

Collaborative Information Retrieval

Chirag Shah

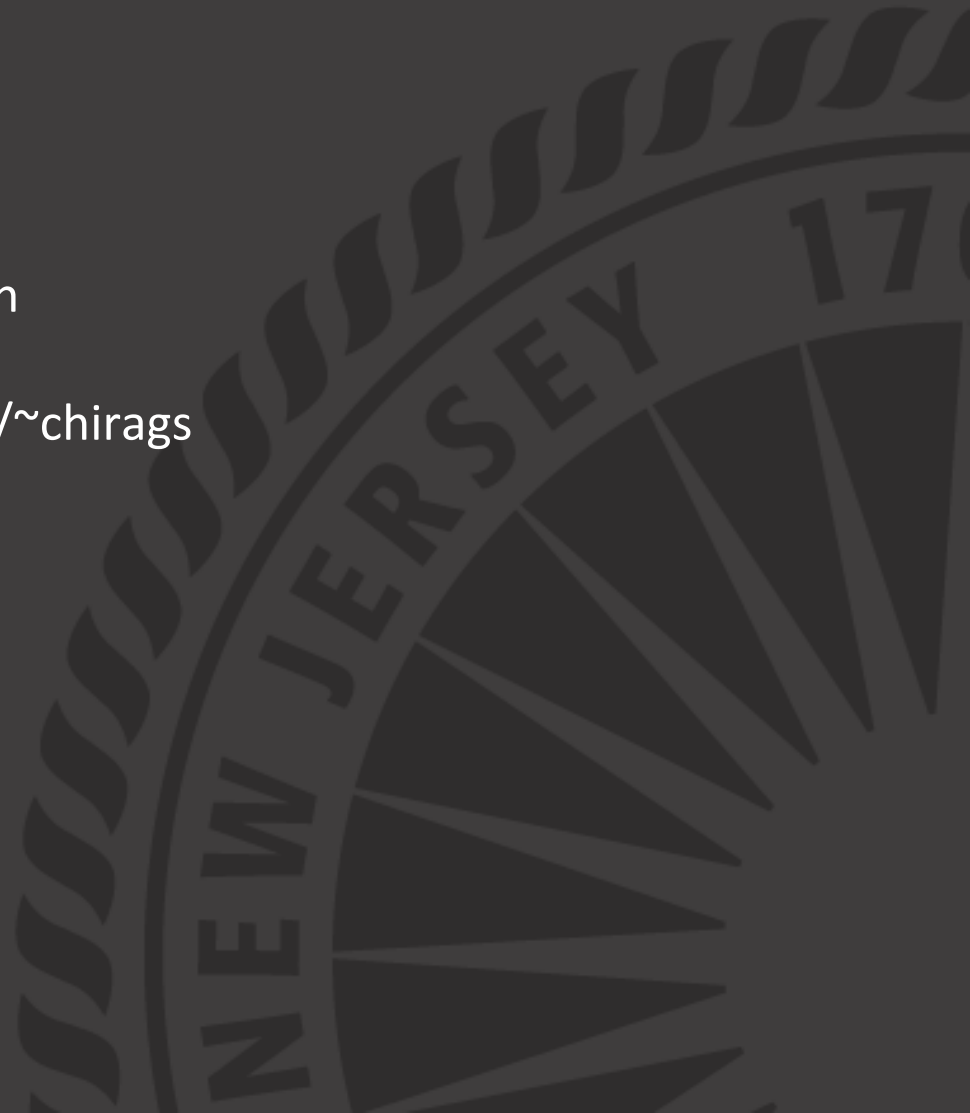
Twitter: #RuSSIR #cir #cis @chirag_shah

Email: chirags@rutgers.edu

Website: <http://comminfo.rutgers.edu/~chirags>

RUTGERS

School of Communication
and Information



Course Objectives

- Define **collaborative IR** and differentiate it with related terms.
- Identify **situations** and **motivations** for collaborative IR.
- List various **dimensions** of collaborative work/systems/method.
- Discuss **system-mediated** and **user-mediated** CIR with examples.
- Enumerate essential and desired **features** of CIR systems.
- Outline an **evaluation** framework for CIR.

Outline of the Course

Day-1

- Collaborative IR – what, why, how?
- Taxonomy of CIR situations and systems
- Algorithmic collaboration

Day-2

- User-centered collaboration
- Evaluating CIR systems
- Directions for further research and development

Day 1

Collaboration – What and Why



Requirement or setup
e.g., merger



Division of labor
e.g., class assignment



Diversity of skills
e.g., co-authorship

Background

- Dumais, Fidel, Grudin, Bruce, Pejtersen, Poltrock
(Microsoft, UWash, Boeing – Collaborative IR)
- Morris, Horvitz, Teevan (Microsoft – SearchTogether, Co-search)
- Twidale, Nichols (UIUC – Collaboration in libraries, Ariadne)
- Reddy, Jensen (PSU – Collaborative info behavior in health)
- Golovchinsky, Pickens (FXPAL – Algorithmic mediation for CES)
- Foley, Smeaton (DCU – Synchronous CIR)
- Foster (Sheffield – Collaborative info behavior)
- Hansen, Jarvelin (SICS – CIR in work environments)
- Shah (Rutgers – CIS in everyday problem-solving)

Background

- Workshops by Golovchinsky, Morris, and Pickens:
 - JCDL 2008
 - CSCW 2010
 - CIKM 2011
- Workshop by Reddy, Jensen, and Twidale:
 - Group 2009
- Workshop by Shah, Reddy, and Twidale:
 - Group 2010
- Workshop by Shah, Hansen, and Capra:
 - ASIST 2011

Terms and Terminology

- Collaborative search
- Social search
- Collaborative IR
- Collaborative exploratory search
- Collaborative information seeking
- Collaborative information behavior
- Co-browsing
- Collaborative navigation
- Collaborative information synthesis
- Collaborative sense-making

Our Definition

Information seeking/retrieval process that is

- **explicitly** defined among participants,
- **intentional**,
- **interactive**, and
- **mutually beneficial**.

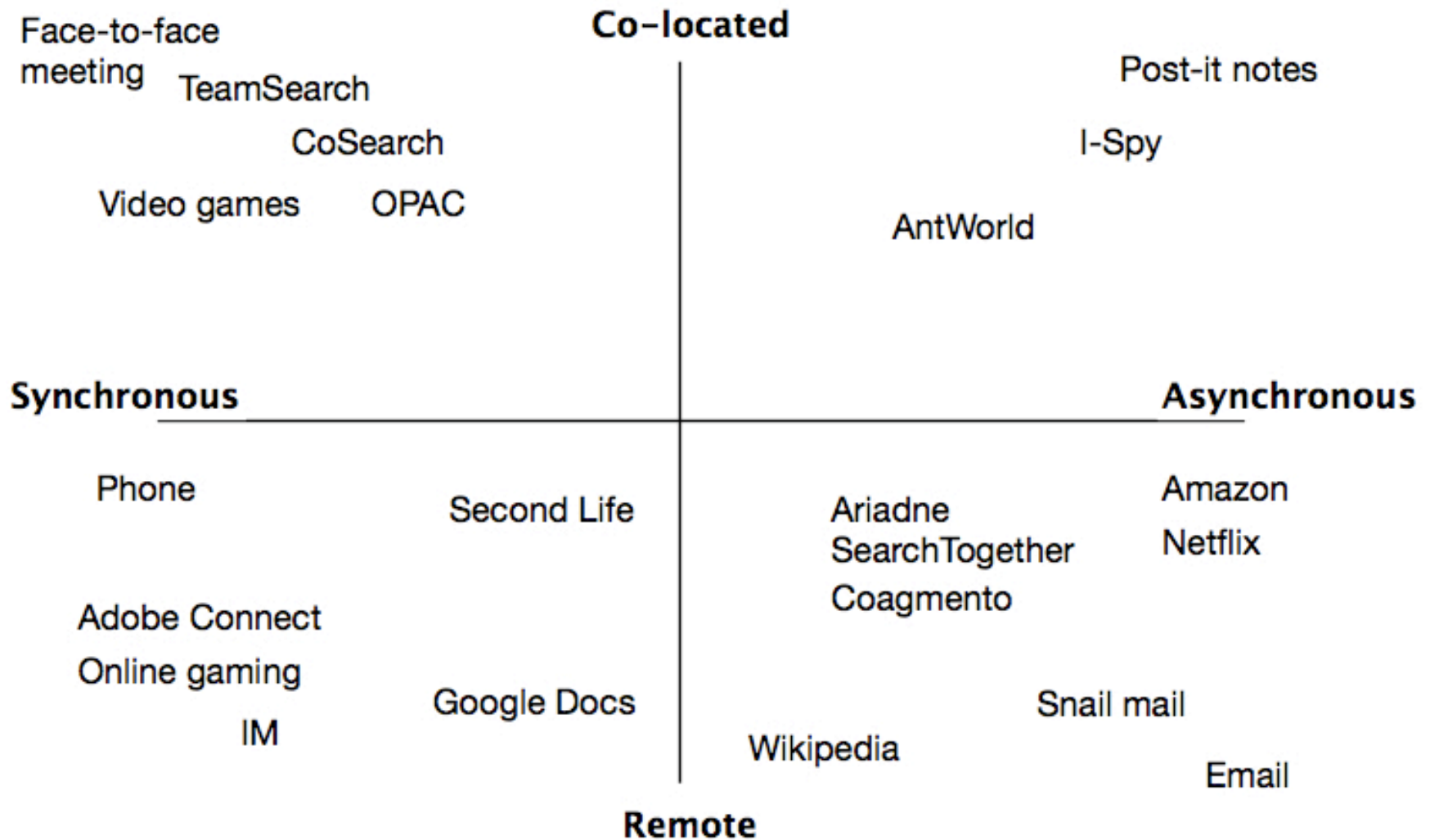
Collaborative IR \neq Collaborative filtering

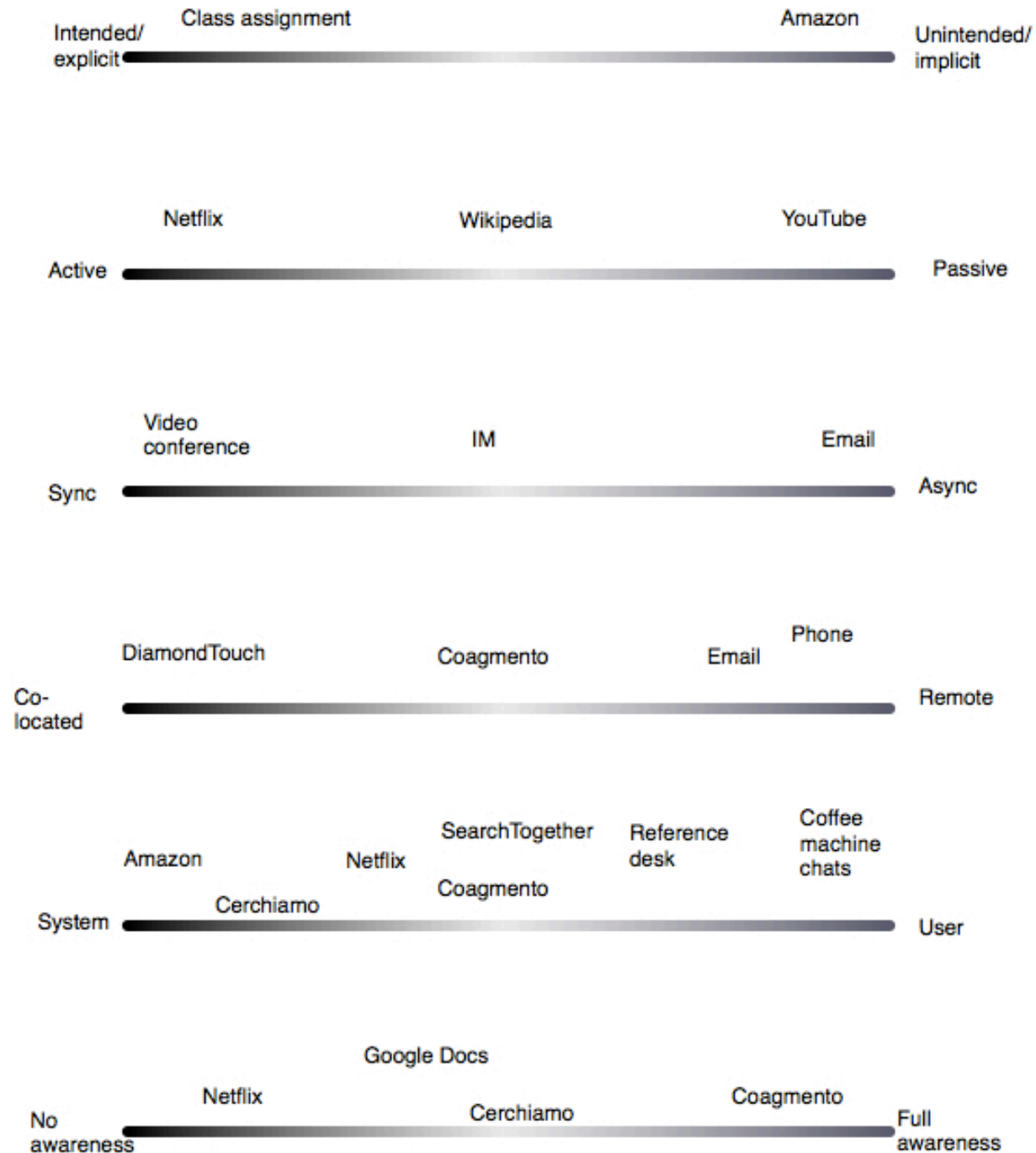
Class Exercise

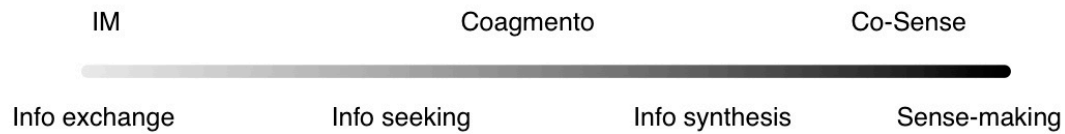
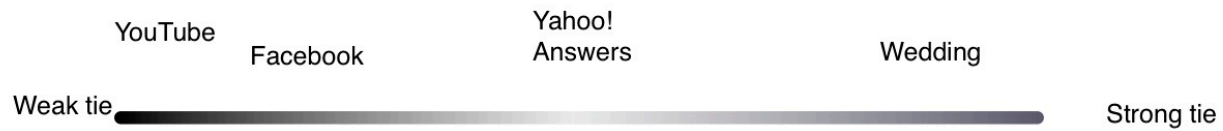
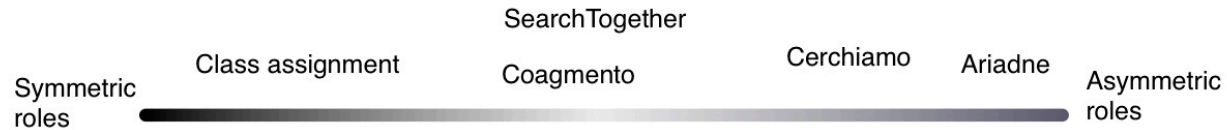
Examples of CIR

- Co-authorship
- Engaged couple doing wedding planning
- Vacation planning
- ???

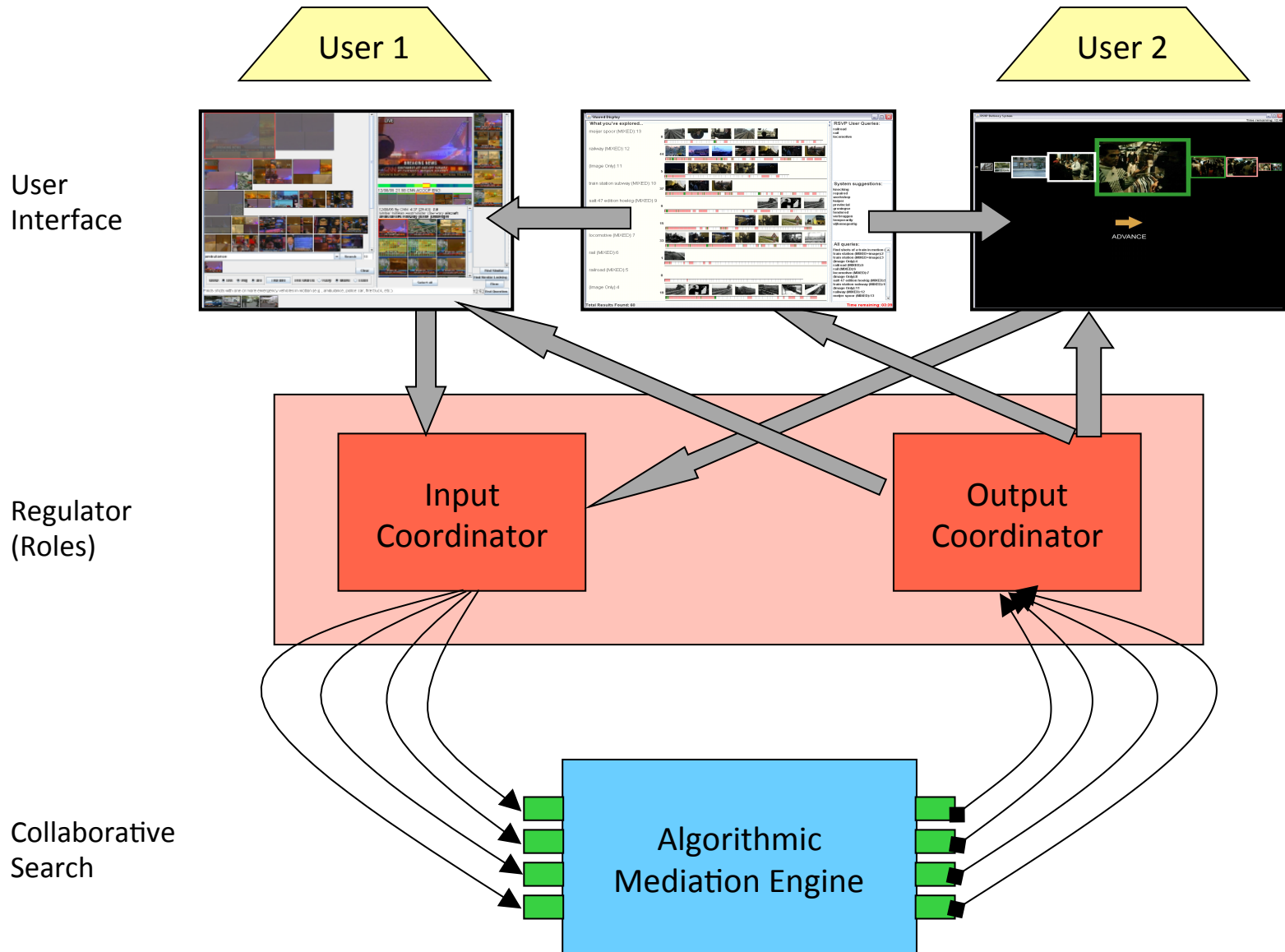
Group Activities in Time & Space

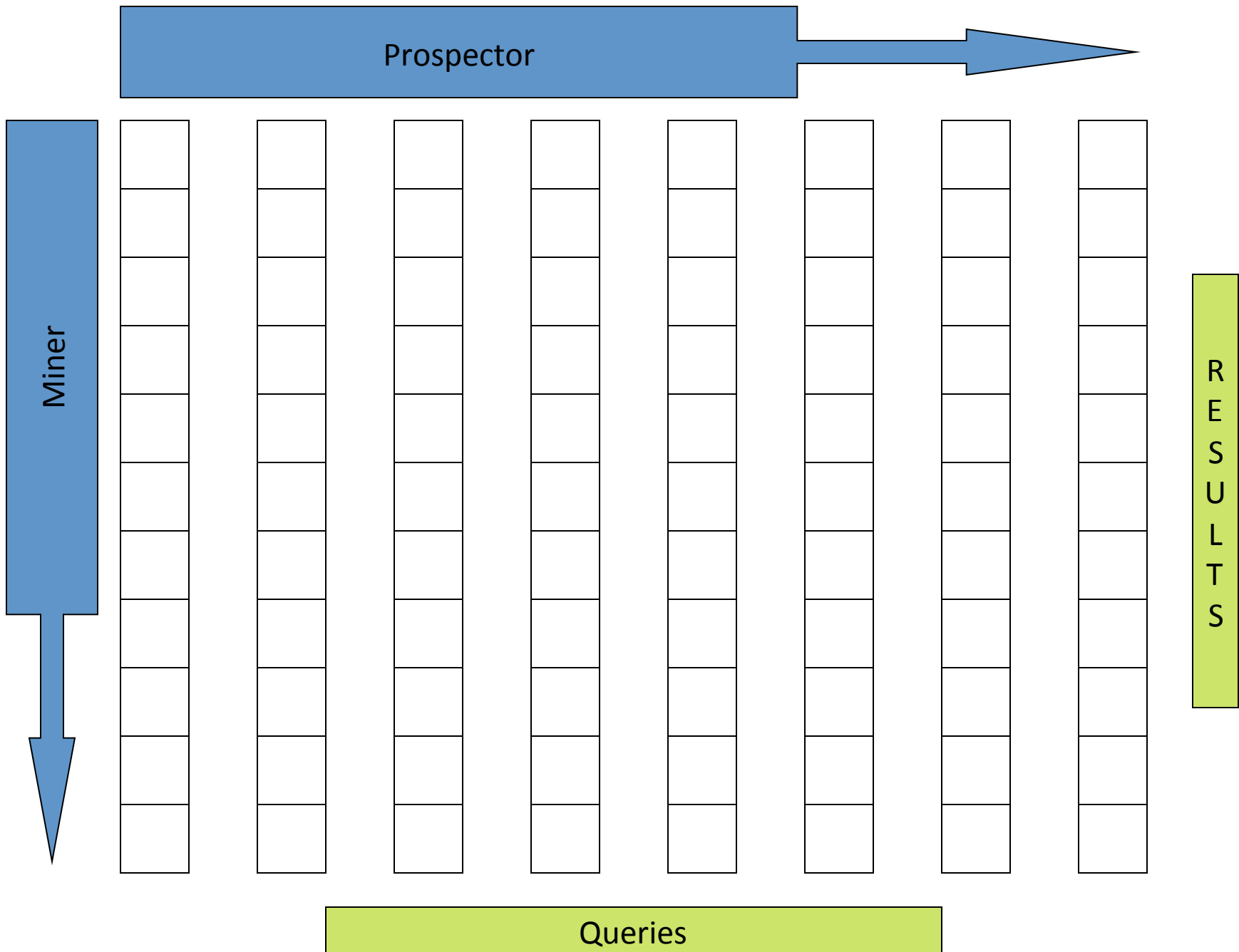


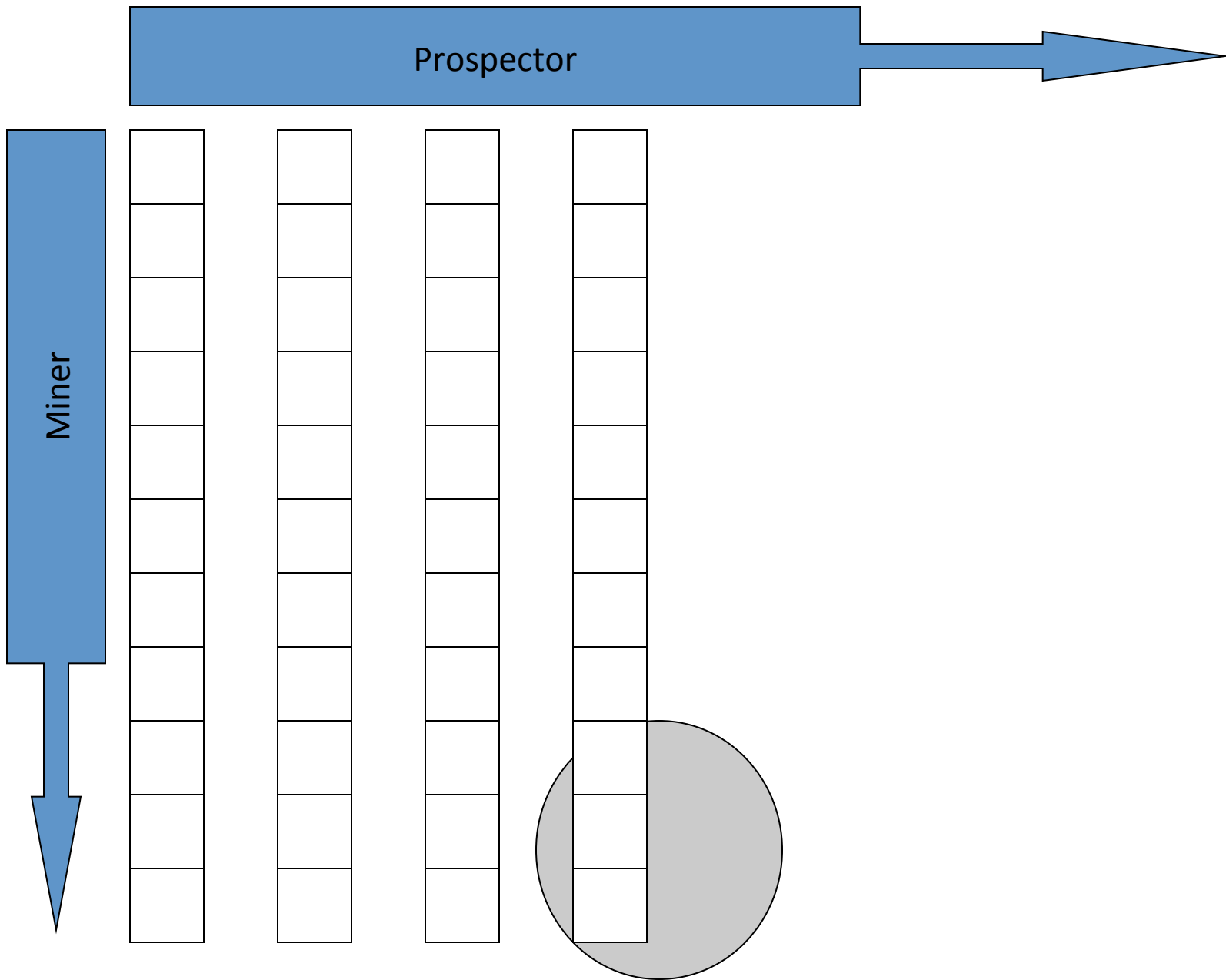


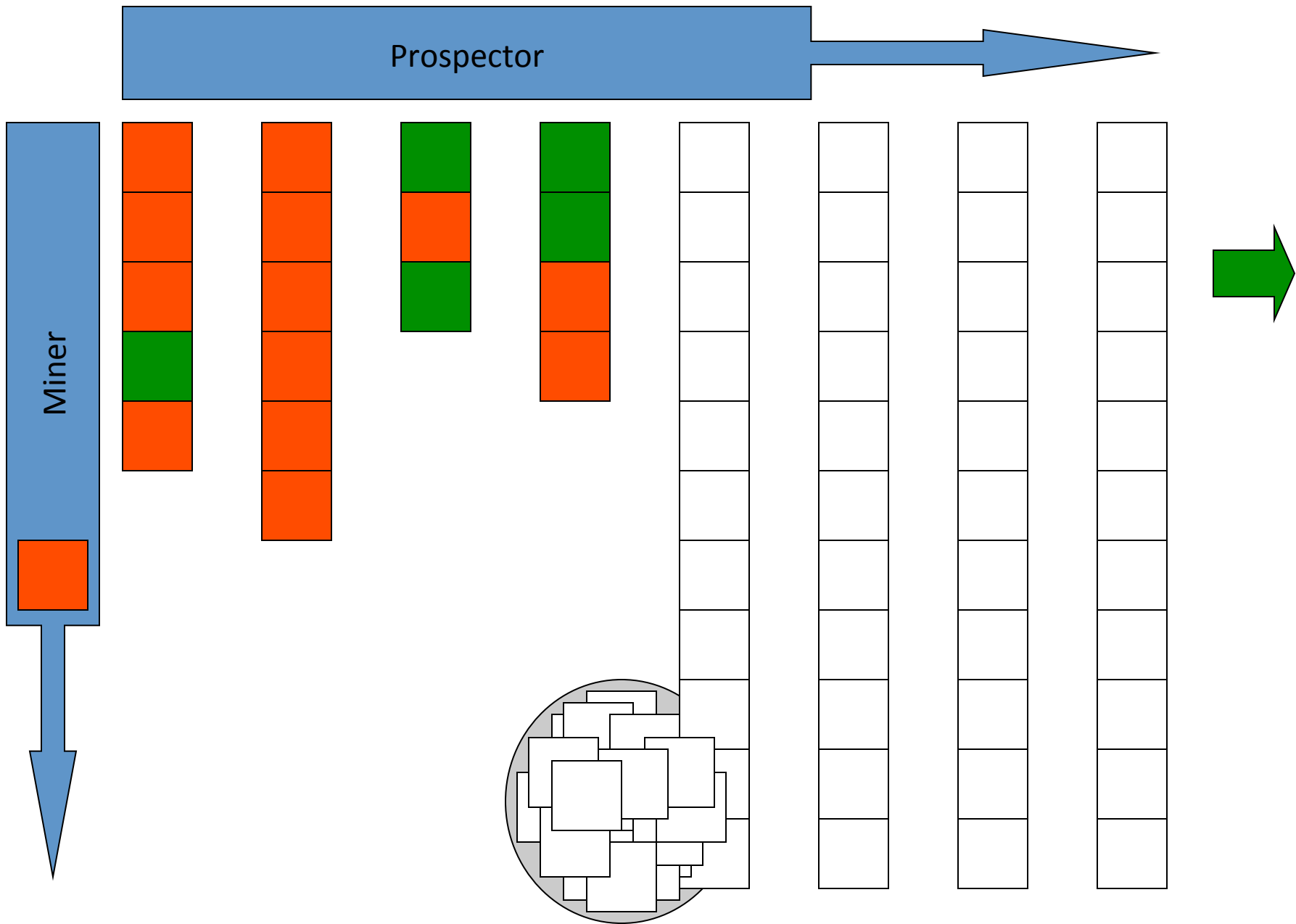


System-mediated Asymmetric-role Collaboration









Miner Queue Priority

Weighted Borda Count fusion

$$priority_{doc} = \sum_q Borda_{doc,q} \cdot w_{seen,q} \cdot w_{rel,q}$$

$$w_{seen,q} = N_{seen,q} / N_{unseen,q}$$

$$w_{rel,q} = N_{rel,q} / N_{nonrel,q}$$

Freshness

Relevance

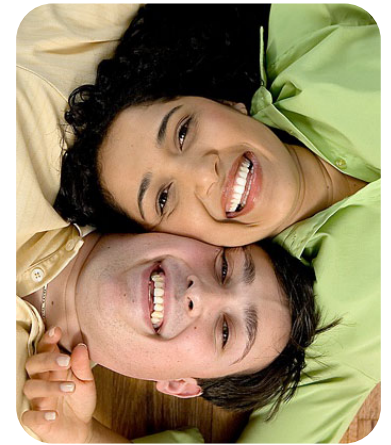
Homework

Think of different kind of roles in collaborative IR.
Come up with a thought/sketch/algorithm to
address them.

Twitter: #RuSSIR #cir #cis @chirag_shah
Email: chirags@rutgers.edu

Day 1 Conclusion

- Collaboratively seeking/retrieving/using information
- Situations and motivations
- Dimensions
- Algorithmic or system mediated collaborative IR



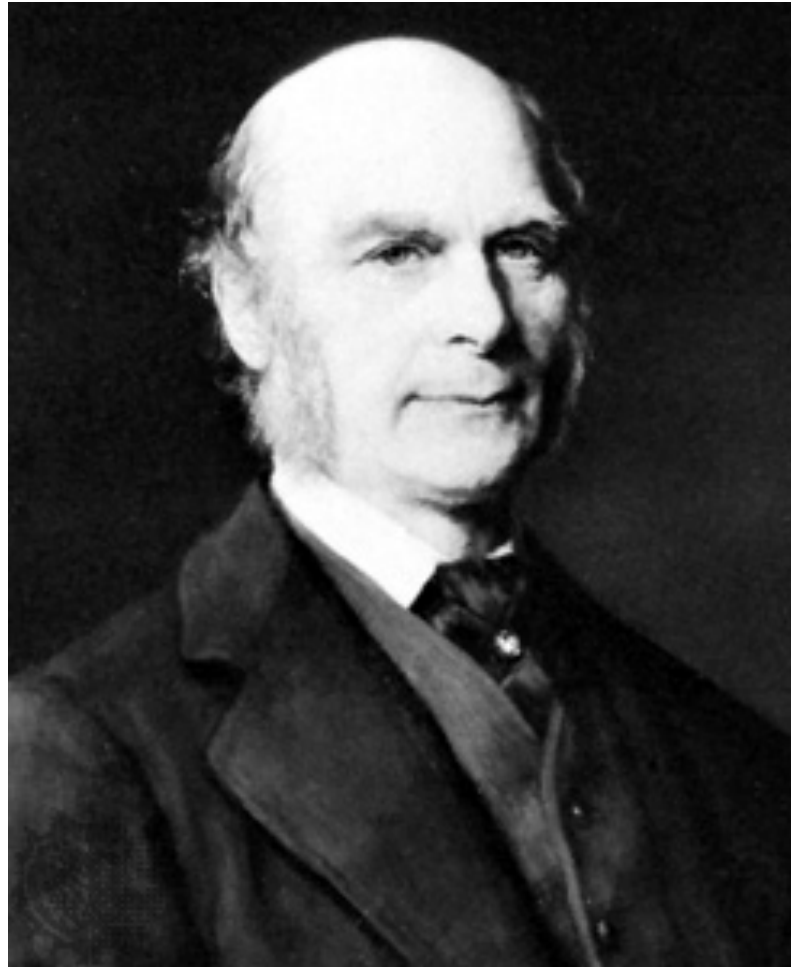
Day 2

Twitter: #RuSSIR #cir #cis @chirag_shah

Email: chirags@rutgers.edu

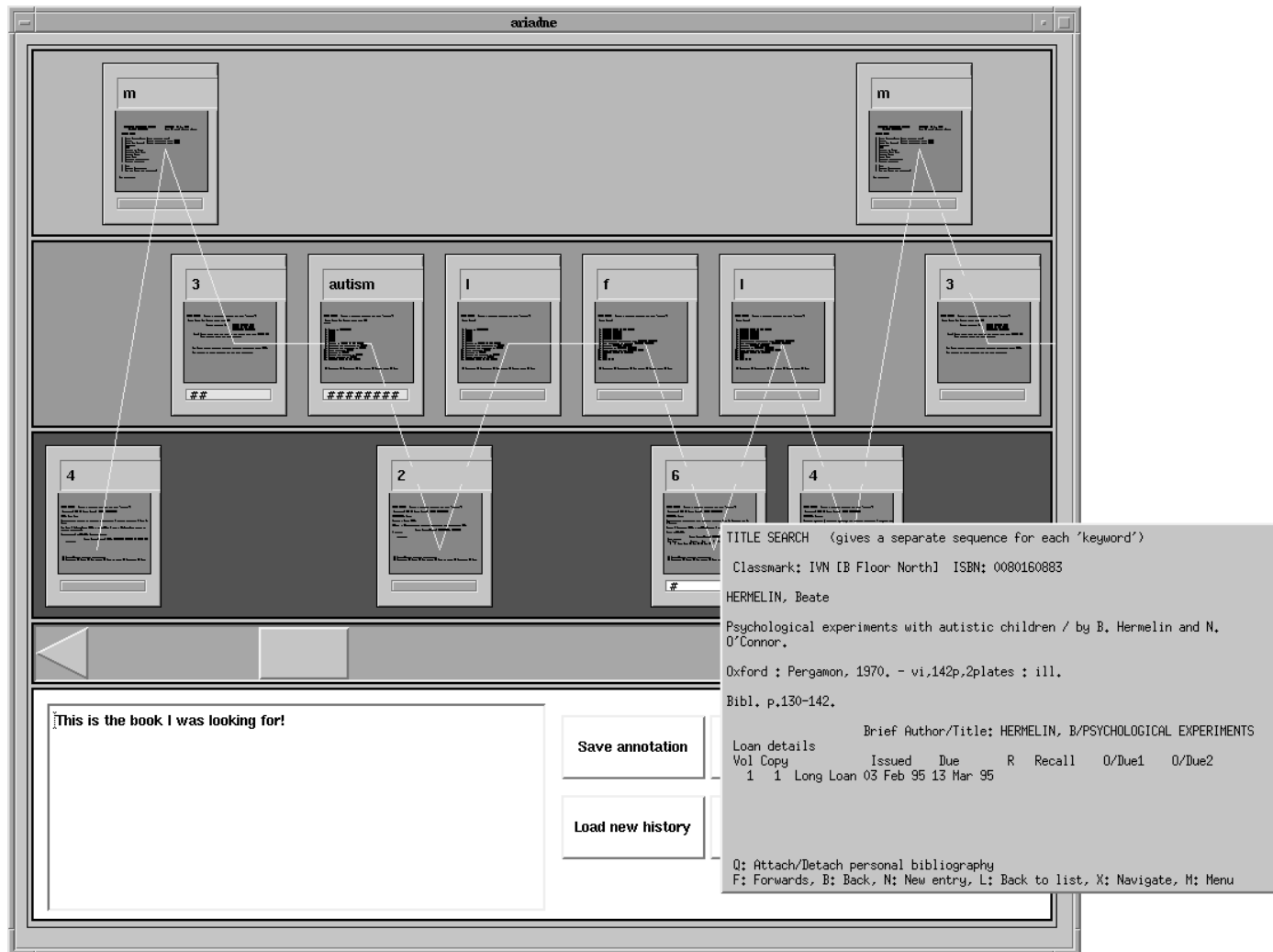
Website: <http://comminfo.rutgers.edu/~chirags>

Galton and the Wisdom of Crowd



Sir Francis Galton (1822–1911)

Ariadne



SearchTogether

The screenshot displays the SearchTogether diabetes (george) interface. On the left, a sidebar shows a list of users: george, rachel, and betty. The main content area displays search results for "diabetes friendly recipes". The results include links to "VGS - Recipes - American Diabetes Association", "Diabetic Recipes", "Diabetes Self Management Recipes", "Children's Diabetes Foundation - Recipes", "Diabetes-Friendly Holiday Recipes", "Diabetes Friendly Holiday Cookie Recipes", "AARP - Segunda Juventud - Diabetes-Friendly Recipes", and "AARP - Segunda Juventud - Diabetes-Friendly Recipes". The right side of the interface shows a "Reader's Digest" article titled "Ho-Ho-Holiday Help! Diabetes-Friendly Holiday Recipes". The article includes a photo of a dish and text about preparing a satisfying holiday meal. The interface also features a "Current Results" section with "Standard Search", "Split Search", and "Multi-Engine Search" buttons. A "Recommend" button is visible at the bottom right. Red annotations with letters a through j are placed around the interface: 'a' points to the search bar, 'b' points to the user list, 'c' points to the search results, 'd' points to the recommendations section, 'e' points to the "Standard Search" button, 'f' points to the "Split Search" button, 'g' points to the "Multi-Engine Search" button, 'h' points to the "Back" button, 'i' points to the "Forward" button, and 'j' points to the "Recommend" button.

Coagmento

The screenshot shows the Coagmento web application in a browser window. The browser's address bar displays <http://www.coagmento.org/CSpace/>. The page title is "Coagmento - Collaborative Information Seeking, Synthesis, and Sense-making". The browser's toolbar includes buttons for "CSpace", "Activate", "Bookmark", "Recommend", "Annotate", "Snip", "Views: 0", "Annotations: 0", "Snippets: 0", and "Project: carrots and sticks". A search bar is located in the top right corner.

The main content area is titled "Coagmento" and features a sidebar on the left with the following sections:

- Collaborators**: Add and manipulate collaborators for your projects. Includes a link to "Add someone" as a collaborator and a list of collaborators for the current project: "Daffy Duck" and "Bugs Bunny". A link "See all your collaborators" is also present.
- Projects**: Add and manipulate your projects.
- Data and Information**: Explore data and information about you and your collaborators.
- Workspace**: Explore your collected information and produce results using the workspace.
- Recommendations**: See the recommendations from your collaborators or the system.
- Settings**: Update your profile, change settings for CSpace and Coagmento plug-in.

The main content area displays a welcome message to "Chirag Shah" to the "CSpace". It states: "Welcome, **Chirag Shah** to your [CSpace](#). You have [6 projects](#) and [9 collaborators](#). Current project: **carrots and sticks**. [Select a different project.](#)" There are links for "Help" and "Logout".

Below the welcome message, there is a section for "Download [Coagmento Beta 2](#) (Release date: 10/08/2009)".

The main content area also includes a section titled "Welcome to your CSpace - short for Collaborative Space or Coagmento Space (because sometimes you don't need a collaborator!). You can visit your CSpace anytime by clicking on 'CSpace' (home) button on your Coagmento toolbar."

From the menu on the left side, you can do a number of things, such as choosing the project that you want to work with, changing settings, and inviting someone to join you for a project.

Clicking on a bar in the menu will open it up with more options. You can find specific help for each option on their individual pages, which will be displayed right here.

Remember, this is work-in-progress, so things will keep changing. But don't worry. All your data and other information will remain intact. And hey, do let us know what you think!

Known issues at the moment:

- Toolbar does not update the information immediately. The update happens once the page is reloaded or a new tab or window is selected.
- Data filtering doesn't always work well. There's some bug there.
- Deletion and moving of data items not working.
- Sometimes clicking on a link shows 'Loading..' picture and doesn't load the actual page. If you see this for more than a few seconds, click on that link again.
- There is an issue with changing the projects. When you choose a different project as your active project, you should see the 'Current project' name changed above. If it doesn't, reload the page.
- Updating profile picture is not functional.
- When collecting objects other than text snippets, source information (e.g., URL) not captured.

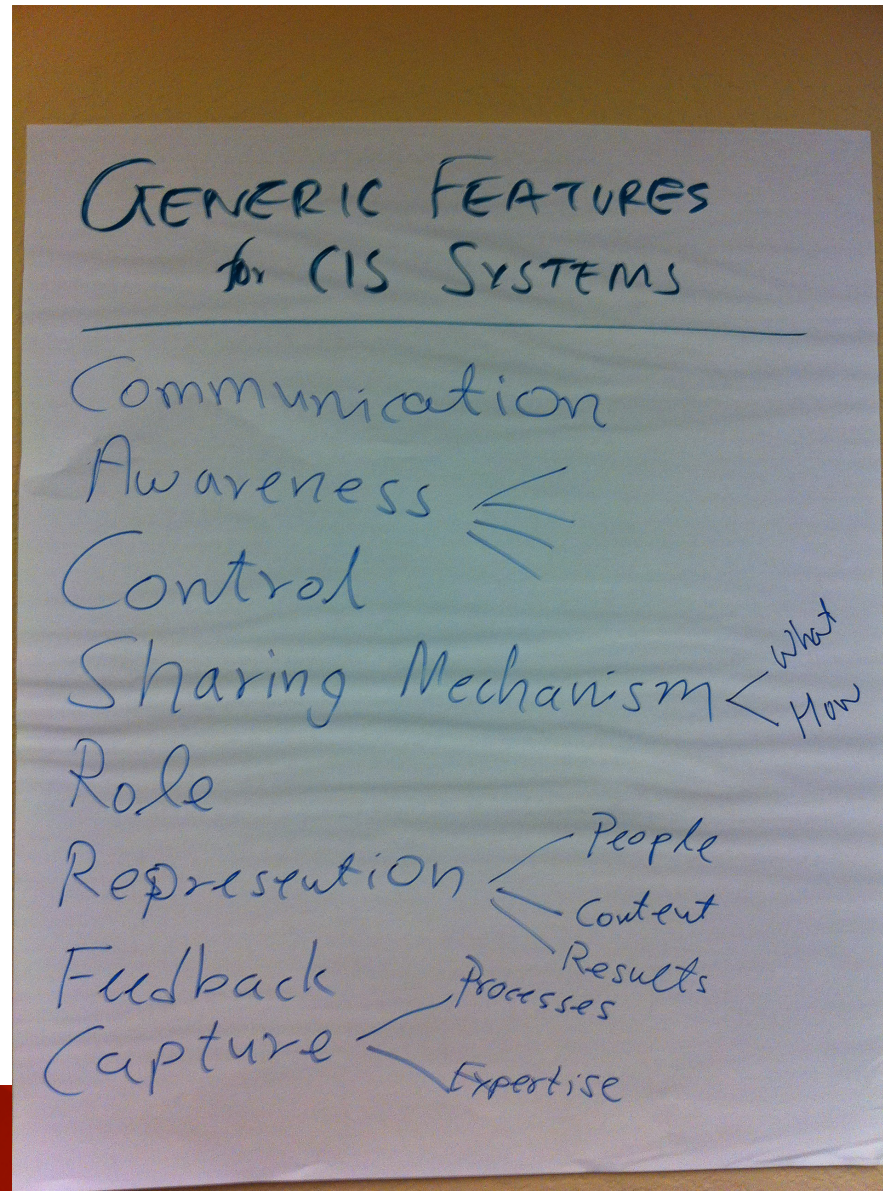
The right sidebar contains the following sections:

- Reload the Sidebar**: A link to reload the sidebar.
- Chat**: A chat window with a "Send" button. The chat history shows: "bugs: hey there", "daffy: what's up?", "bugs: looking for carrots", "daffy: of course you are!", "daffy: good luck! ;)".
- History**: A section with tabs for "Searches", "Bookmarks", and "Objects". The "Searches" tab is active, showing a list of searches: "bugs bunny daffy duck animation", "daffy: movie (google)", "daffy: daffy duck (live)", "bugs: bunny (cnn)", and "bugs: bugs bunny (yahoo)".
- Notepad**: A section for taking notes.
- Notifications**: A section for notifications.

The bottom of the browser window shows a "Done" button and a small icon in the bottom right corner.

Class Exercise

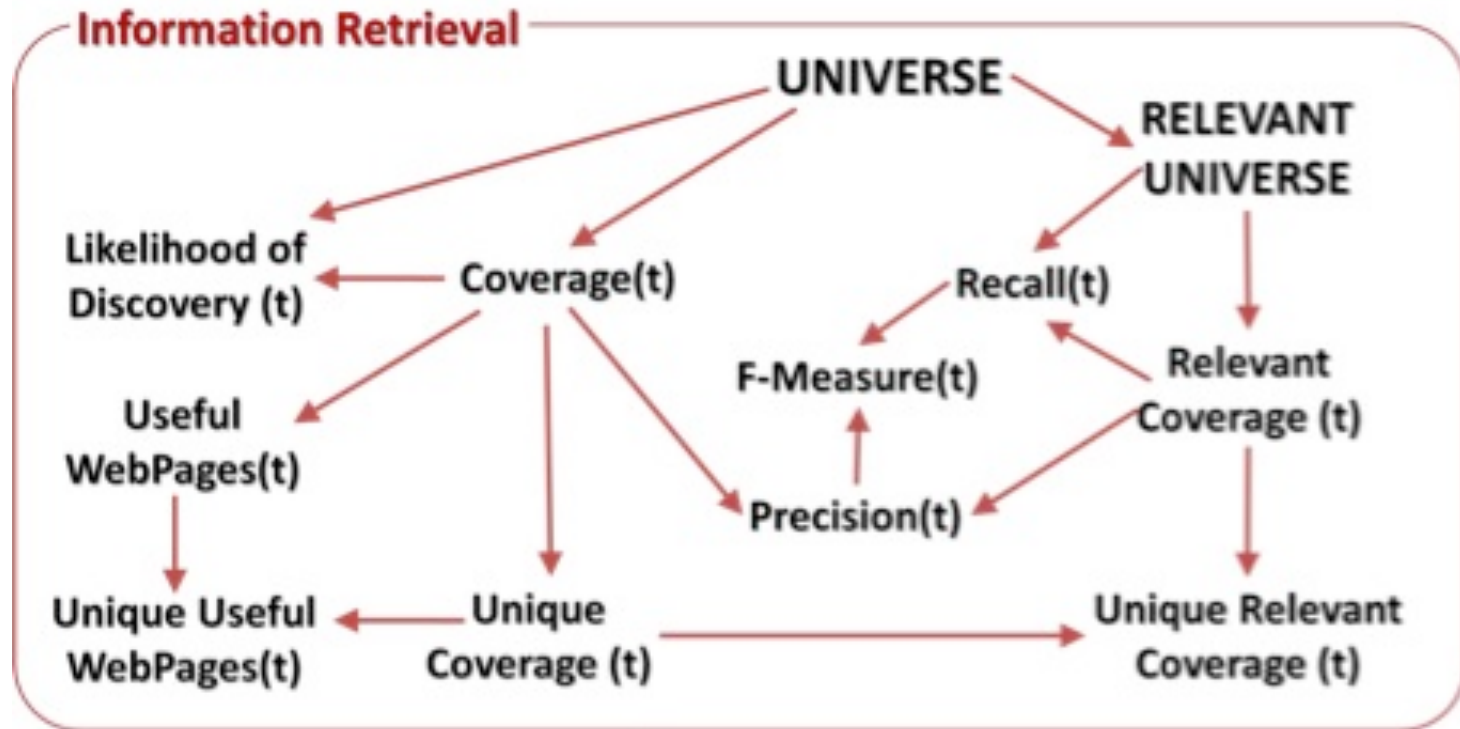
Discuss essential and desired features of a CIR system



Evaluation

- **Traditional measures:** precision, recall, coverage, novelty, diversity
- **Productivity measures:** effectiveness, efficiency
- **Usability measures:** ease of learning, ease of use, satisfaction, cognitive load
- **Other:** engagement, awareness

Evaluation



Query Diversity

Lavenshtein distance

User Measures

- Cognitive Load
 - NASA's Task Load index (TLX)
- Engagement

Evaluation

$$U = \bigcup_t Coverage(t) \quad \dots (1)$$

$$U_r = \bigcup_t RelevantCoverage(t) \quad \dots (2)$$

$$Precision(t) = \frac{RelevantCoverage(t)}{Coverage(t)} \quad \dots (3)$$

$$Recall(t) = \frac{RelevantCoverage(t)}{U_r} \quad \dots (4)$$

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad \dots (5)$$

Evaluation

$$Coverage(t) = \{wp_i : wp_i \text{ was visited by } t \wedge wp_i \in U\} \quad \dots (6)$$

$$UniqueCoverage(t) = Coverage(t) \setminus \bigcup_{t_i \in (T \setminus \{t\})} Coverage(t_i) \quad \dots (7)$$

$$RelevantCoverage(t) = Coverage(t) \cap U_r \quad \dots (8)$$

$$UniqueRelevantCoverage(t) = UniqueCoverage(t) \cap U_r \quad \dots (9)$$

Shah, C., and Gonzalez-Ibanez, R. (2011). Evaluating the Synergic Effect of Collaboration in Information Seeking. In *Proceedings of ACM SIGIR 2011*. Beijing, China. July 24-28, 2011.

Further Research & Development

Theory

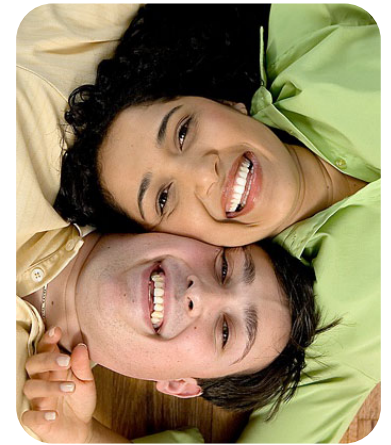
- Framework for CIR
- Models for CIR
- Social aspect of IR

Practice

- Design principles for CIR systems
- Building specialized tools for CIR
- Evaluation

Day 2 Conclusion

- User-mediated CIR
- Examples of systems: Ariadne, SearchTogether, Coagmento
- Essential and desired features of a CIR system
- Evaluation of users and systems in CIR environments
- CIR – a fairly new and fertile research area



Wrap-up and Takeaways

- IR doesn't need to be a solitary pursuit.
- Need more studies and more support for collaborative IR.
- Mediating collaboration – algorithmic or system-based and UI-based or user-centric
- Two major dimensions – time and space
- Evaluation – beyond traditional IR measures

<http://collab.infoseeking.org>

<http://www.coagmento.org>

<http://comminfo.rutgers.edu/~chirags>

Twitter: @chirag_shah #cis

Email: chirags@rutgers.edu

