



Generating Recommendations in OPACs

Initial Results and Open Areas for Exploration

Colleen Whitney

California Digital Library
and
UC Berkeley School of Information



Melvyl Recommender Project

Exploratory project, probing several ideas:

- _ Text-based discovery system: eXtensible Text Framework (XTF)
- _ Enhanced relevance ranking
- _ Spelling correction
- _ User interface strategies
- _ **Recommending**



The Testbed

- _ Bibliographic records
 - ~4.2 million UCLA records
 - MARC export from Melvyl
- _ Circulation data
 - ~9 million UCLA circulation transactions
 - September 1999 – May 2005
 - Data from two systems: Taos, Voyager



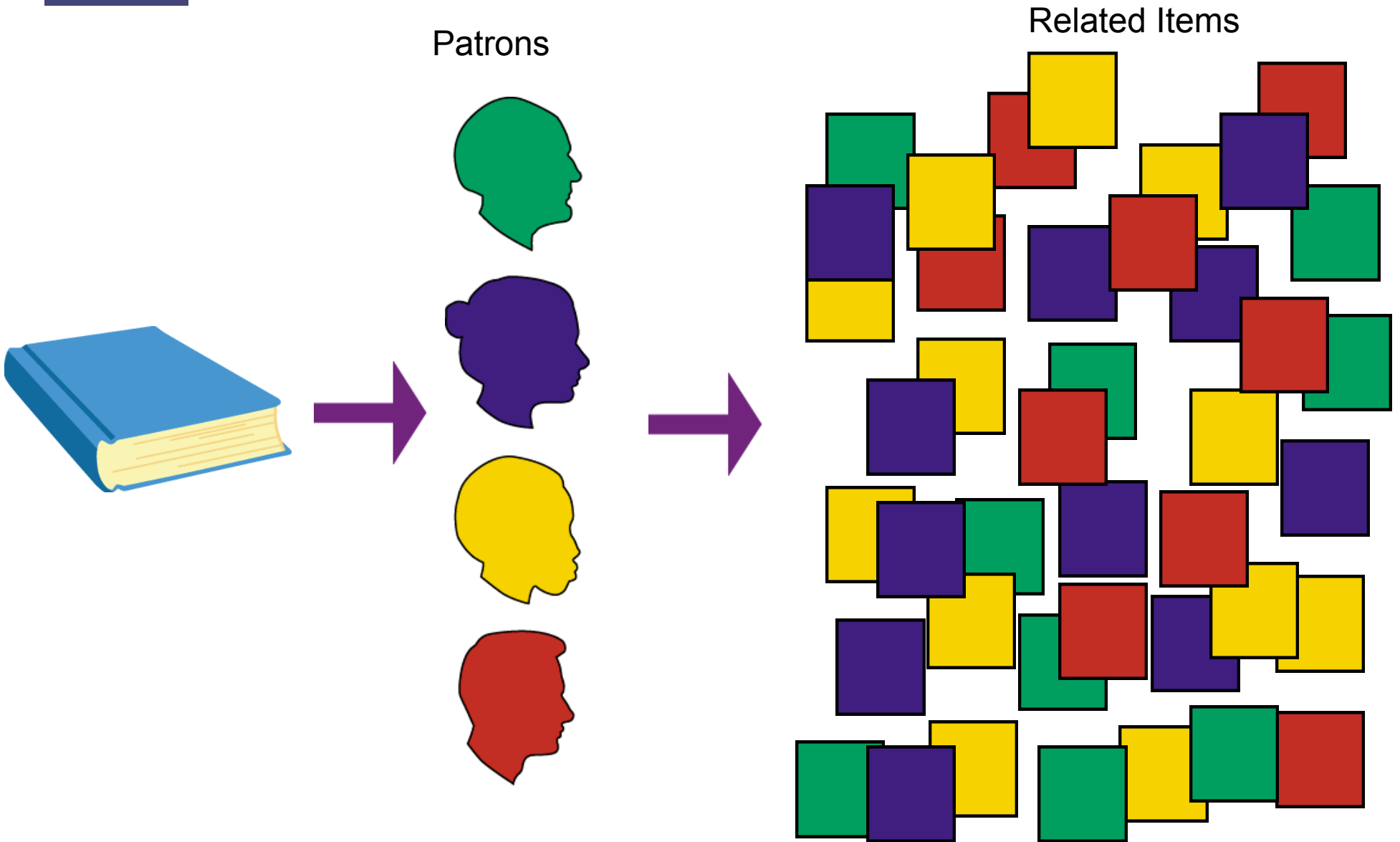
The Basic Idea

- “Patrons who checked this out, also checked out....”
- != Collaborative Filtering





Finding Relationships





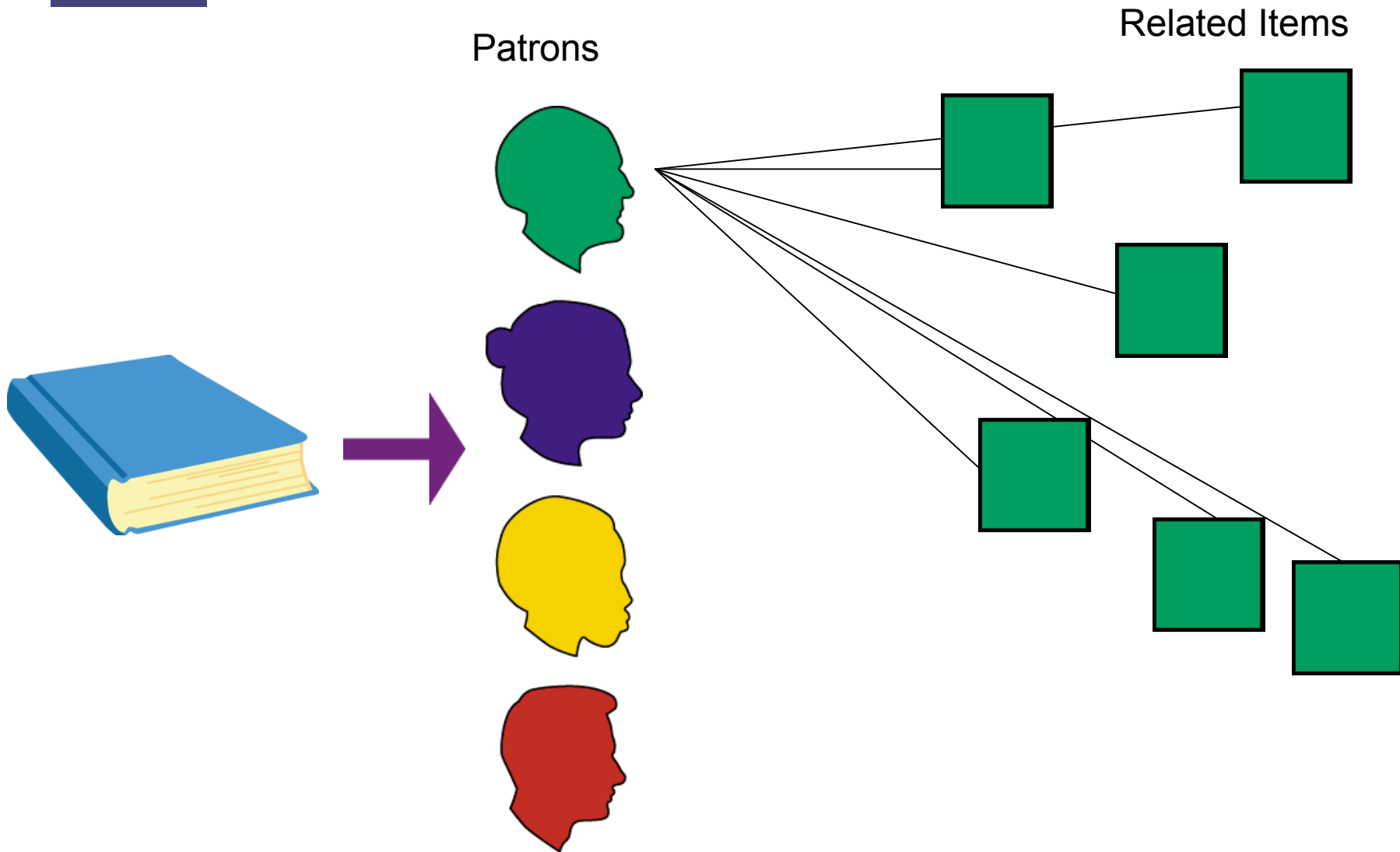
Creating Sets

Patrons



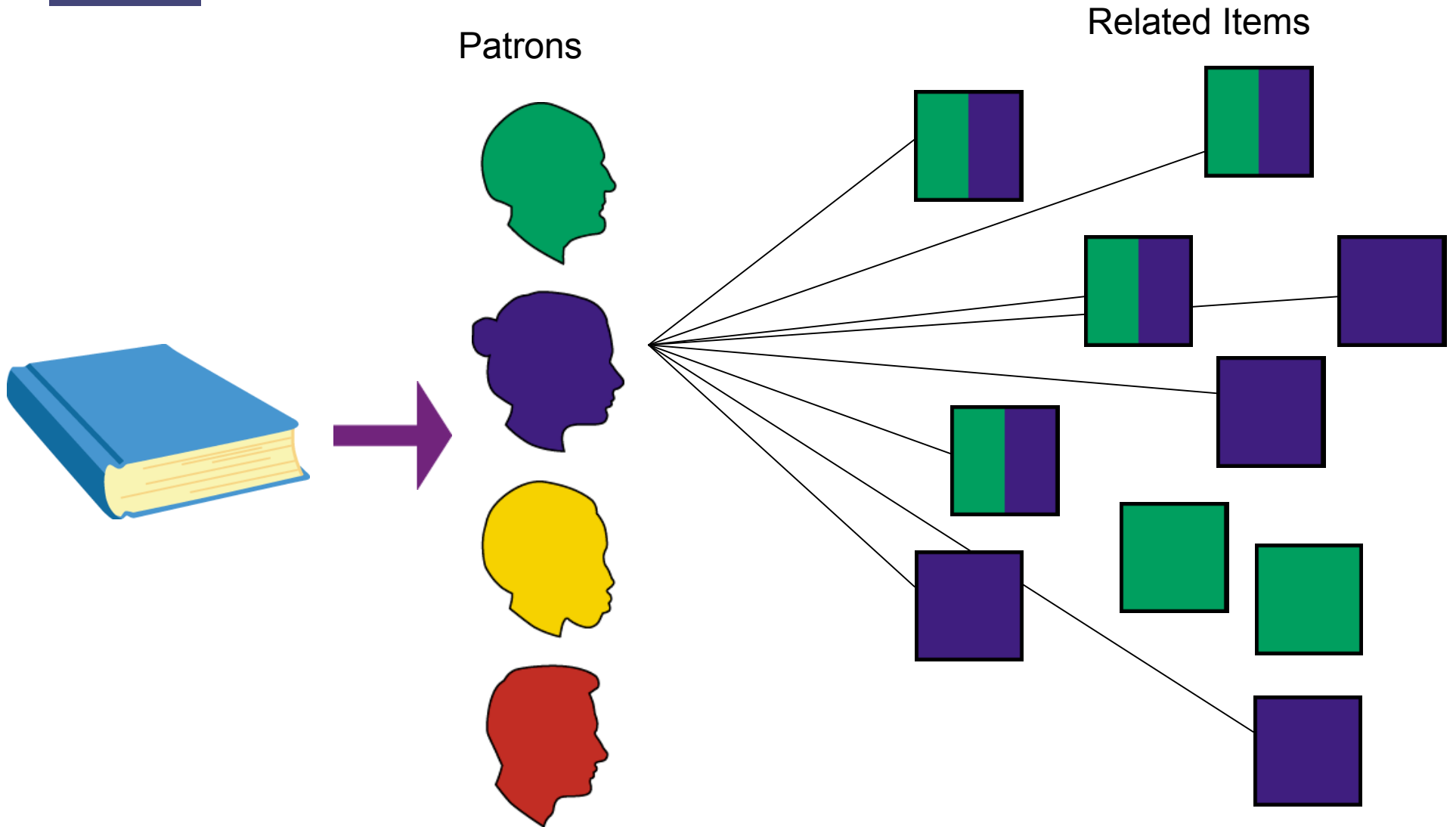


Creating Sets



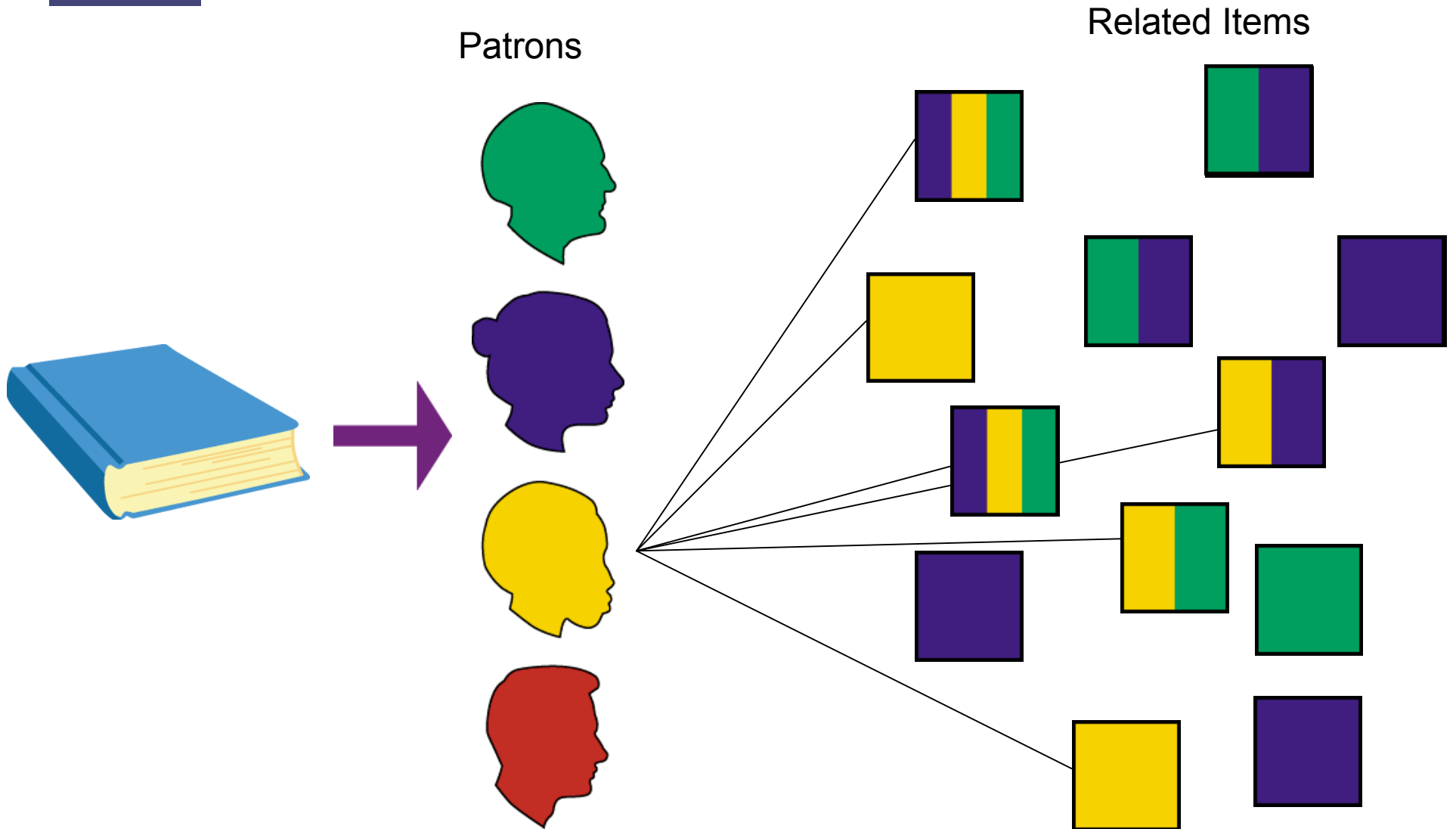


Creating Sets



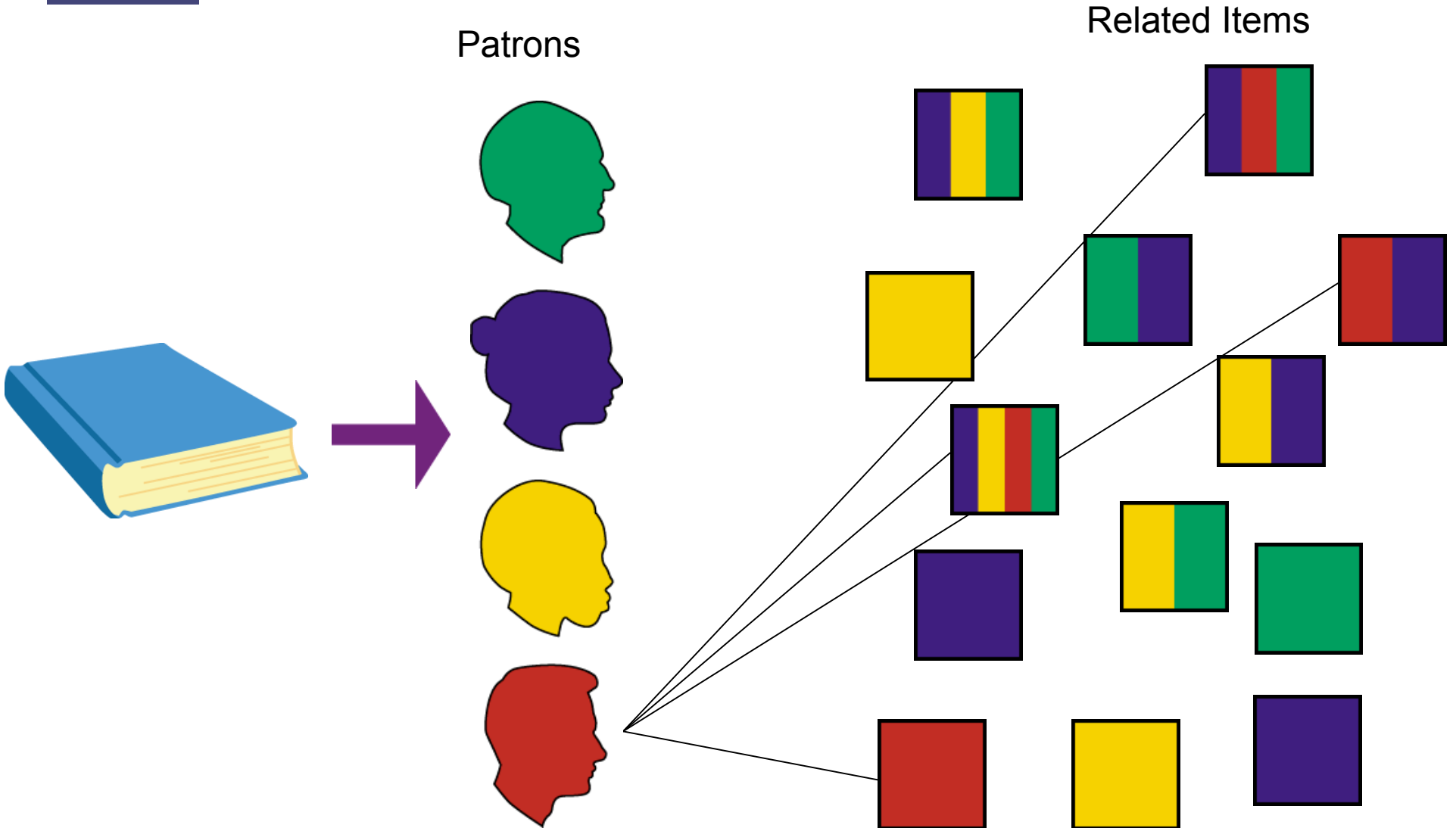


Creating Sets



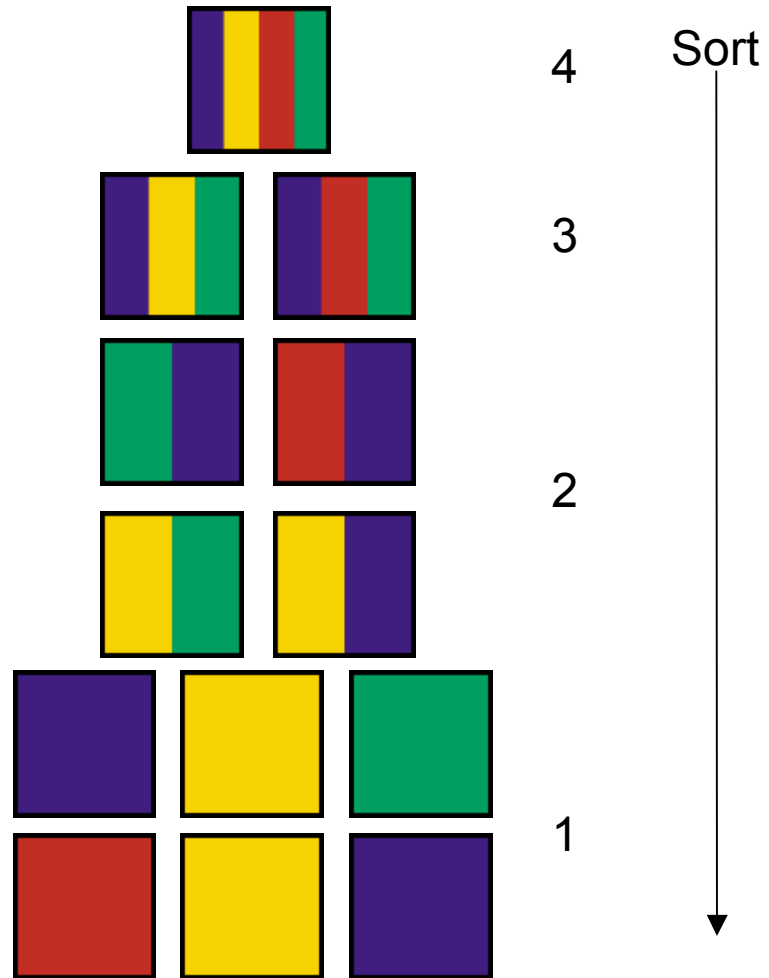


Creating Sets



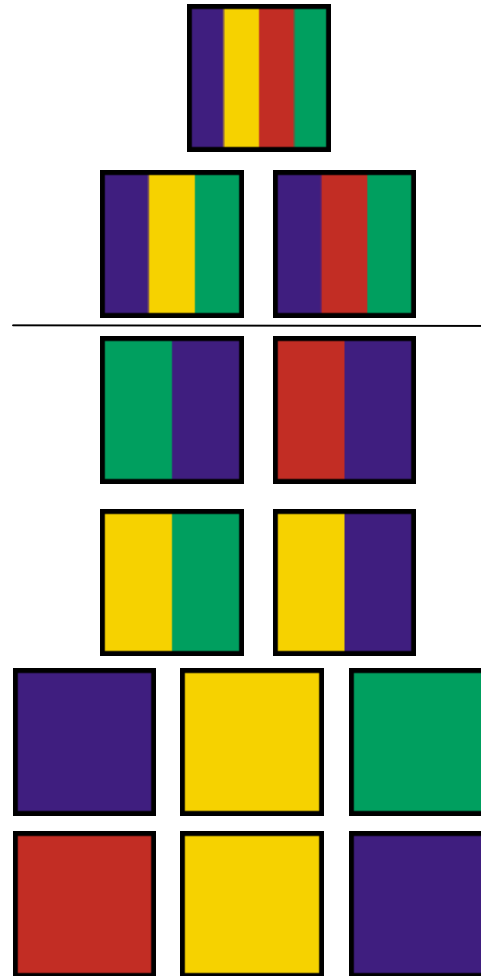


Creating Sets



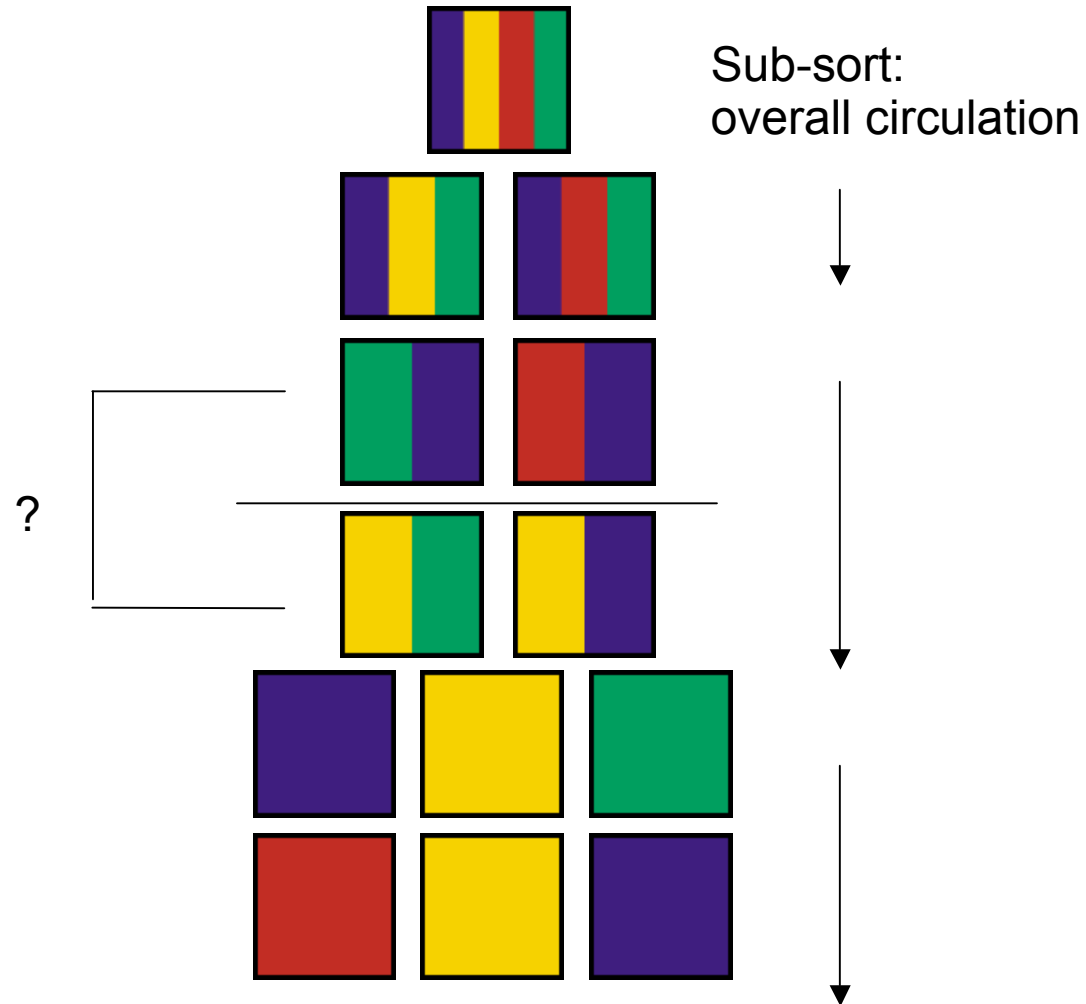


Creating Sets





Creating Sets





Examples (v. 1)

Title: Celebrating a collection: the work of Dorothea Lange

Author (Format, Date): Heyman, Therese Thau. (Book, 1978)

Language: English

Subjects: Lange, Dorothea | Photography, Artistic -- Exhibitions | Photographers -- United States -- Biography

1. The critical vision: a history of social and political art in the U.S
2. The great transformation
3. The face on the cutting room floor: the story of movie and television censor
4. The art of social conscience
5. Iron cages: race and culture in 19th-century America



Examples (v. 1)

Title: As you like it

Author (Format, Date): Smallwood, R. L. (Robert Leo) (Book, 2003)

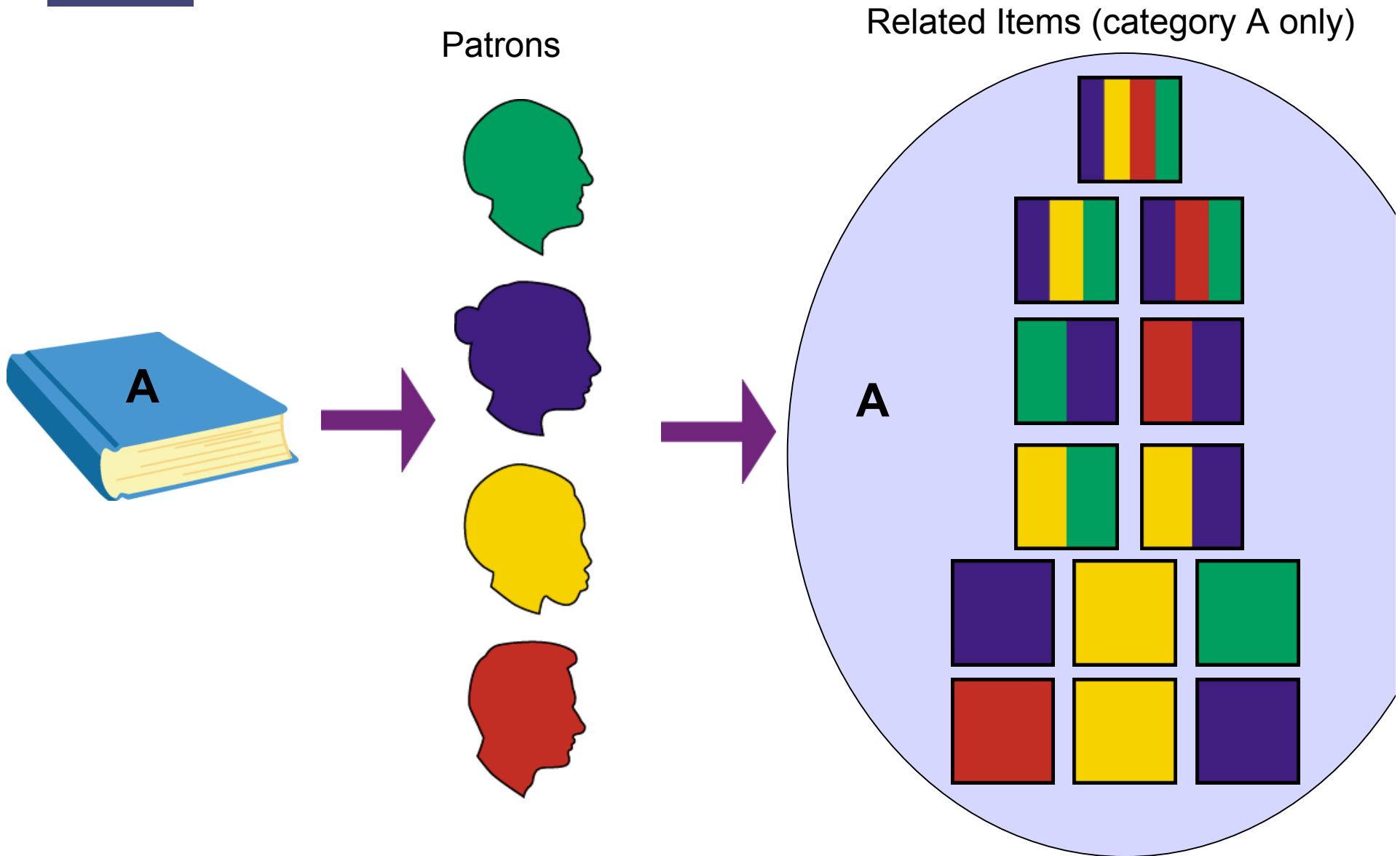
Language: English

Subjects: Shakespeare, William, -- 1564-1616. -- As you like it |
Shakespeare, William, -- 1564-1616 -- Dramatic production |
Shakespeare, William, -- 1564-1616 -- Stage history -- 1950-

1. Calculus
2. Sexual perversity in Chicago; Sexual perversity in Chicago and
The duck variations: two plays
3. Plays; Beth Henley: collected plays
4. Don Quixote. English; The ingenious hidalgo Don Quixote de la
Mancha
5. A decade of new comedy: plays from the Humana Festival



Filtering





Examples (v. 2)

Title: Celebrating a collection: the work of Dorothea Lange

Author (Format, Date): Heyman, Therese Thau. (Book, 1978)

Language: English

Subjects: Lange, Dorothea | Photography, Artistic -- Exhibitions | Photographers -- United States -- Biography

1. Photojournalism
2. Dorothea Lange: a photographer's life
3. In this proud land: America 1935-1943 as seen in the FSA photographs
4. Truth needs no ally: inside photojournalism
5. The Royal Society of Arts, 1754-1954



Examples (v. 2)

Title: As you like it

Author (Format, Date): Smallwood, R. L. (Robert Leo) (Book, 2003)

Language: English

Subjects: Shakespeare, William, -- 1564-1616. -- As you like it |
Shakespeare, William, -- 1564-1616 -- Dramatic production |
Shakespeare, William, -- 1564-1616 -- Stage history -- 1950-

1. Sexual perversity in Chicago; Sexual perversity in Chicago and The duck variations: two plays
2. Plays; Beth Henley: collected plays
3. Don Quixote. English; The ingenious hidalgo Don Quixote de la Mancha
4. A decade of new comedy: plays from the Humana Festival
5. Barefoot in the park: a new comedy



User Assessment

- _ April 2006
- _ Are the recommendations generated using circulation data appropriate for an academic audience?
- _ Stay tuned for results!



The Underbelly

- _ User queries XTF
- _ AJAX request to Perl/mySQL
- _ Perl returns XML response
- _ Display updated in browser if recommendations are present



What works?

- _ Overall quality of recommendations seems good, especially in the humanities (but subject to user testing).
- _ Severely off-topic recommendations are limited by the content filter.
- _ AJAX allows for smooth presentation.



What doesn't work?

- _ Serendipity is limited by content filter.
- _ The content filter is crude, eliminating some good recommendations.
- _ Quality and quantity of recommendations are lower in science and technology.
- _ Missing/bad call number class data eliminates recommendations for some items.



Ideas for Further Development

- _ Refine this algorithm:
 - Improve content filter (underway)
 - Experiment with “Binning” (underway)
 - Backfill missing data



Ideas for Further Development

- _ Experiment with other methods:
 - “More like this...” similarity-based recommendations (underway)
 - Same author
 - Same subject heading(s)
 - Connections identified using other external data sources....citations, reading lists...



Many thanks to...

- _ Mellon Foundation (funding)
- _ RLG (holdings data for ranking experiments)
- _ OCLC (holdings data for ranking experiments)
- _ UCLA Library (circulation data)
- _ UC Berkeley Library (circulation data)
- _ CDL Team and Contributors
(Including Peter Brantley, Lynne Cameron, Rebecca Doherty, Randy Lai, Jane Lee, Martin Haye, Erik Hetzner, Kirk Hastings, Patricia Martin, Felicia Poe, Lisa Schiff, Roy Tennant, Brian Tingle, Steve Toub...and many more.)



Links:

Project Home:

http://www.cdlib.org/inside/projects/melvyl_recommender/

Prototype:

<http://recommend-dev.cdlib.org/melrec/>