

directory entries used to categorize files. Files can be tagged with any number of contexts as long no file system rule is violated. The system extracts metadata into tags and stores it in a relational database. These tags can be file attributes such as size, type, name etc. as well as extracted metadata such as author, content title, etc. Categorizing files by metadata allows linking a file in multiple ways while being able to search it using its context. This enables the users to find relevant information in as few searches as possible.

Virtual directories are used to display stored files in a semantic organization. Search results are displayed through dynamically created listings, which correspond to semantic segregation. The entire implementation is based on a virtual file system which manages only the data organization. The underlying file system takes care of storage. This allows it to be ported in future to any file system.

#### IV. APRIORI ALGORITHM

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness [8]. Based on the concept of strong rules, Rakesh Agrawal [9] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection and bioinformatics [10]. As opposed to sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions.

Apriori [11] is a classic algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by apriori can be used to determine association rules which highlight general trends in the database. This has applications in domains such as market basket analysis.

Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation),

where the algorithm attempts to find subsets which are common to at least a minimum number of the item sets.

Adapting the same for a semantic database, we can utilize apriori to show us files which are common to various tags (or directories). Having a cut-off percentage, the apriori results can determine whether two tags are related by looking at files common to them. This way, the user can be shown related information through links to tags obtained from the apriori algorithm.

This example demonstrates a music player like capability where the user is suggested similar music artists based on his organization of files. This is similar to a music player application or on-line service which uses data collected from hundreds of users to show similar tags. The in-built metadata in a music file is also utilized to show similar artists by genre.

The user has various audio files tagged according to their metadata which results in the following setup.

Tag Artists	Tag Rock	Tag Party
Linkin Park Coldplay Metallica Enrique Pitbull Rihanna Green Day Taylor Swift	Linkin Park Coldplay Metallica Green Day	Enrique Pitbull Rihanna

Tags based on Artist and Genre

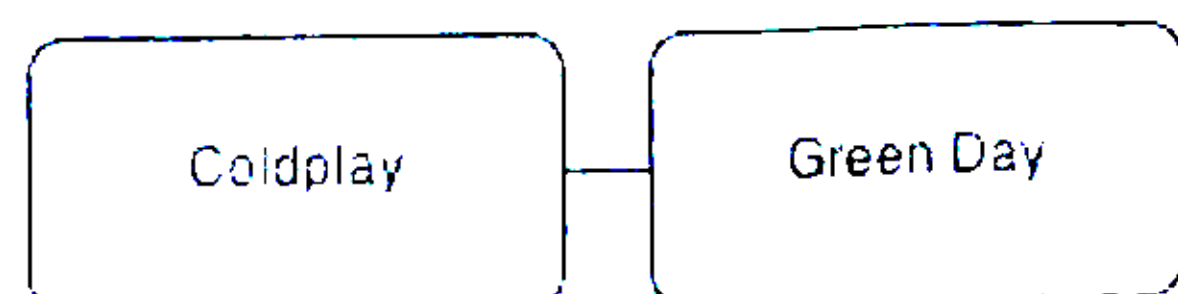
The user then creates new tags to form playlists based on these artists. The user manually (or through some process) adds artists under the new tags created to form the following playlists.

Tag MyBest	Tag Evening	Tag Travel
Linkin Park Coldplay Enrique Green Day	Linkin Park Coldplay Metallica Enrique Pitbull Rihanna Green Day Taylor Swift	Coldplay Metallica Green Day

Playlists (Tags) based on user's preferences

Running an instance of the apriori algorithm on these item sets, we can form the following association rules –

##### 1. Common occurrence



Whenever the tag of files tagged as 'Coldplay' occurs, the tag 'Green Day' is also in the same item set. Therefore, Coldplay and Green Day can be safely assume to be related. Thus, whenever the user wants to create a new play list, and selects one of them, the system can suggest the other to be included as well.