHENA VAKALI, Aristotle University

LDAP directories have proliferated as the appropriate storage framework for various and heterogeneous data sources, operating under a wide range of applications and services. Due to the increased amount and heterogeneity of the LDAP data, there is a requirement for appropriate data organization schemes. The LPAIR & LMERGE (LP-LM) algorithm, presented in this article, is a hierarchical agglomerative structure-based clustering algorithm which can be used for the LDAP directory information tree definition. A thorough study of the algorithm's performance is provided, which designates its efficiency. Moreover, the *Relative Link* as an alternative merging criterion is proposed, since as indicated by the experimentation, it can result in more balanced clusters. Finally, the *LP and LM Query Engine* is presented, which considering the clustering-based LDAP data organization, results in the enhancement of the LDAP server's performance.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—*Clustering; information filtering; query formulation;* H.2.1 [Database Management]: Logical Design—*Schema and subschema*

General Terms: Algorithms, Performance

Additional Key Words and Phrases: LDAP services, DIT organization, clustering, merging criteria, query and retrieval engine

ACM Reference Format:

Koutsonikola, V. and Vakali, A. 2011. A clustering-driven LDAP framework. ACM Trans. Web 5, 3, Article 12 (July 2011), 34 pages.

DOI = 10.1145/1993053.1993054 http://doi.acm.org/10.1145/1993053.1993054

1. INTRODUCTION

The explosive growth of Web has increased the need for more robust and scalable distributed networks that are characterized by high performance, high capacity, secure, and reliable services which can be rapidly scaled and managed. The Lightweight Directory Access Protocol (LDAP) [Whal et al. 1997] is an important technology which enables data sharing under an extendable framework for the centralized storage and management of information that needs to be available for today's distributed systems and services. As the name suggests, LDAP is the lightweight version of the Directory Access Protocol and a direct descendent of the heavyweight X.500 [Chadwick 1994], the most common directory management protocol. It is an open industry standard that is gaining wide acceptance due to its flexibility in supporting the storage of heterogeneous data, providing at the same time optimized response times in read operations [Koutsonikola and Vakali 2004].

© 2011 ACM 1559-1131/2011/07-ART12 \$10.00

DOI 10.1145/1993053.1993054 http://doi.acm.org/10.1145/1993053.1993054

ACM Transactions on the Web, Vol. 5, No. 3, Article 12, Publication date: July 2011.

The original idea for this work was presented by the authors in their work "A structure-based clustering on LDAP directory information" at ISMIS'08.

Authors' address: V. Koutsonikola (corresponding author) and A. Vakali, Department of Informatics, Aristotle University, 54124 Thessaloniki, Greece; email: vkoutson@csd.auth.gr.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from the Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.