Tag based File System

Monday, 08.05.2017

Team Members

Utsav Chokshi , *201505581*

Deeksha Singh Thakur, *201505627*

Problem Statement / Idea

- Tagging contents is common practice today. We use tags online for photos, bookmarks and much more. Social networks let us tag almost everything. HTML or other documents contain keywords which are basically tags.
- Despite the heavy use of tags, there is not substantial advances in desktop file system. It is still have same hierarchical (directory based) system.
- So our idea is to create visually appealing tag based file system for desktop for linux operating system.

Scope

Tag Based File System(referred to as TBFS hereafter) allows user to create set of tags and apply those tags to files. Each tag associated with a file is a recognizer for the file, the user can search all files matching subset of tags, get tag suggestion when searching for files and also get suggestion for tags when assigning tags to new files.

Features of the system

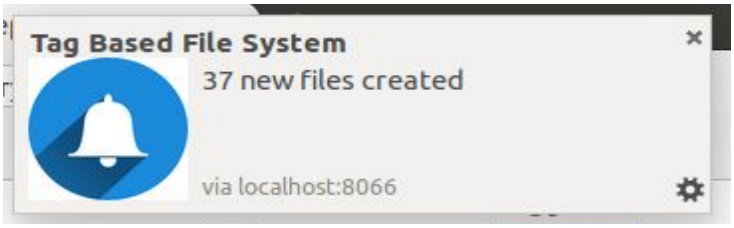
- **Tag assignment :** Assign tags to file. This process can be manual or automatic.
    - In manual process , user can assign tags at the time of creation of file and later it can change tags.
    - In automatic tagging process, tags are assigned to file based on content of file and file type.
    - Few common tags can be assigned to every file. Like extension, directory hierarchy

- **Tag Suggestion :** Tag Suggestion is based on 3 parameters
    1. **File Type** - the file extension gives idea about the type of file, also the file metadata gives inside information about file.
    2. **File Content & Metadata** - The object in images, the content of text files, metadata about music and video files etc are used for suggesting tags.
    3. **Apriori Based** - Frequently co-occurring tags is suggested by mining interesting association rules using apriori algorithm.
    4. **Popular Tags -** TBFS maintains the frequency count of tags assigned to files. The tags that have been most popularly assigned are shown in the suggestion list.

- **File retrieval :** retrieve files based on tags assigned to it. This process is two fold :

- **Showing Related Tags** : When user searches for file using some particular tag, tags which are related to that tag are shown (using graph). So that list of filtered files can be minimized.
- **Retrieving Filtered Files :** Based on set of tags selected by user, list of files matching those tags are shown.

- **Notification for newly created/modified files:** The home page displays the list of newly created/updated files. User can choose to assign tags to these files. Redirection from home page to assign tags page is possible.



Notifications are shown on top right corner of the user screen. It shows the count of newly created/modified files in the system. On clicking the notification, user is redirected to home screen of TBFS.

- **Update of file :**
  - Update tags for file - If the file has previously assigned tags, the tags are loaded and displayed to user for editing, adding new tags/ delete existing tags. Update tag suggestions automatically based on change of file content (in case of automatic tagging).

- **Deletion of file :** TBFS allows to delete tags assigned to file. In case file is deleted from system, the cron job handles the purging of TBFS database after every 24 hours. The complete record of file is deleted from system
- **Cron Jobs :** The TBFS system has 3 cron jobs, processes running in the backend.

CRON1 : Creating Association Rules

- ❖ Whenever new tags are added to the system, new association rules are created. A cron process looks after updation of 'rules.pkl' file where association rules for tag are stored.
- ❖ Using Apriori data mining algorithm, we find relation between tags in the system. These association rules are then used to suggest tags when assigning tags to files, and give search suggestions to user. Assumption, it is highly likely that if tag1 and tag2 are assigned to some file, the tag assignment would be useful of file of same type. Hence the association rules for tags.

CRON2 : Record for New Files in system

- ❖ For all new files in the system, a record needs to be created with the TBFS. Our cron process runs every 120 minutes to monitor any modification/creation of new files in the system. List of modified files is maintained and their tagged status is set to 'false' till tags are assigned to these files.
- ❖ On every modification of file, TBFS would set tagged status of file to false.(to allow tags to be added/removed as per change in content of the file). Browser based html5 notification are generated to inform the user of modified files in system demanding his attention to assign tags to modified files.

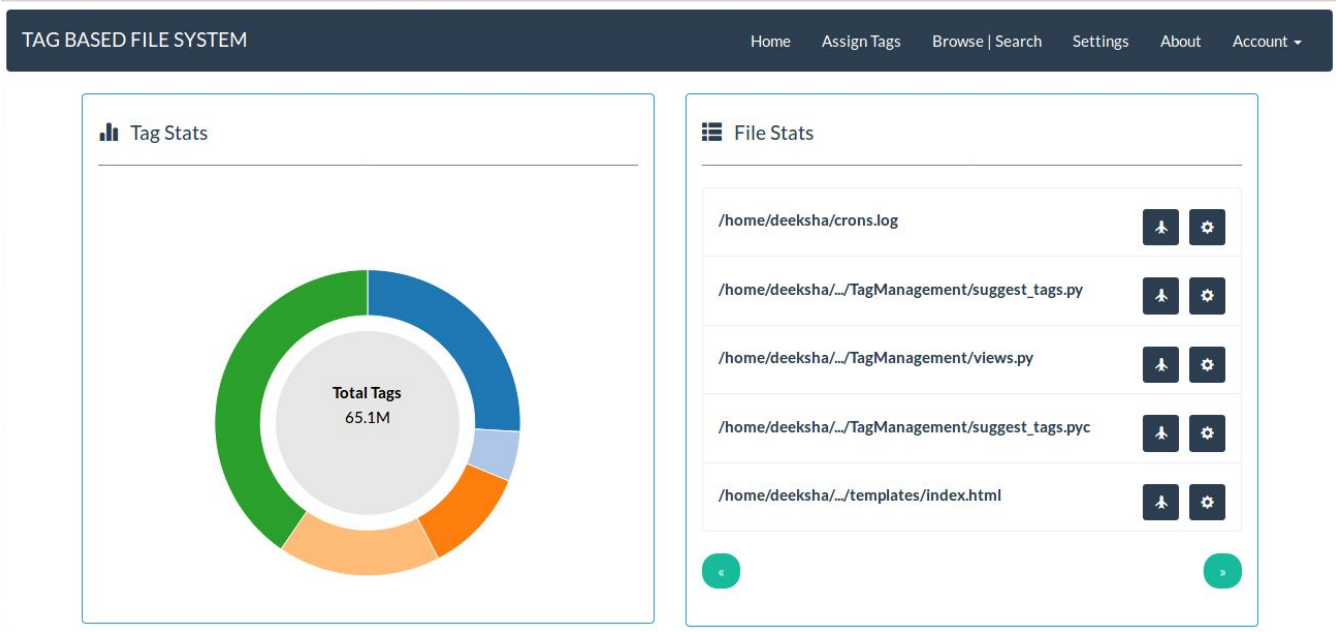
CRON3 : Purging file_info database

- ❖ To handle deletion of files from system, a DB purging operation runs every 24 hours. This cron job has the responsibility of deleting the file records from TBFS database which no more exist in the system. On

every deletion of file, the file record is deleted from file_info table, the tag frequency of associated tags is decremented by 1.

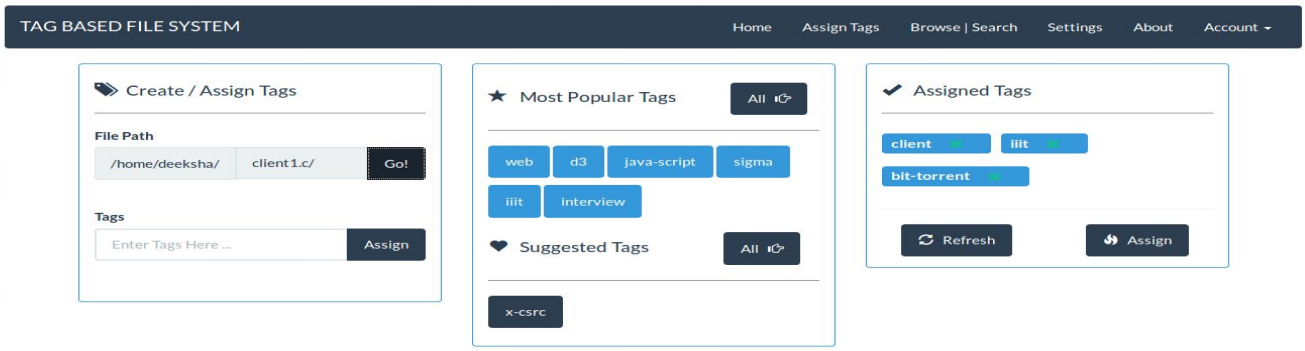
User Interface

1. Home Screen



- Showing tag Statistics of System
- List of modified files in system waiting to be assigned tags, with buttons for auto assign and manual assign in each row.

2. Assign Tags



The following input options are available : Part 1

1. Input file path, filepath of the file you wish to assign tags to.

   Feature of autocomplete filename is provided for browsing through directory for appropriate file

   Feature of 'Assign All' is provided to allow assigning all popular tags to list of assigned tags

2. Tags that you wish to assign to file.

   Feature of autocomplete tag deriving from the list of previously assigned tags is provided to user.

   Feature of 'Assign All' is provided to allow assigning all suggested tags to list of assigned tags

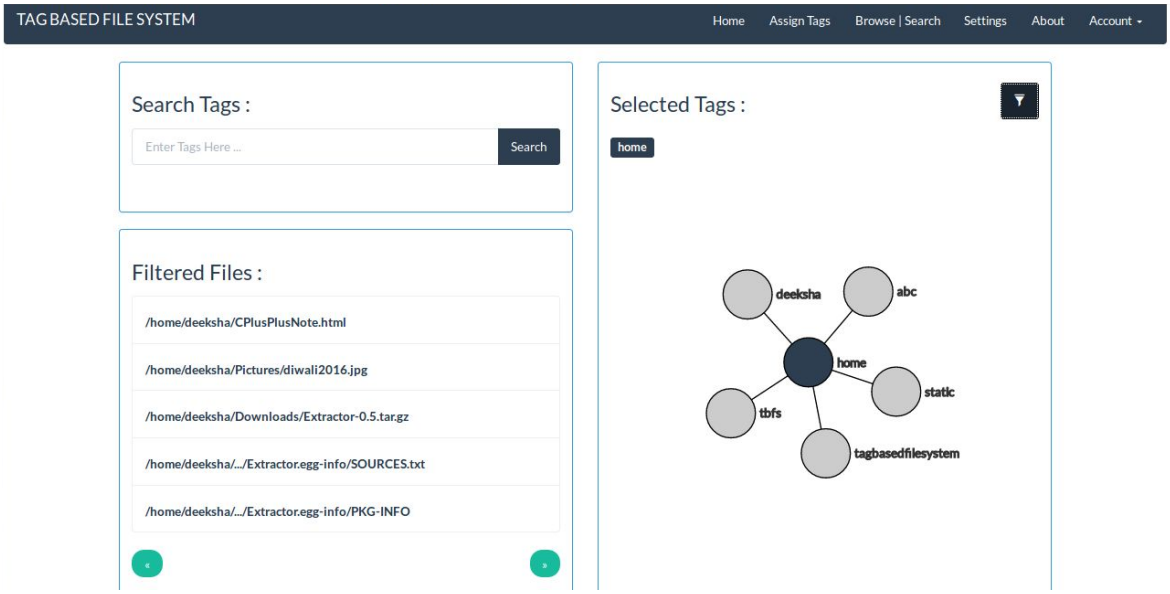
The following suggestion list is shown to user :

1. Popular Tags - Showing list of most popular tags in system
2. Suggested Tags - By asking for tag suggestion on basis of file type, ts content and other related tags in system.

The list of assigned tags / to be assigned tags is displayed in part3 of the page. On clicking on Assign - All tags get assigned to the file.

On clicking refresh the page is reloaded.

3. Browse and Search



Search tags with tag names.

- Use of d3.js to view connected tag with your tag, to filter more files based on your interest.
- Filter button retrieves all files with the tags chosen by the user. Pagination is implemented to view multiple paged results. (5 entries per page is shown)

Deployment/Configuration

PERMISSIONS :

The following set of permission shall be required to run TBFS :

1. TBFS needs permission to read user's file content. (The content shall remain local to user's system and will not be shared over the network.)
2. TBFS shall run on a specific port. Currently set to 9000.
3. Allow TBFS to send browser notifications to monitor updates in system and inform user about the same.

USER SETTINGS : On first time use of TBFS, the user is required to set the following parameters in user settings -

1. HOME : Location where TBFS is deployed, "path to manage.py to DJANGO project"
2. MOUNT_DIR = "mount directory of the system"
3. SAVE_NOTIFICATION_FOR_X_HOURS = “the duration in hours for which user wants to receive notifications”
4. FOLDER_HEIRARCHY_AS_TAG_FLAG

   True, if user wants the folder hierarchy of file to be saved as tags by default

   False, if user folder hierarchy should not be assigned as tags to each file.

Link to github repo :
https://github.com/ChokshiUtsav/TagBasedFileSystem