

File indexing and searching

Francisco J. Ballesteros

Laboratorio de Sistemas

nemo@lsub.org



Motivation

- Where is this mail?
- Where is this C function?
- How could Plan 9 do this?



Tools

- APE tools:
 - Take time to start
 - Expensive indexing
- Plan 9's:
 - Pre-loaded file system
 - Detect changed files



Data structures

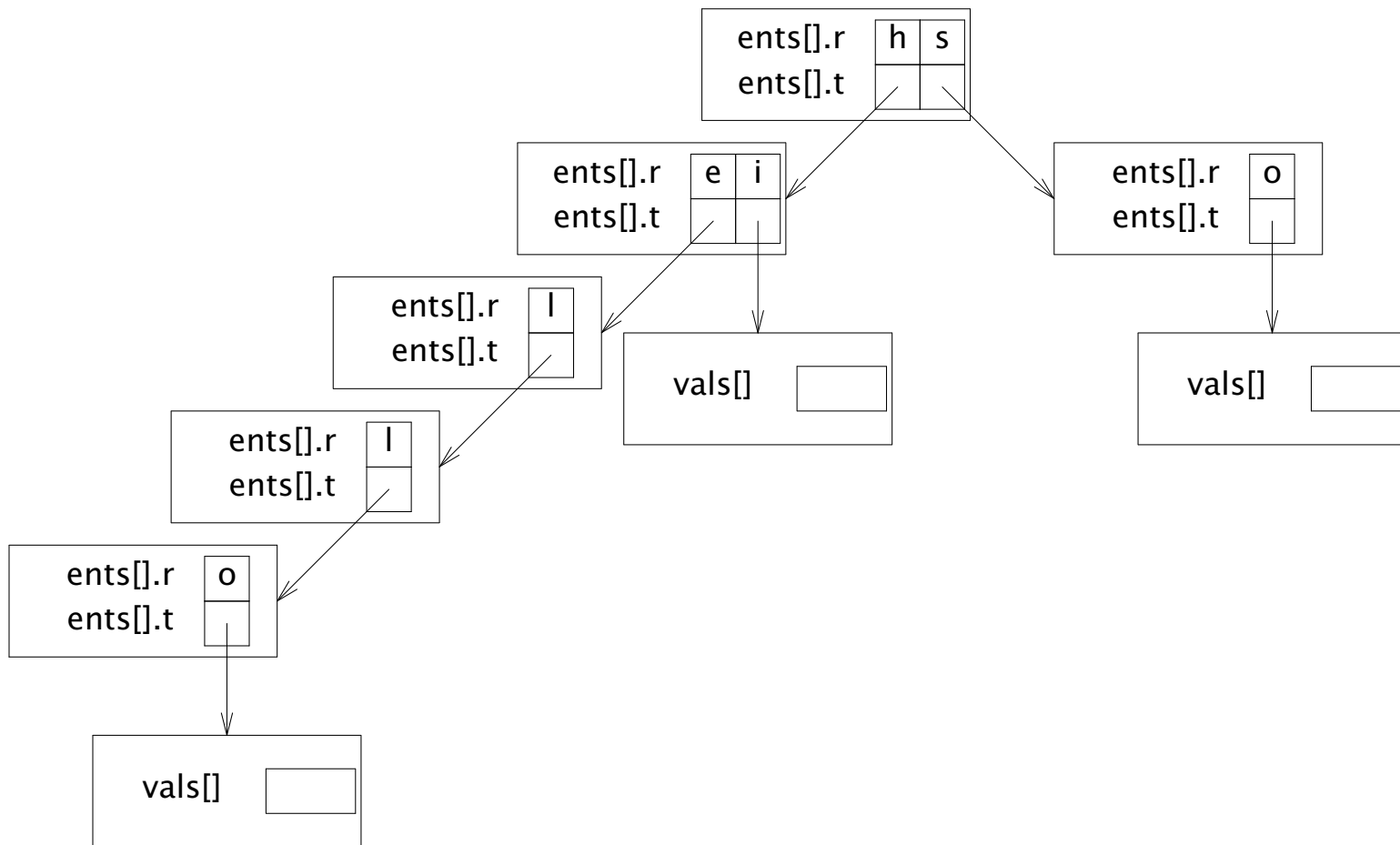
- Per file-system DB
- A file server:
 - provides files
 - provides search



Data structures

- Trie
 - Prefix \Rightarrow Qids
- Hash
 - Qid \Rightarrow path





Tools

`mktags dbath file...`

`[DB=path] looktags [-n] file...`

`tagfiles trie file...`

`rdtrie trie tag...`

`qhash hash qid...`

`qhash -a hash qid path...`

`tagfs trie`



Examples

```
mktags $home/lib/$user $home /mail/...
```

```
files=`{changes $home /mail/.../msgs}  
tagfiles /mnt/tags $files
```

```
looktags list append : queue append
```



Search process

- Search the trie for tags
- Find paths for qids
- If they exist
 - list
 - or use grep to list a few lines



Indexing process

- Assume text in most cases
- Use external tools
 - tagc
 - taglimbo
 - etc.
- What is a word?



Sharing the DB

- Locating the DB:
 - Try /mnt/tags
 - Use ndb to reach tagfs
 - Use rdtrie



Security

- Per file system DB.
- Don't index what you don't want to find.
- We do index.



Experience

- 148736K system DB
- 236808K personal DB
- searches in 1 to 5 seconds.

