

---

# Perspective: Semantic data management for the home

Brandon Salmon,  
Steve Schlosser\*, Lorrie Cranor,  
Greg Ganger  
\* Intel Research Pittsburgh

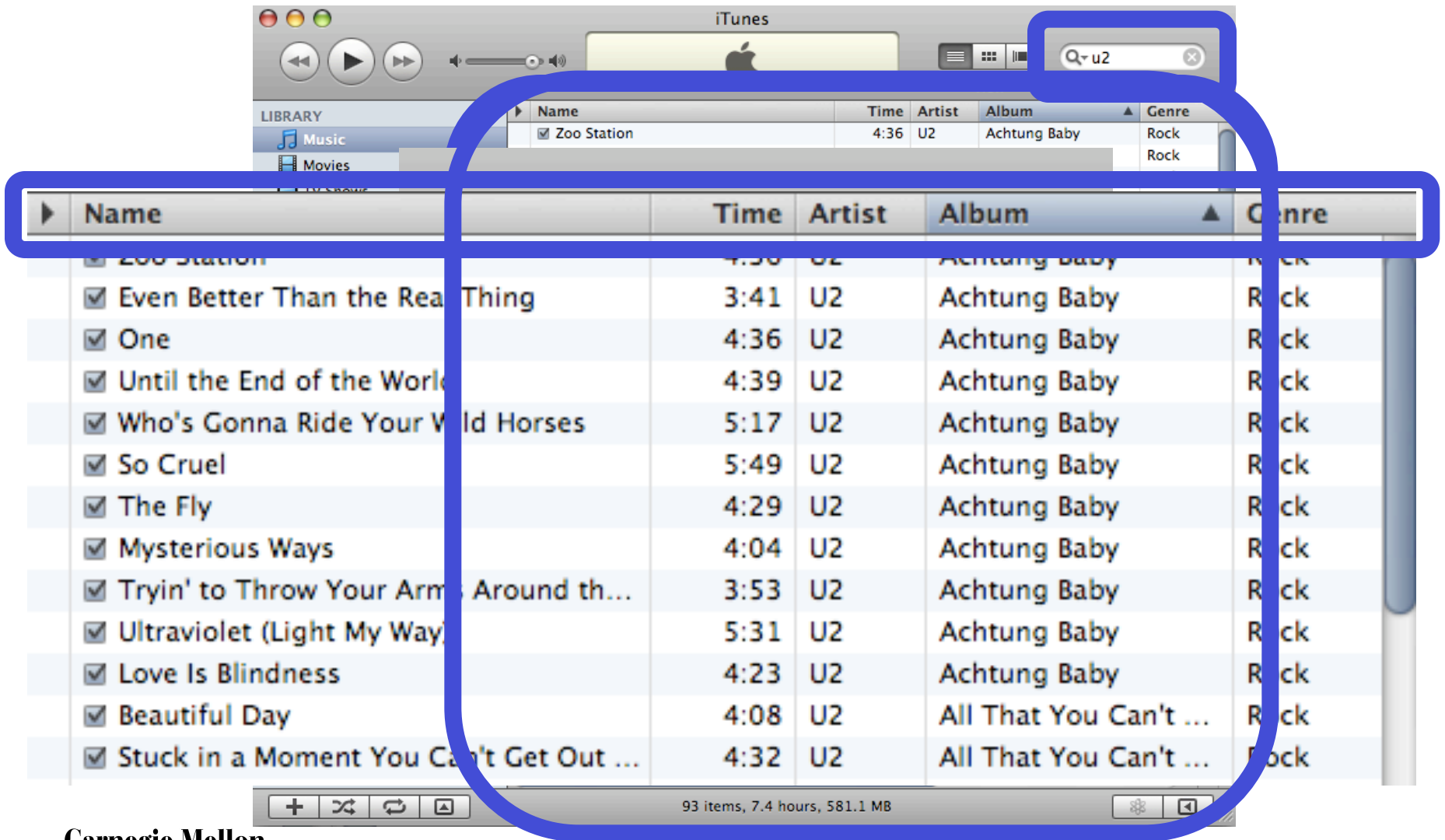
# Mary's music

---

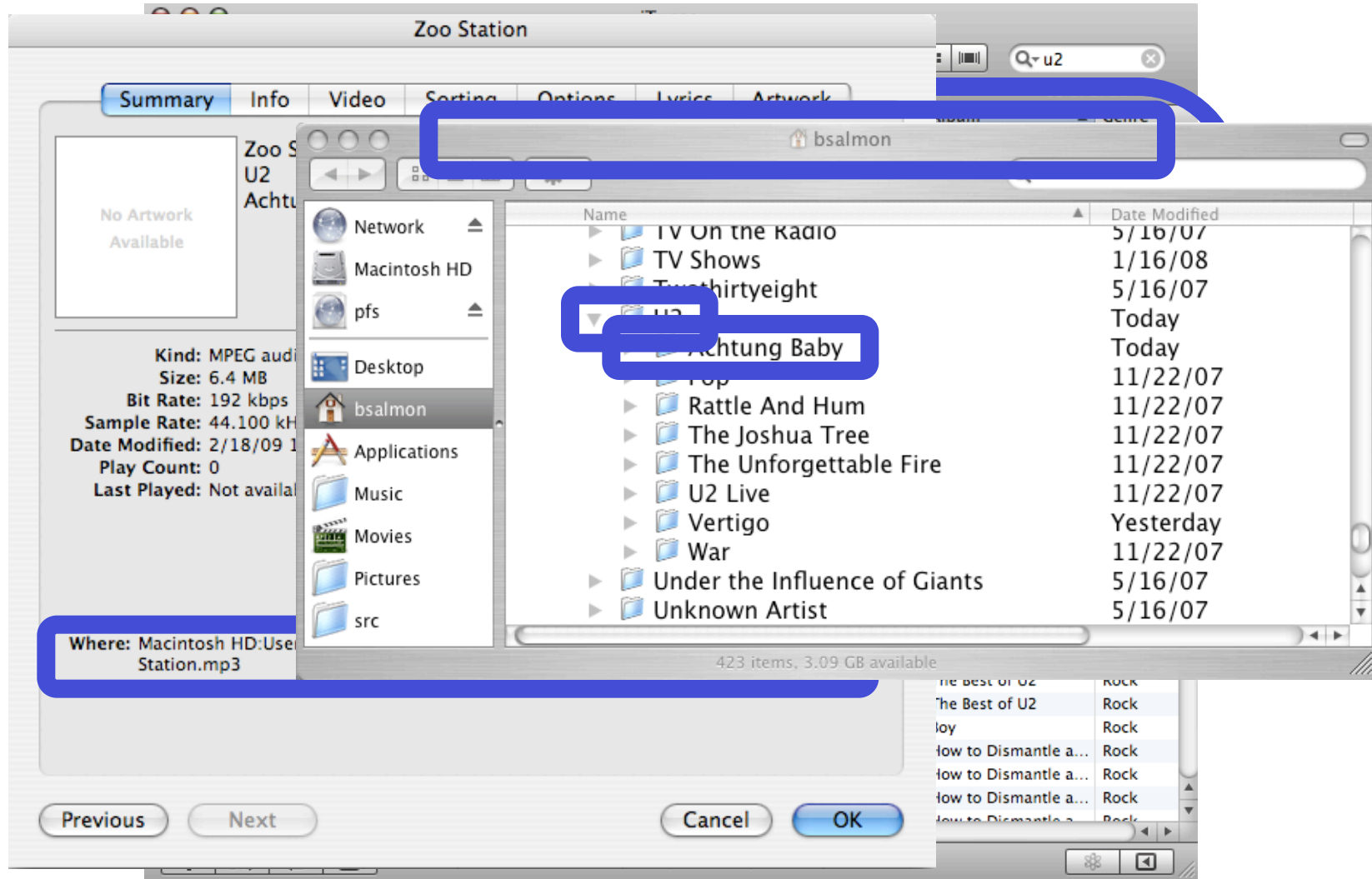
- Roommates listen to music on Carol's desktop
  - It has better speakers
- Mary wants U2 for an upcoming party
- So, she'll copy U2 files to Carol's desktop
  - Using a USB dongle



# Moving U2

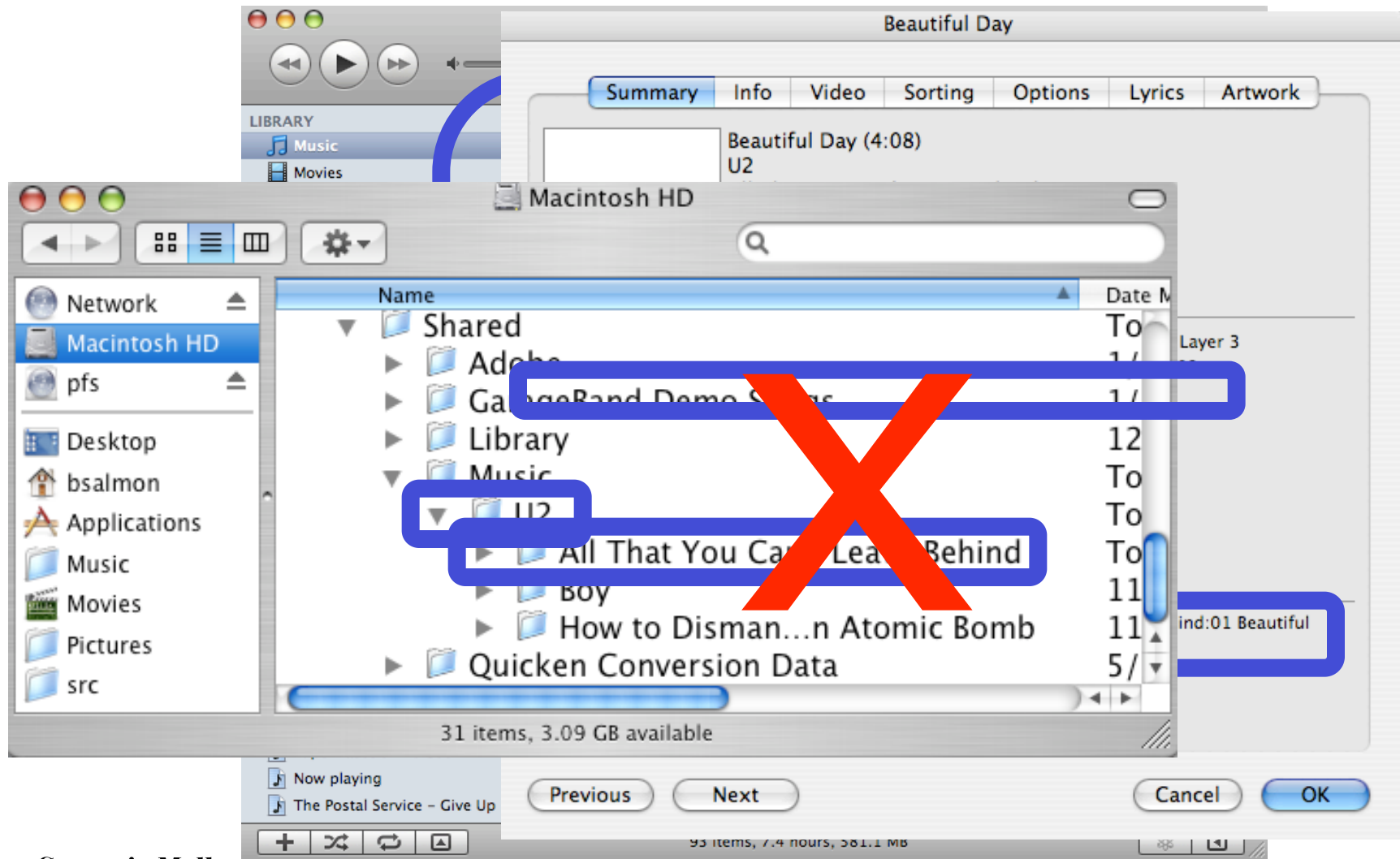


# Moving U2





# U2 moves in mysterious ways



# How did this get so complicated?

---

- Naming mismatch between app and tool
  - Application (iTunes): semantic naming
  - Management tool (Finder): hierarchical naming
  - Not just an interface problem
    - Design of the system is critical
- The big problem is management
  - Problems occur when system and usage diverge
  - Even simple tasks can become complex
  - End user is a “real person”, not a techie

# Our methodology

---

- Explored what home users need
  - In-situ, semi-structured interviews
- Built a prototype filesystem: Perspective
  - Addresses discovered management challenges
- Tested Perspective's usability in lab
  - Have shown view-based management is more usable than directory and volume approaches
  - We tested “real users” – not tech gurus

# Contextual analysis

---

- In-situ semi-structured interviews
  - Interviewed 24 people in 8 households
  - Asked about data management practices
  - Also asked about data and device specifics

For example:

Describe the last time you backed up your data. Can you step me through a backup right now?



**Carnegie Mellon**  
**Parallel Data Laboratory**

<http://www.pdl.cmu.edu/>

# Core home management tenets

---

- Users are comfortable with semantic naming
- Devices are decentralized and dynamic
- Users require control over automation
- User place data infrequently and explicitly
- Users require low cost / high utilization

# Our methodology

---

- Explored what home users need
  - In-situ, semi-structured interviews
- **Built a prototype filesystem: Perspective**
  - Addresses discovered management challenges
- Tested Perspective's usability in lab
  - Have shown view-based management is more usable than directory and volume approaches
  - We tested “real users” – not tech gurus



# Perspective addresses core tenets

---

- Peer-to-peer architecture
  - Loosely coupled devices
- Semantic management
  - Move data, set reliability on semantic groupings
- Rule-based data placement
  - Placement rules automatically place new data
- Transparent automation
  - Human-understandable language for tools

# Perspective is a distributed fs

---

- Global namespace
  - Can read/write data from any accessible device
  - Can search for data from any device
- Files accessible through FUSE
  - Can be accessed by unmodified applications
- Eventual consistency
  - Data can always be modified from any device
  - All replicas of file converge to same version
  - Topology independent
  - Conflicts are handled similarly to previous systems

# Views

---

- View specifies the files stored on a device
  - Uses a semantic query to specify a set of files
  - Perspective ensures view files are stored on device
- Semantic query
  - Query on name-value pair attributes
  - Example: files where ***album = The White Album***  
files where ***task = Taxes and create time > January 1, 2008***
- Use queries to manage data, not just locate it
  - Views provide data placement, fault tolerance

# View provide: update routing



Brian laptop: *owner=Brian*

name = "Galileo"  
type = "TV Show"  
series = "West Wing"  
owner = "Brian"



Mary cell: *type=address book  
and owner=Mary*



Family desktop: *all files*



DVR: *type=TV Show*

DVR: *sharing=Family*

Also used to limit costs on synchronization,  
and direct searches to appropriate devices

# View-based management

---



# Rule-based management

---

Mary laptop: *owner=Mary*





# Semantic management

---



Mary laptop: *owner=Mary*

Mary laptop: *owner=Mary*  
*and create time > 2007*

Mary's laptop is short on disk space, but the Desktop has extra disk space. Should I put Mary's old files on the Desktop to free up space?



Desktop: *owner=Mary*  
*and create time <= 2007*

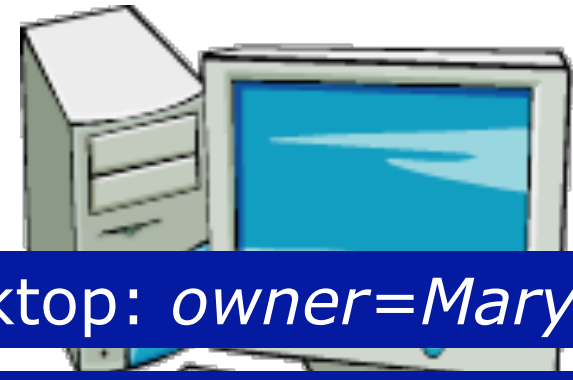
# Transparent automation

---



Mary laptop: *owner=Mary  
and create time > 2007*

Mary laptop: *owner=Mary  
And type=Music*



Desktop: *owner=Mary*

Desktop: *owner=Mary  
and create time <= 2007*

# Loosely coupled devices

---



Mary laptop: *owner=Mary*  
*and create time > 2007*

Mary laptop: *owner=Mary*  
*And type=Music*



Desktop: *owner=Mary*



Mary laptop: *owner=Mary*  
*and create time > 2007*

Mary laptop: *owner=Mary*  
*And type=Music*

# Our methodology

---

- Explored what home users need
  - In-situ, semi-structured interviews
- Built a prototype filesystem: Perspective
  - Addresses discovered management challenges
- Tested Perspective's usability in lab
  - Have shown view-based management is more usable than directory and volume approaches
  - We tested “real users” – not tech gurus

# Usability evaluation

---

- Is view-based management easier than directory or volume-based management?
- Between-group lab usability study
  - Non-technical university students and staff
  - Tested 30 users, each using one system
  - Each user did 10 tasks in latin square ordering
- Use the same interface, change system
  - View-based system (semantic, p2p)
  - Directory-based system (hierarchical, p2p)
  - Volume-based system (hierarchical, client/server)

# View manager interface

- None of these files stored here
- Some of these files stored here
- All of these files stored here
- Not all protected from failure
- All protected from one failure





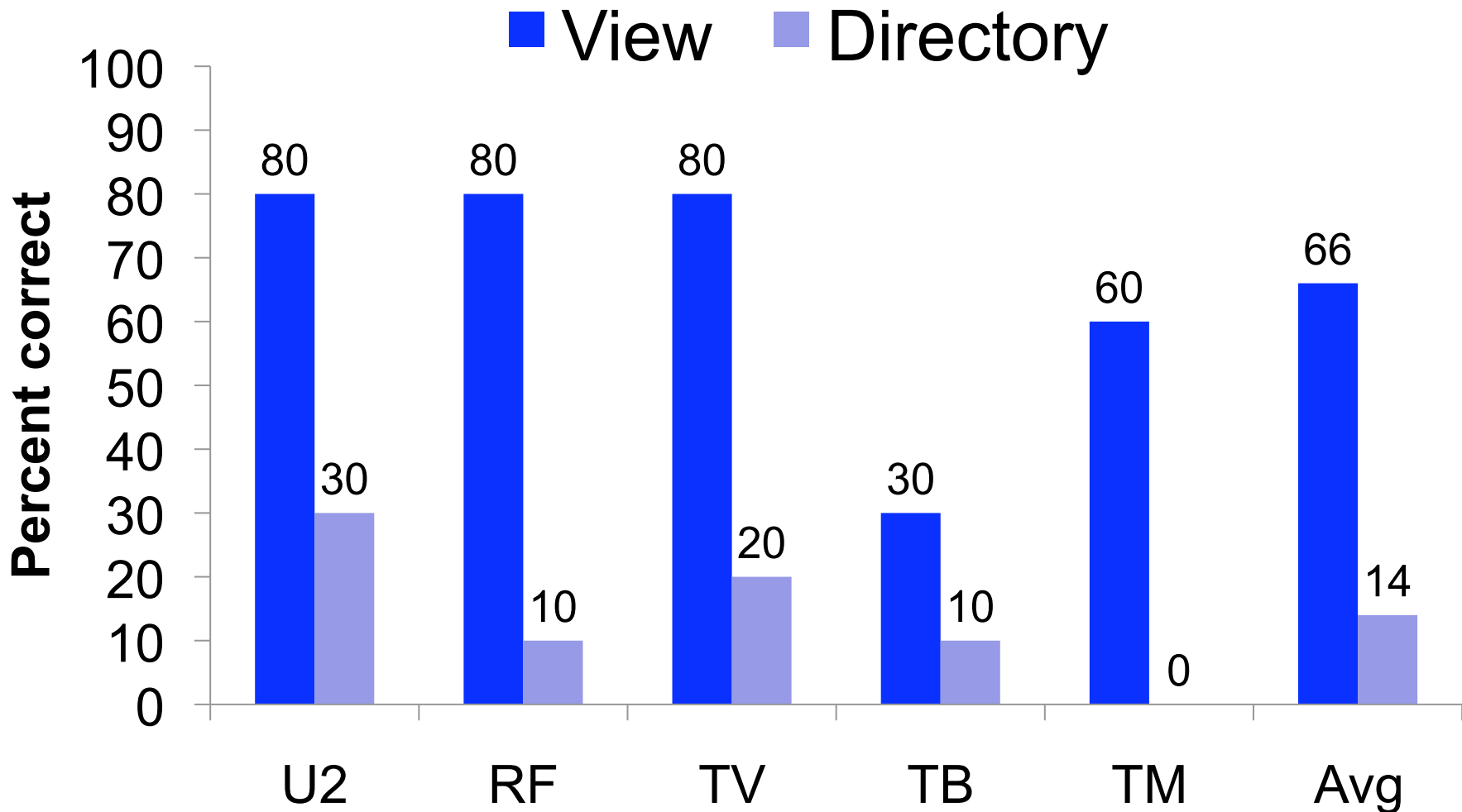
# Data organization tasks

---

- Test semantic/hierarchy mismatch
  - Applications are semantic
  - View-base system matches application naming
  - Directory / Volume systems are hierarchical
    - Require user to map from semantic to hierarchy
- Example task:

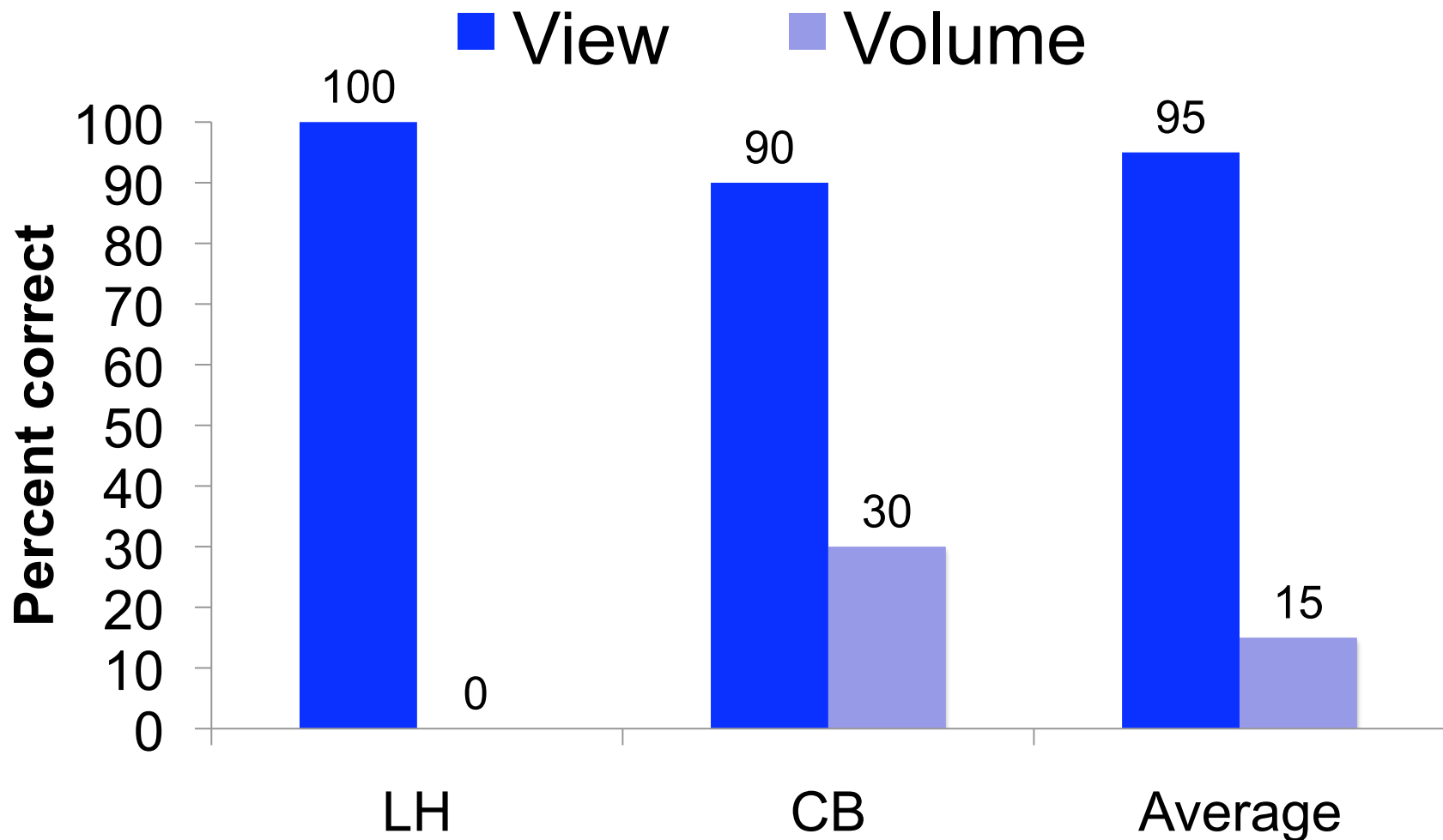
“Mary and Brian share music at home. However, when Mary is on trips, she finds that she can't listen to all the songs by U2 on her laptop. ...  
Make sure she can listen to all music by the artist U2 on her trips.”

# Data organization tasks



Users found semantic model simpler than hierarchical

# Client/server tasks



Users found peer-to-peer model easier to use

# Other results

---

- Tasks w/ clear but tedious mappings
  - Example: playlists, photo albums
  - Completion time varied, not accuracy
  - Views users 70s compared to 428s for directory
  
- Performance acceptable for day-to-day use
  - It has stored my data for 1 1/2 years
  - It also supports several multi-tuner DVRs
  - Latency acceptable, throughput ok for HDTV

# Conclusion

---

- We present analysis of home management
  - Found that environment was decentralized
  - Found disconnect in application and tool naming
- Perspective simplifies management tasks
  - Through semantic, decentralized management
  - User studies show up to 6x gain in accuracy