Search Engines

Victor Milenkovic

Department of Computer Science
University of Miami

CSC220 Programming II – Spring 2019

Hard Disk as Map

A hard disk is really a Map from Long to File.

Hard Disk as Map

A hard disk is really a Map from Long to File.

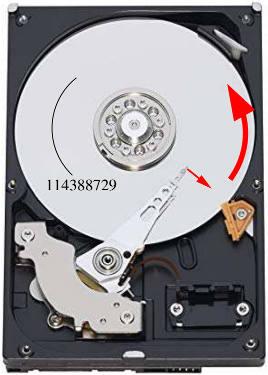
- Disk files are one or more *blocks*.

Hard Disk as Map

A hard disk is really a Map from Long to File.

- Disk files are one or more *blocks*.
- Disk controller takes block *index*.

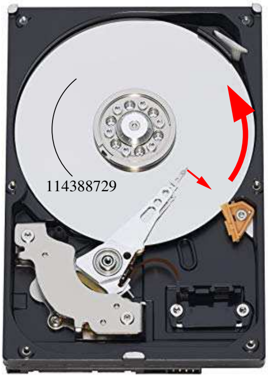
Hard Disk as Map



A hard disk is really a Map from Long to File.

- Disk files are one or more *blocks*.
- Disk controller takes block *index*.
- Moves read head to correct track.

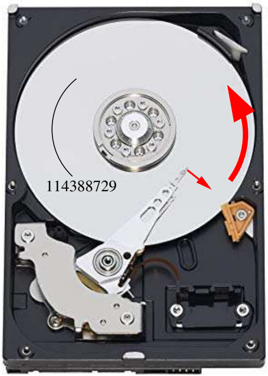
Hard Disk as Map



A hard disk is really a Map from Long to File.

- Disk files are one or more *blocks*.
- Disk controller takes block *index*.
- Moves read head to correct track.
- Waits for disk to rotate to beginning of file.

Hard Disk as Map



A hard disk is really a Map from Long to File.

- Disk files are one or more *blocks*.
- Disk controller takes block *index*.
- Moves read head to correct track.
- Waits for disk to rotate to beginning of file.
- In about one millisecond.

Hard Disk as Map

A hard disk is really a Map from Long to File.

- Disk files are one or more *blocks*.
- Disk controller takes block *index*.
- Moves read head to correct track.
- Waits for disk to rotate to beginning of file.
- In about one millisecond.
- Read or writes file.

Page and Word Files

Page file:

Page and Word Files

Page file:

- page file index (redundant)

Page and Word Files

Page file:

- page file index (redundant)
- URL

Page and Word Files

Page file:

- page file index (redundant)
- URL
- reference count

Page and Word Files

Page file:

- page file index (redundant)
- URL
- reference count
- 27(edu.miami.cs.www/˜vjm/csc220/google/little.html)2

Word file:

Page and Word Files

Page file:

- page file index (redundant)
- URL
- reference count
- 27(edu.miami.cs.www/˜vjm/csc220/google/little.html)2

Word file:

- list of page file indices

Page and Word Files

Page file:

- page file index (redundant)
- URL
- reference count
- 27(edu.miami.cs.www/˜vjm/csc220/google/little.html)2

Word file:

- list of page file indices
- 16(water)[3, 7, 11, 14, 16, 19, 20, 27]

Page and Word Files

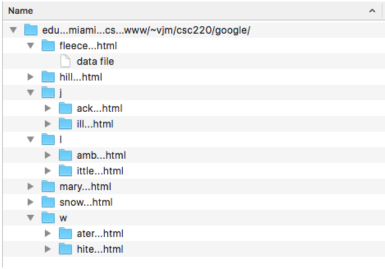
Page file:

- page file index (redundant)
- URL
- reference count
- 27(edu.miami.cs.www/˜vjm/csc220/google/little.html)2

Word file:

- list of page file indices
- 16(water)[3, 7, 11, 14, 16, 19, 20, 27]
- We get its index and string from context.

External Compressed TRIE

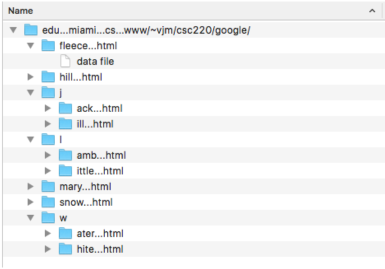


Name

- edu...miami...cs...www/~vjm/csc220/google/
  - fleece...html
    - data file
  - hill...html
  - j
    - ack...html
    - ill...html
  - l
    - amb...html
    - ittle...html
  - mary...html
  - snow...html
  - w
    - ater...html
    - hite...html

Map from URL to index

External Compressed TRIE

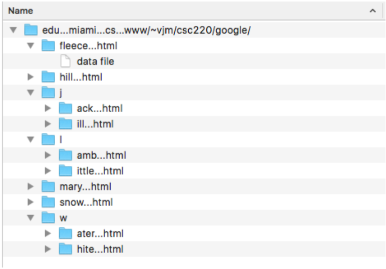


```
Name
▼ edu...miami...cs...www/~vjm/csc220/google/
    ▼ fleece...html
          data file
    ▶ hill...html
    ▼ j
        ▶ ack...html
        ▶ ill...html
    ▼ l
        ▶ amb...html
        ▶ ittle...html
    ▶ mary...html
    ▶ snow...html
    ▼ w
        ▶ ater...html
        ▶ hite...html
```

Map from URL to index

- Compressed External TRIE

External Compressed TRIE

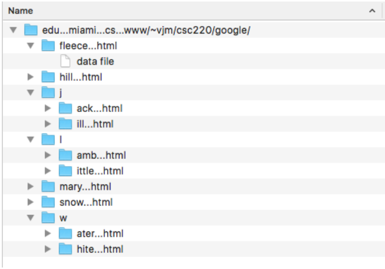


```
Name
edu...miami...cs...www/~vjm/csc220/google/
    fleece...html
        data file
    hill...html
    j
        ack...html
        ill...html
    l
        amb...html
        ittle...html
    mary...html
    snow...html
    w
        ater...html
        hite...html
```

Map from URL to index

- Compressed External TRIE
- . replaced by ... for technical reasons

External Compressed TRIE



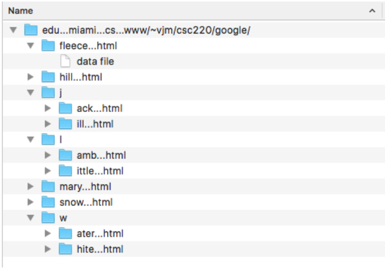
Name

- edu...miami...cs...www/~vjm/csc220/google/
  - fleece...html
    - data file
  - hill...html
  - j
    - ack...html
    - ill...html
  - l
    - amb...html
    - ittle...html
  - mary...html
  - snow...html
  - w
    - ater...html
    - hite...html

Map from URL to index

- Compressed External TRIE
- . replaced by ... for technical reasons
- sub from internal TRIE is now folder name

External Compressed TRIE



```
Name
edu...miami...cs...www/~vjm/csc220/google/
  fleece...html
    data file
  hill...html
  j
    ack...html
    ill...html
  l
    amb...html
    ittle...html
  mary...html
  snow...html
  w
    ater...html
    hite...html
```

Map from URL to index

- Compressed External TRIE
- . replaced by ... for technical reasons
- sub from internal TRIE is now folder name
- value is stored in data file in folder