

# CLiMB

## Computational Linguistics for Metadata Building



Using Computational Linguistic  
Techniques to Harvest Image Descriptors

# The CLiMB Project

- What are the problems that CLiMB addresses?
- What are the potential CLiMB solutions?
- How did CLiMB come about?
- Who have been CLiMB participants?
- What are the next steps and goals for this meeting?

# Problems in Image Access

- Cataloging digital images

- Traditional approach:

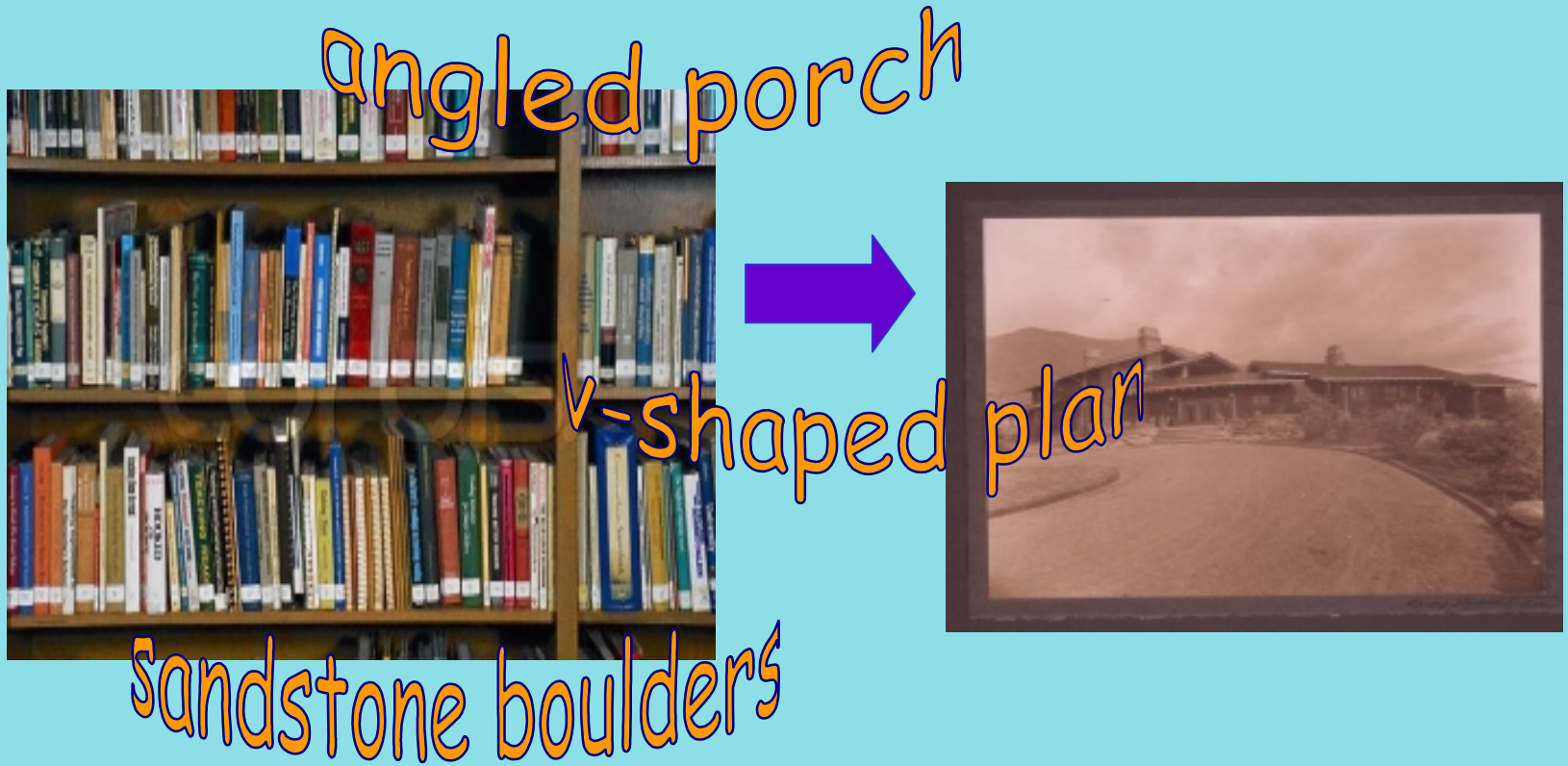
Manual expertise

- labor intensive
- expensive



- Can automated techniques help?

# Can we harvest image descriptors?



# CLiMB Technical Contribution



CLiMB will identify and extract

- proper nouns
- terms and phrases

from text related to an image:

By September 14, 1908, the basis of the **Greenes'** final design had been worked out. It featured a radically informal, **V-shaped plan** (that maintained the original **angled porch**) and interior volumes of various heights, all under a constantly changing **roofline** that echoed the rise and fall of the mountains behind it. The **chimneys** and **foundation** would be constructed of the **sandstone boulders** that comprised the local geology, and the **exterior** of the house would be sheathed in **stained split-redwood shakes.**

— Edward R. Bosley. *Greene & Greene*. London: Phaidon, 2000. p.127.

# CLiMB Overall Goals

The essence of CLiMB:

- Use scholars themselves as “catalogers” by utilizing scholarly publications
- Enhance existing descriptive metadata
- Help cataloger with subject access
- Improve retrieval of images

Outcomes:

- **Research:** Impact on retrieval through increased numbers of descriptors
- **Practice:** Suite of CLiMB tools
- **Resources:** Vocabulary list(s) that can be used by other visual resource professionals

# CLiMB History

- CLiMB-1 2002-2004 Columbia University
  - How did it start?
  - What were goals?
  - What were achievements?
  - Next steps
- CLiMB-2 – 2005 University of Maryland
  - Pursuing next steps

# How Did CLiMB Start?

- Within the context of interdisciplinary digital library research at Columbia
- Center for Research on Information Access
  - Mission – link Libraries with DL Technologies
- Search for a libraries-centric project
- Took 2 years to conceptualize



# CLiMB Then and Now

- Evolution over time
- Shift from use of
  - diffuse text to
  - more tightly coupled text and image and back to
  - more diffuse relationships between text and image
- Better understanding of criteria
- Clearer idea of possible techniques

# CLiMB-1 Project Team

Judith Klavans, PI

Stephen Davis

Angela Giral

Patricia Renfro

Bob Wolven

Roberta Blitz

Rebecca Passonneau

David Elson



# CLiMB-1 Achievements

- Created collection selection criteria
- Built and tested prototype CLiMB image cataloging toolkit
- Evaluated use of cataloging toolkit

# Key Open Issues

- Processing texts with more diffuse relation between text and image
- Refining CLiMB toolkit
- Utilizing existing vocabularies
- Incorporating into image search platform
- Testing with users, both catalogers and image searchers

# CLiMB-2 Project Team

Judith Klavans, co-PI

Marilyn Domas White, co-PI

Jimmy Lin, UMD

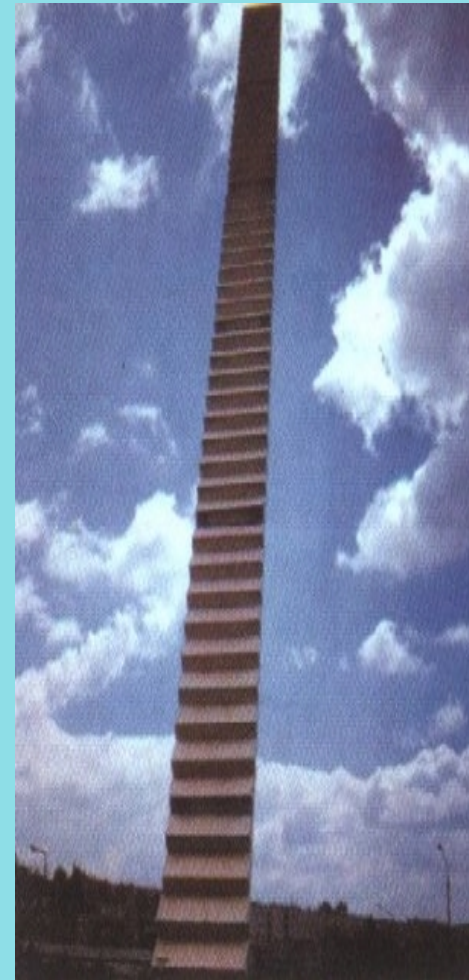
Eileen G. Abels, UMD

Dagobert Soergel, UMD

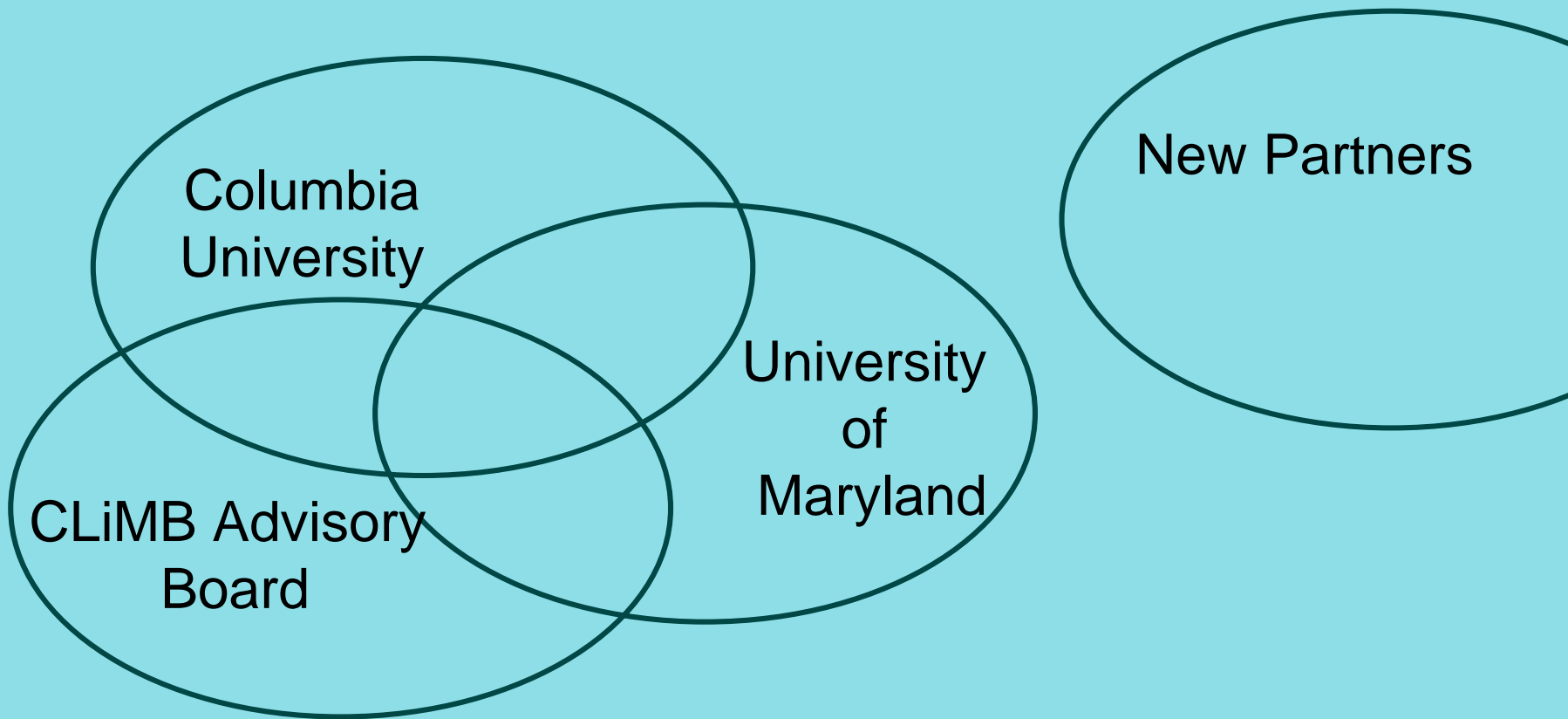
Brooke Rosenblatt, UMD

Angela Giral, CU (ret)

Rebecca Passonneau, CU



# Building CLiMB Partnerships



# Building Partners

- Collections
  - Museums
  - Libraries
- Users and user groups
  - Catalogers and end users of many types
- Tool builders for digital libraries
- Evaluation, evaluation, evaluation

# Next Steps

- Initial project – Ended June 2004
- Planning grant at UMD – January 2005
  - College of Information Studies (CLIS)
- Schedule for 2005:
  - March – Partner meeting
  - April – Coalition for Networked Information (CNI)
  - Submit prospectus May 2005



[www.umiacs.umd.edu/~climb](http://www.umiacs.umd.edu/~climb)

[www.columbia.edu/cu/cria/climb](http://www.columbia.edu/cu/cria/climb)

# Meeting Schedule



- CLiMB Overview and Problem  
Judith Klavans
- CLiMB Image Cataloging Toolkit  
Rebecca Passonneau and Jimmy Lin
- Effective Collection Selection  
Angela Giral
- Evaluation and Users  
Rebecca Passonneau
- Research in CLiMB-2  
Marilyn Domas White
- Partners in CLiMB-2  
Judith Klavans