

# A Semantic Cloud for File System Annotation

Tyronda Strong<sup>1</sup>, Pavani Akundi<sup>2</sup>

<sup>1</sup>Computer Science and Engineering, Southern Methodist University, Dallas, TX, United States

<sup>2</sup>Computer Science and Engineering, Southern Methodist University, Dallas, TX, United States

**Abstract** - *There is no standard method to relate topics between documents to retrieve information for a file system. The amount of data in file systems is on the rise and corporate intranets benefit from annotating their files to improve search performance. Search is affected when employees create new information instead of reusing information that is already available or do not mindfully associate sites, folders, and files within the file system. Creating the semantic cloud for file systems with search patterns in mind improves the value of the data on file servers. Semantic annotation and meta-tags methods support the creation of networks of files towards information relevancy for meaningful searches.*

**Keywords:** File systems, meta-data, meta-tag, RDF, semantic annotation, semantic cloud

## 1 Introduction

Web standards, like schema.org, [1] (adopted by Google, Bing, Yahoo, and Yandex) have been helpful in considering annotation of metadata for databases and file systems that are within private networks such as schools and corporations. In a database, content may need metadata annotation or tags to make it machine readable; however, the extraction, analysis, and editing process for annotation is manual. Annotated files allow querying, searching, and reporting of patterns within a dataset, but will be a full time job for anyone attempting to enforce tagging on all the files in their organization. File systems are typically unstructured, without metadata, and do not have a pre-defined data model for searching and organizing which contribute to the challenges involved in structuring files with annotation.

Web search engines have been the primary focus of structuring data, but intranet environments within corporations lack robust search capabilities [2]. Corporate file systems are nonrepresentational file stores of numerous file types and information. Creating, storing, and accessing files in an organization will vary and grow based on the needs of users, teams, and departments. As these data collections grow, improving data processing and search methods is important.

This paper proposes a semantic cloud to configure a file system for search. Although, search may already be an available function in many applications and browsers, defining a standard protocol will assist all users innately manage the

company file structure and assist in the greater goal of building an information network for projects and groups. Utilizing semantic web technologies like information extraction, annotation, and data mapping will help create a semantic cloud for file system search engines. The significance of using a semantic approach in capturing and organizing file system data is to be usable, searchable, and modifiable.

## 2 Problem

Folders and files may lack naming conventions, metadata requirements, and organization standards. Configuration guidelines may or may not exist for users to reference at the time they create folders and files leading users to make up their own definitions. Another matter is to help employees make informed decisions about how to find information in historical documents instead of opting to start from scratch [3]. Any one of these actions could result in duplication of work and general frustration with searching for a file.

For example, Alice works for a pet corporation canine research division. Alice creates File 1 about Labrador retriever illnesses and names it, "Pet Research". Three years later, Bob creates File 2 called "Labrador Retriever Illnesses". When he searched the file system, Alice's paper could not be found because of the vague file name and no metadata about the file. If Alice's document had included meta-tags, then Bob could have saved time by modifying her paper.

## 3 Standard Process

This paper proposes creating a semantic cloud for a file system. The cloud will contain the library of key-terms, recommended meta-tags for new files, and the relationships between files. Search engines match user-entered keywords and phrases to those found within the texts or meta-tag fields of documents, but these are still just word matches and do not necessarily go after the meaning of a document [4].

Windows search returns numerous files that do not meet the intent of the search. In previous versions of Microsoft Office, the search capability was not as robust as a web search engines like Google. When users searched for specific terms the results were too numerous or vague to be useful. With Windows 7, Office now has the ability to search tags and by utilizing tags in our process users will be able to find the data