USING ASSOCIATION RULE LEARNING IN A SEMANTIC FILE SYSTEM

[1]HARSHVARDHAN PANDIT , [2]ASEEM GOGTE, [3]SAHIL GUPTA, [4]ROHIT SHARMA

Department of Computer Engineering, RSCOE, Tathawade, Pune University, Pune
Email: hpandit86@gmail.com, aseem2691@gmail.com, euphoric.sg@gmail.com, rohitc.dude@gmail.com

Abstract— The main concern in information-rich systems is to efficiently navigate and access desired information. Traversing a file system using long pathnames is cumbersome and requires the user to accurately remember where each file can be found. Semantic file systems help in finding a file under various contexts (called as tags) which act like directories. However, the user still has to traverse these paths to reach the files. It is left up to the user to create and manage an efficient system of tags. Since all data items and relations are stored in a database structure, association rule learning can be used to form useful relations between various tags and files. Using various algorithms, we can create associations (links) between data sets that can help the user traverse related data without the burden of long pathnames. Popular data mining algorithms can be readily adapted to a semantic file system's database. Thus utilizing associations, a semantic file system can offer a more efficient and contextual way to search, store and organize data.

Keywords— semantics, indexing, classification, database, tagging, virtual file system, information access, metadata

I. INTRODUCTION

The boom in information has created a situation where it becomes difficult to categorize and search relevant information. Compared to a file system, the web has highly active services and algorithms for data navigation. Tools such as Google, DuckDuckGo [1], and Apple's Siri etc. allow the user to search using keywords and show relevant information by utilizing data mining concepts. Developers creating web services think up of innovative ways to provide easy access to data. However, file system developers are still mostly focusing on stability and performance. While this becomes a necessity in a company server, the home user is more concerned with efficient navigation of data. Semantic file systems address this concern by providing access based on context. However, there is still scope for a richer and rewarding experience of navigating data in a file system.

This paper introduces the concept of using association rule learning in a semantic file system. Association rule learning is a popular and well researched method for discovering interesting relations between variables in databases. By utilizing this, it becomes easy to understand the relativity between different data sets. Since most semantic file systems utilize a database to store and manage meta-information, algorithms such as 'Apriori' can be easily adapted for such applications.

II. RELATED WORK

Over the years, organizing and retrieving information accurately and efficiently has attracted lot of attention. While few have been successful, a number of innovative implementations [2] have emerged. The idea of using a file's semantics as the means to categorize it has been around for quite some time. This section discusses the various implementations made in the field of semantic file system. An efficient implementation of keyword based searching was brought to the desktop by Google's Desktop Search [3] and Apple's Spotlight [4]. Both allow efficient and quick file retrieval based on keywords. They support many file types and have a simple interface which attracts a large number of users. However, both of them are limited to returning search results without any way to organizing contents. In addition, they do not provide any provision to the user for classification of data. This limitation prevented the user from having a personalized way to retrieve data stored by them.

Semantic systems depend on data stored inside the files rather merely relying on a file's attributes. Most implementations use common methodologies like content recognition [5], tagging [6], extracting metadata, etc. to categorize files by using various algorithms.

Each of these has the drawback that although they make organizing data easier, the task of searching relevant data is not tackled. Although Semantic file systems allow the user to browse data by context, but there is simply no provision to recognize the relativity between various contexts.

Using data mining in the semantic-aware file-system, one can easily come up with relations that can help the user get to related data quickly and efficiently.

III. SEMANTIC FILE-SYSTEM

A semantic file system [7] is a virtual file system that uses a database to store and manage file meta-information. It is capable of performing its own interpretation of common file system functionality.

The database is utilized to store tags which are virtual