TagFS

Tag Filesystem

TagFS: Organizing Information Using Tags

http://code.google.com/p/tagfilesystem

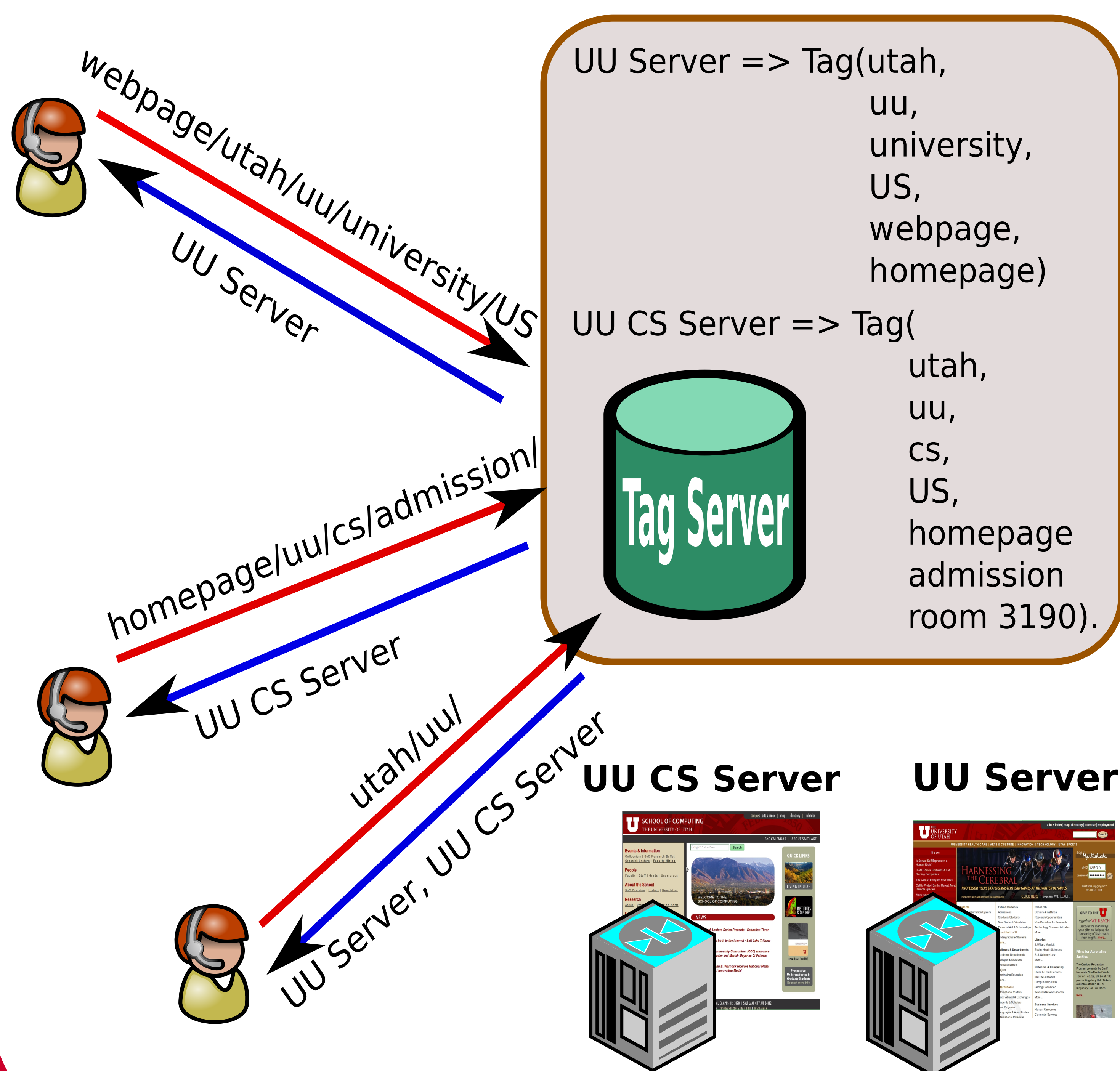Weibin Sun, Robert Ricci

THE UNIVERSITY OF UTAH

Motivation

Bringing search into computing infrastructure

- When creating information, tag it for search
- Categorize information using tags
- Locate information by tag query

Enabling search at different scales:

- **Within local machine: a tag-based filesystem, which is the TagFS here**
- **The Internet scale:**
  - An Internet with search engines built in the infrastructure. (Relax, Google! :))
  - Using tag-servers instead of DNS naming servers

webpage/utah/uu/university/US

UU Server

UU Server => Tag(utah, uu, university, US, webpage, homepage)

UU CS Server => Tag(utah, uu, cs, US, homepage admission room 3190).

homepage/uu/cs/admission/

UU CS Server

Tag Server

utah/uu/

UU Server, UU CS Server

UU CS Server   **UU Server**

What is TagFS?

A filesystem in which:

- Directories are tags, or tags are directories.
- Files are associated with tags that can reflect their properties and contents.
- A directory path, say /Photo/Utah/Spring , is a query that has the result similar to 'select files that have tags: Photo AND Utah AND Spring'.
- Tags have no hierarchical structure like directory tree in traditional filesystem. So /Photo/Utah/Spring and /Utah/Photo/Spring have the same files.

Different from desktop search applications:

- TagFS is a real filesystem. Existing applications can benefit from it without modification. No extra syscalls and APIs introduced.
- TagFS searches user-annotated tags while desktop search applications search file contents.

Filesystem APIs

mkdir: create a new tag

rmdir: remove a tag and untag all files associated with it

creat/open: when creating a new file, the file is also tagged with the tags in the given path.

unlink: untag the file by removing tags in the given path, if no tags are left, the file is deleted.

readdir: readdir, which is the core function of 'ls', lists:
- files with the given tags
- the remaining tags of the listed files
as the following example shows:

Tags: Utah, Arch, Photo

Tags: Utah, Campus, Photo

>ls Utah/Photo/

Arch/   Campus/

>ls Utah/Arch/

Photo/

TagFS Implementation

Challenges:

- Integrating tag query into filesystem APIs without helps from extra interfaces and special path formats
- Dealing with files with the same name in a tag query
- Mapping tag semantics to standard UNIX commands
- Efficient data structures to do tag query

Technical details:

- A FUSE filesystem written in Python
- Support all standard Linux filesystem system calls
- Single threaded(for now)
- Hashtable to store associations between tags and files
- Running well on Linux
- May crash on Mac or some GUI file browsers

Try TagFS

Download and install:
Go to *http://code.google.com/p/tagfilesystem* to download the released TagFS packages or check out the most recent sources to play with it.

Current status:
Some brave users are using TagFS to store/organize papers, music, photos and videos.