

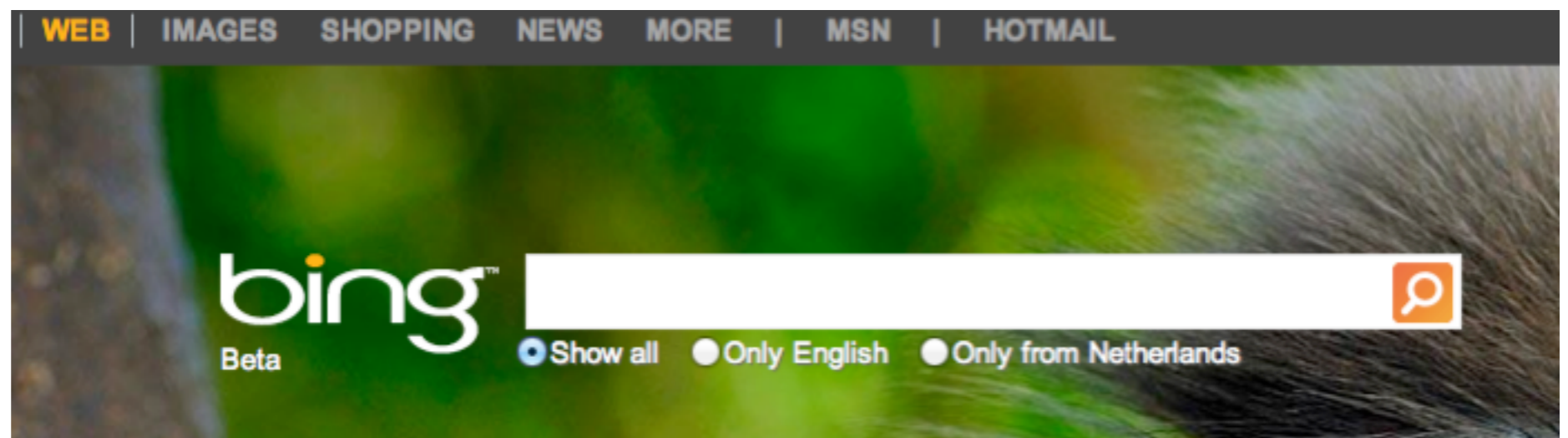
# **(Interactive) Information Retrieval for Complex Search Tasks**

Jiyin He  
(Interactive) Information Access Group  
CWI scientific meeting  
Feb. 8, 2013

# (Interactive) Information Retrieval for Complex Search Tasks

- The problem
- User, system, and their interactions
- Two examples of system- and user-end studies
- Moving forward

# Is search a solved problem?



# To some extent ...

The image displays two side-by-side screenshots of search results for the query "johnny depp".

**Left Screenshot (Search Engine 1):**

- Search query: johnny depp
- Results: 9.340.000 RESULTS
- Navigation: Narrow by language, Narrow by re
- Section: [Images of johnny depp](#) (bing.com/images)
- Images: A row of four small images showing Johnny Depp in different styles.
- Section: [Johnny Depp - Wikipedia](#) (Translate this page)
- URL: [nl.wikipedia.org/wiki/Johnny\\_Depp](http://nl.wikipedia.org/wiki/Johnny_Depp)
- Text: Levensloop · Prijzen en nominaties · Filmografie · Trivia · Exter John Christopher (**Johnny**) **Depp** II (Owensboro (Kentucky) , 9 Amerikaans acteur . **Depp** werd geboren als zoon van stedenb Christopher ...
- Section: [Dutch Johnny Depp fan site, always up to...](#) (Trans)
- URL: [www.johnnydeppfan.nl](http://www.johnnydeppfan.nl)
- Text: **Johnny Depp** bij Aerosmith. gepost op 09-08-2012 door Ilse . S plaatsing maar ook ik was er even tussen uit, maar ...
- Section: [Johnny Depp - IMDb](#)
- URL: [www.imdb.com/name/nm0000136](http://www.imdb.com/name/nm0000136)
- Text: Born 9-6-1963 · 49 years old · [News](#) · [709 photos](#) · [Biography](#) · Actor: Edward Scissorhands (1990) · Sweeney Todd: The Dem (2007) · Alice in Wonderland (2010) · Pirates of the Caribbean:

**Right Screenshot (Search Engine 2):**

- Search query: johnny depp
- Navigation: Web, Images, Maps, Shopping, News, More, Search tools
- Results: About 76,400,000 results (0.26 seconds)
- Section: [Johnny Depp - IMDb](#)
- URL: [www.imdb.com/name/nm0000136/](http://www.imdb.com/name/nm0000136/)
- Text: **Johnny Depp**, Actor: Edward Scissorhands. Born John Christopher Depp in Owensboro, Kentucky, on June 9, 1963, **Johnny Depp** was raised in Florida. [Filmography by type](#) - [Biography](#) - [Filmography by year](#) - [Blow](#)
- Section: [Johnny Depp - Wikipedia, the free encyclopedia](#)
- URL: [en.wikipedia.org/wiki/Johnny\\_Depp](http://en.wikipedia.org/wiki/Johnny_Depp)
- Text: John Christopher "**Johnny**" **Depp** II (born June 9, 1963) is an American actor, film producer, and musician. He has won the Golden Globe Award and Screen ... [Vanessa Paradis](#) - [The Rum Diary](#) - [The Lone Ranger](#) - [Johnny Depp filmography](#)
- Section: [JohnnyDeppWeb.com](#)
- URL: [www.johnnydeppweb.com/](http://www.johnnydeppweb.com/)
- Text: Photos, news and information, interactive materials, multimedia, and links are included in this fan site.
- Section: [Johnny-Depp.org » an admirers' site dedicated to Johnny Depp and ...](#)
- URL: [johnny-depp.org/](http://johnny-depp.org/)
- Text: 13 hours ago – **Johnny Depp** fan website with news, biography, filmography, and actor contact information.
- Section: [News for johnny depp](#)

- When your query is about **simple factual** information

# What if it is not as simple?

the big bang theory

10.800.000 RESULTS Narrow by language ▾ Narrow by region ▾

[The Big Bang Theory \(televisieserie\)...](#) Translate this page  
nl.wikipedia.org/wiki/The\_Big\_Bang\_Theory\_(televisieserie) ▾  
Verhaal · Rolverdeling · Afleveringen · Dvd's · Waardering · Bijzonderheden  
The **Big Bang Theory** is een Amerikaanse komische televisieserie van CBS , gesta  
2007. De serie werd bedacht door Chuck Lorre (die ook Two and a Half Men bedacht

[The Big Bang Theory: Watch Episodes and Video and Join the ...](#)  
www.cbs.com/primetime/big\_bang\_theory ▾  
Watch video, browse photos and join the ultimate fan community for The **Big Bang Theory**

[The Big Bang Theory - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/The\_Big\_Bang\_Theory ▾  
Production · Main cast · Recurring themes and ... · Reception · Episodes  
The **Big Bang Theory** is an American sitcom created by Chuck Lorre and Bill Prady both of whom serve as executive producers on the show, along with Steven Molaro.

[The Big Bang Theory \(TV Series 2007– \) - IMDb](#)  
www.imdb.com/title/tt0898266 ▾  
User rating: 8,6/10 · Comedy · TV-14 · 21 min · 179.972 ratings  
Created by Chuck Lorre, Bill Prady. Actors: Johnny Galecki: Leonard Hofstadter · Jim Parsons: Sheldon Cooper · Kaley Cuoco: Penny · Simon Helberg: Howard Wolowitz

[Big Bang - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/Big\_Bang ▾  
Overview · History · Observational evidence · Related issues in physics  
The **Big Bang theory** is the prevailing cosmological model that describes the early development of the Universe. According to the **theory**, the **Big Bang** occurred ...

the big bang theory

Web Images Maps Shopping News More ▾ Search tools

About 105,000,000 results (0.24 seconds)

[The Big Bang Theory: Watch Episodes and Video and Join the ...](#)  
www.cbs.com/shows/big\_bang\_theory/  
Watch video, browse photos and join the ultimate fan community for **The Big Bang Theory**.  
Episodes - Video - Ultimate Fan - Photos

[The Big Bang Theory - Wikipedia, the free encyclopedia](#)  
en.wikipedia.org/wiki/The\_Big\_Bang\_Theory  
The **Big Bang Theory** is an American sitcom created by Chuck Lorre and Bill Prady, both of whom serve as executive producers on the show, along with Steven ...  
List of The Big Bang Theory ... - Characters - Penny - Season 5

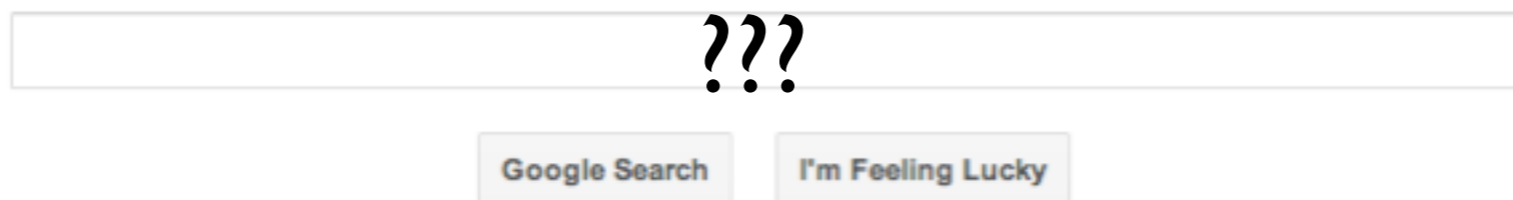
[The Big Bang Theory \(TV Series 2007– \) - IMDb](#)  
www.imdb.com/title/tt0898266/  
★★★★★ Rating: 8.6/10 - 189142 votes  
A woman who moves into an apartment next door to two brilliant but socially awkward physicists shows them how little they know about life outside of...  
Starring Johnny Galecki, Jim Parsons, Kaley Cuoco.

[The Big Bang Theory Site](#)  
the-big-bang-theory.com/  
A fan site for the hit CBS TV show **The Big Bang Theory**. Get news, episode information and more.

- For example, what if your query is **ambiguous**?

# What if it is not as simple?

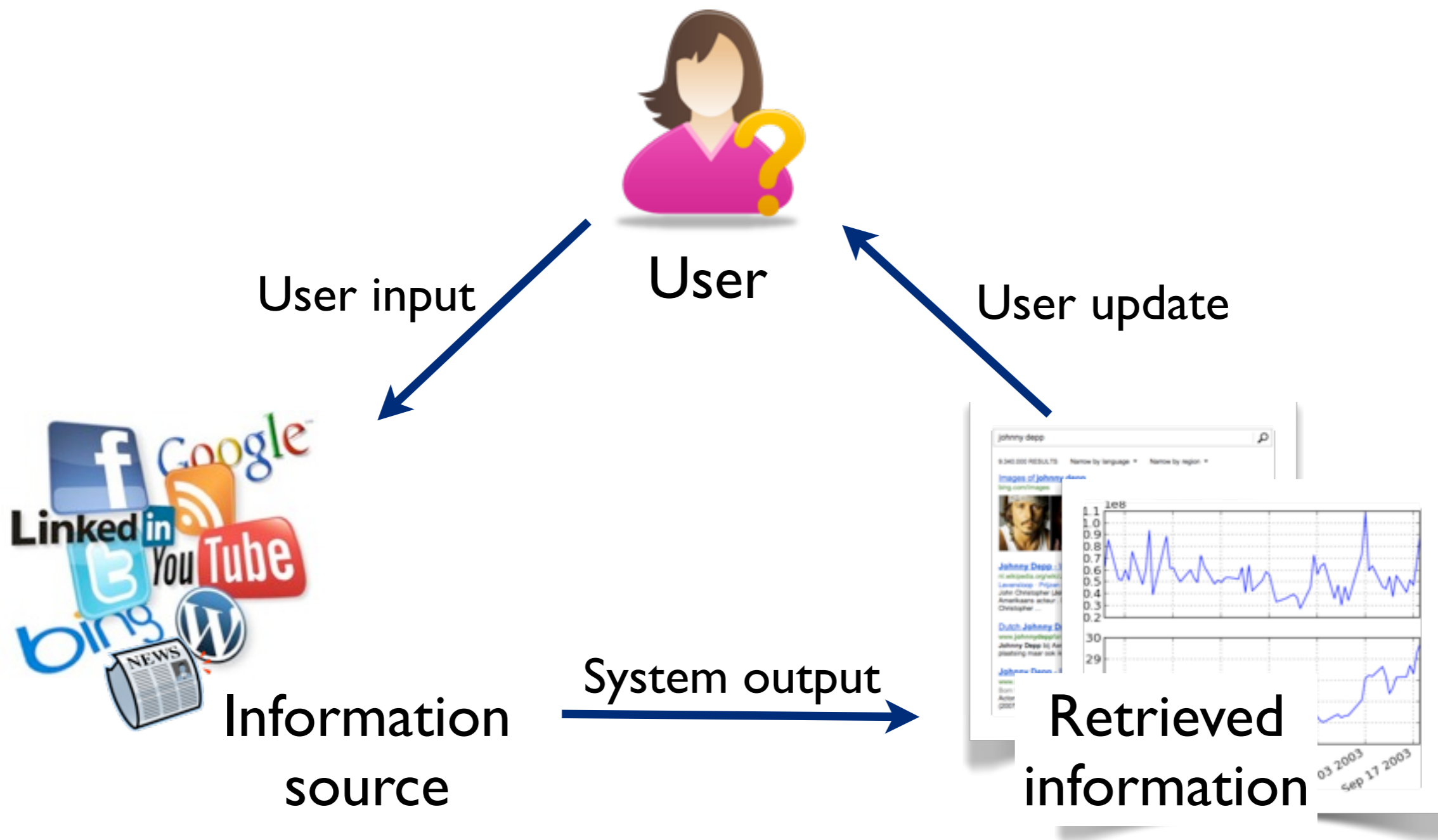
- What if one does not know **how to formulate** his/her query?

The Google logo is displayed in its characteristic multi-colored font.A screenshot of the Google search interface. It features a search bar with three question marks '???' inside. Below the search bar are two buttons: 'Google Search' and 'I'm Feeling Lucky'.

“There was this pizza restaurant in Rome that my friend told me about, but I don’t remember its name or precise location....”

# What if it is not as simple?

- e.g., I'm writing an article on topic X

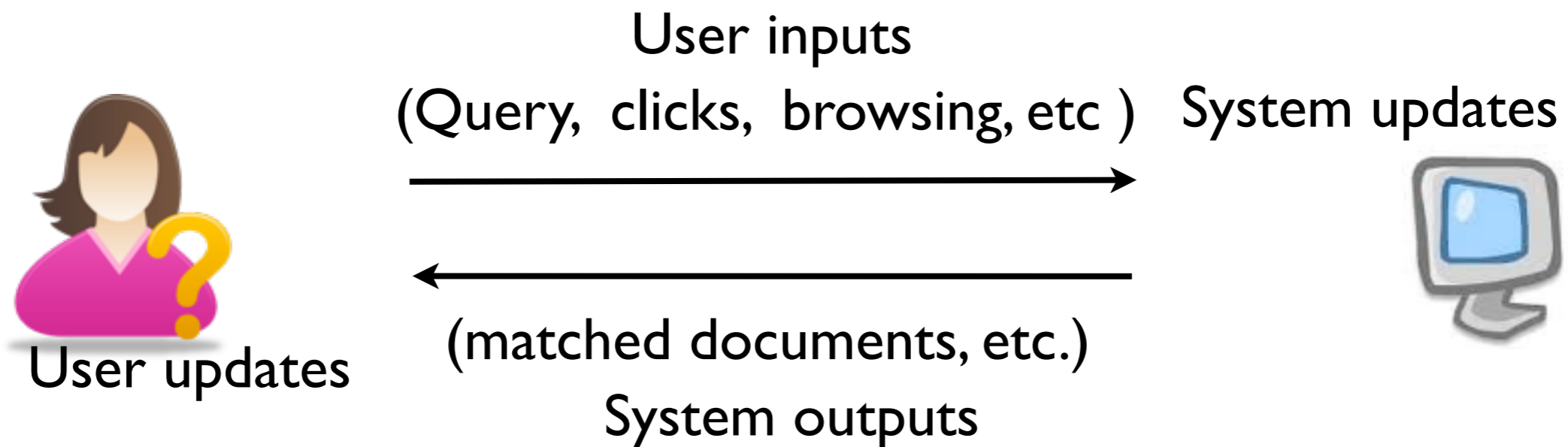


# Complex search tasks

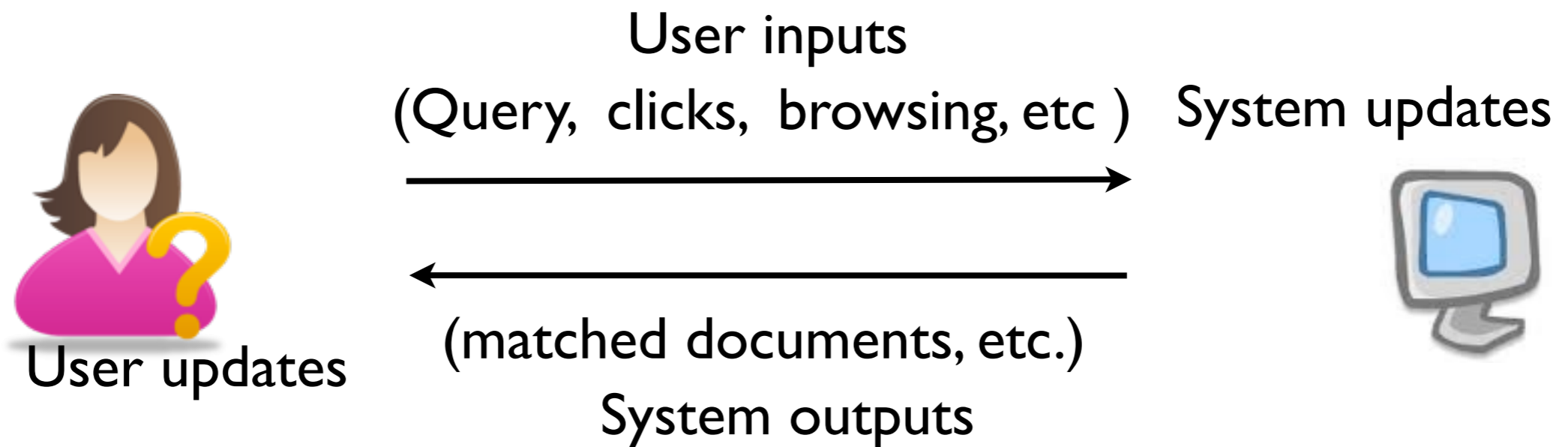
- Users are engaged in multiple iterations of interactions with the system
  - their need is broad/vague/underspecified
  - their need can not be solved in a single shot search
  - their need may evolve during search
- How can we reduce users' effort in this process?



# User, system, and interaction



# User, system, and interaction



For a given assumption about the user's need, what system output should be present?

- adhoc search vs. exploratory search
- topic distillation vs. result diversification
- ...

# User, system, and interaction

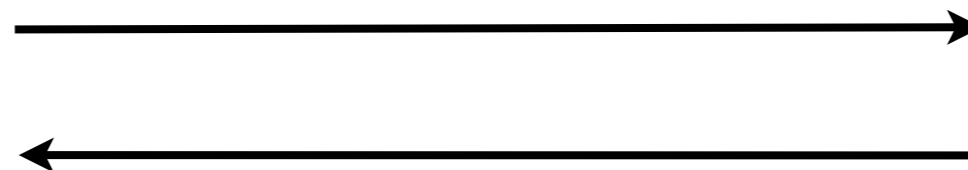
Which assumption about the user is valid?

- know-item seeking vs. exploration
- focused info vs. diverse info
- ...



User updates

User inputs  
(Query, clicks, browsing, etc )



System updates



(matched documents, etc.)



System outputs

For a given assumption about the user's need, what system output should be present?

- adhoc search vs. exploratory search
- topic distillation vs. result diversification
- ...

# An example of a system-end approach for ambiguous queries: result diversification

- Without knowing the actual intent of the user, try to cover **as many** interpretations **as possible** within the **shortest** ranked list

- Where do we find different interpretations?

- How do we construct the ranked list?



Query: python

# Mining interpretations of a query from multiple sources

- A network constructed over interpretations of a query from multiple sources
  - Nodes: interpretations
  - Edges: weighted by similarity between interpretations
- A random walk over the network to find relatedness of interpretations

Q: python  
2000

Q: horror  
movie python



click logs: co-clicks

A: python  
documentation

A: python tutorial

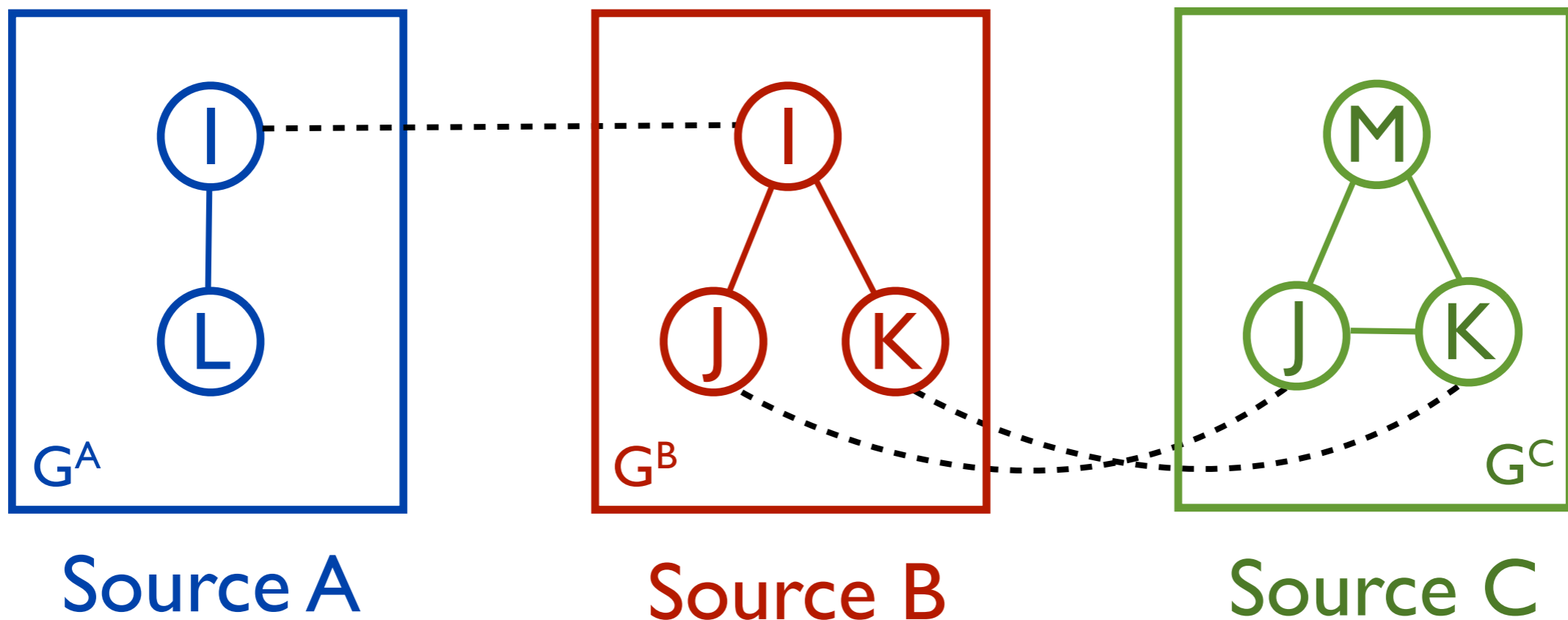


anchor texts: co-reference

(J. He, V. Hollink, and A. P. de Vries, SIGIR 2012)

# Mining interpretations of a query from multiple sources

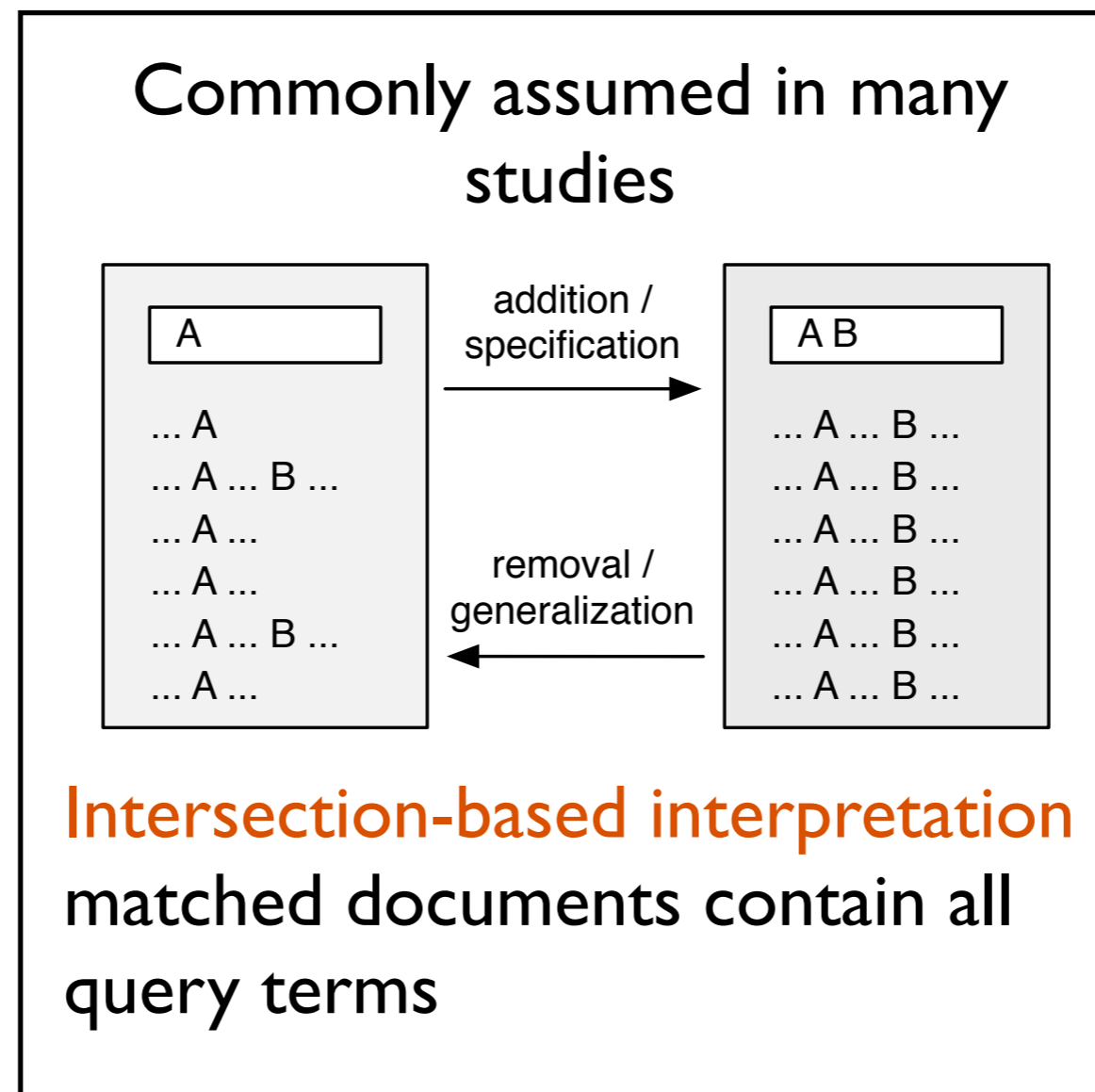
- A network constructed over interpretations of a query from multiple sources
  - Nodes: interpretations
  - Edges: weighted by similarity between interpretations
- A random walk over the network to find relatedness of interpretations



(J. He, V. Hollink, and A. P. de Vries, SIGIR 2012)

# An example of a user-end approach for understanding users' motivation

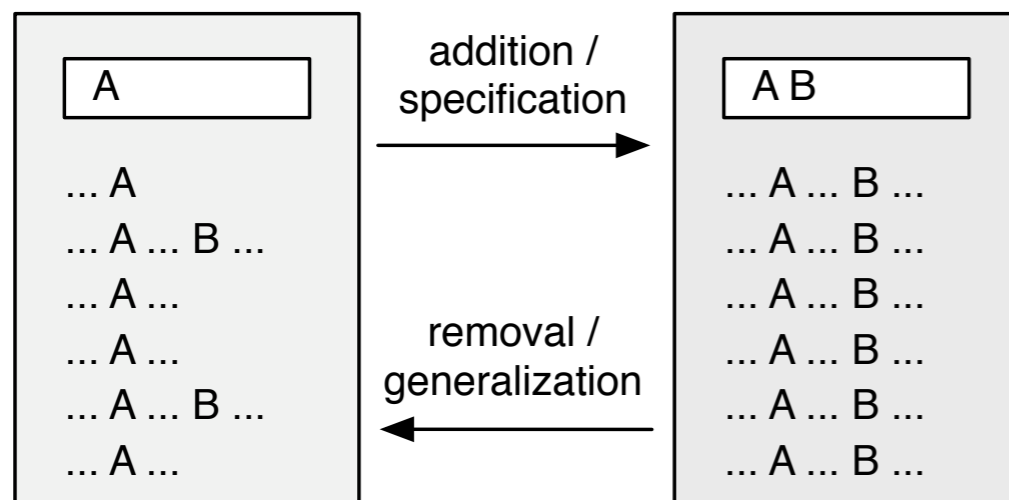
- When users add/remove terms from their queries, what do they attempt to achieve?



# An example of a user-end approach for understanding users' motivation

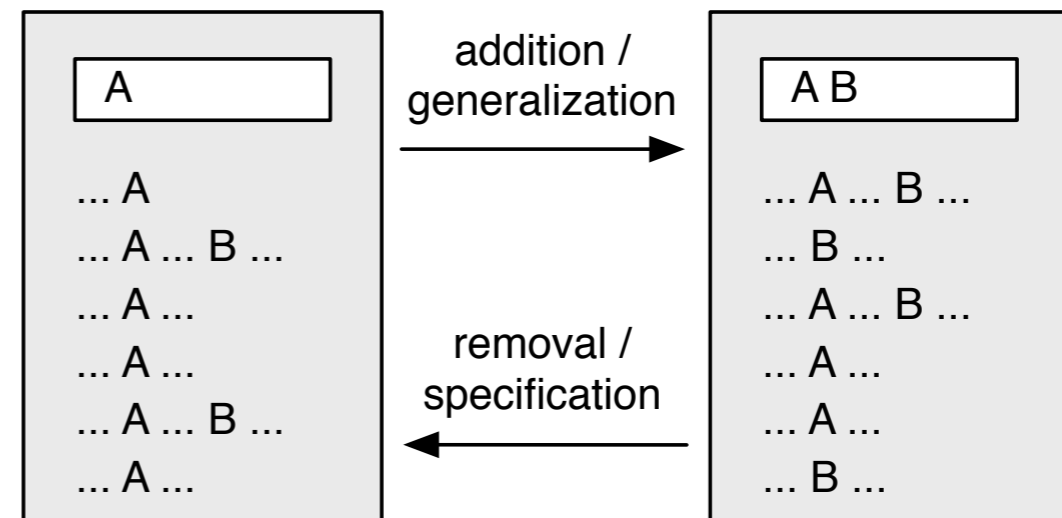
- When users add/remove terms from their queries, what do they attempt to achieve?

## Commonly assumed in many studies



**Intersection-based interpretation**  
matched documents contain all query terms

## An alternative proposal

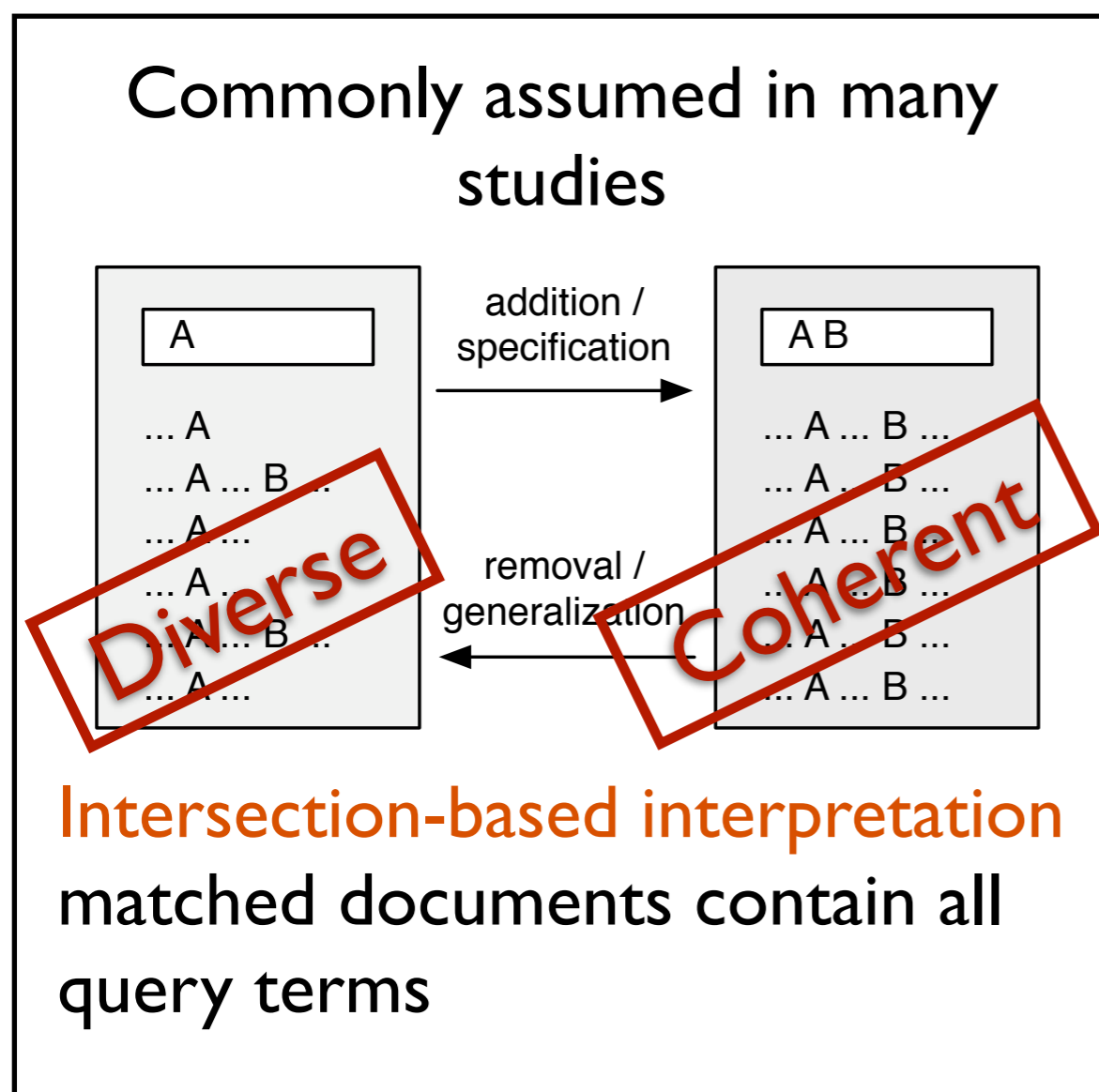


**Union-based interpretation**  
matched documents contain some of the query terms

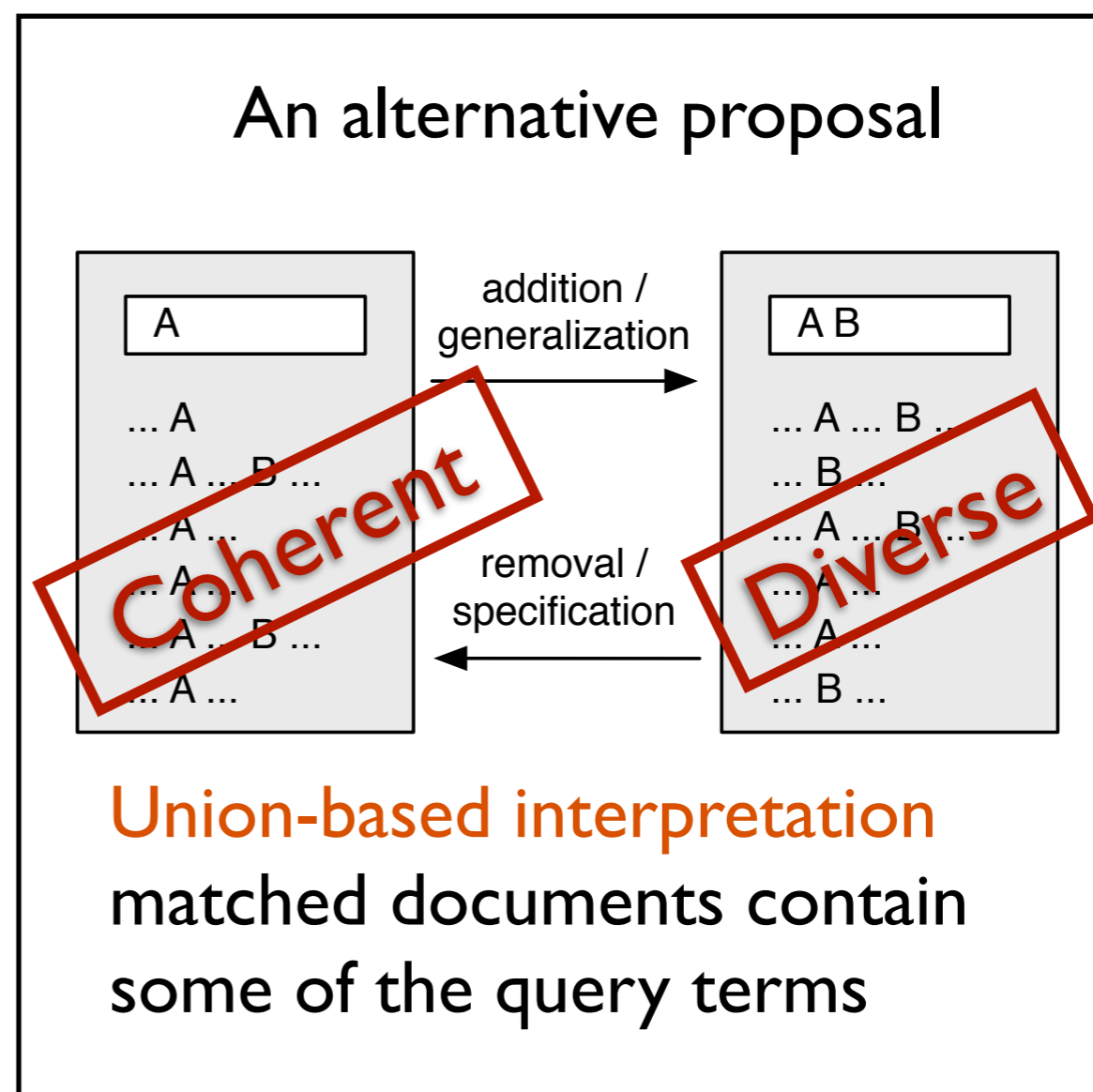


# An example of a user-end approach for understanding users' motivation

- When users add/remove terms from their queries, what do they attempt to achieve?



(V. Hollink, J. He, and A. P. de Vries, ECIR 2012)



# Validation of the two interpretations

- Do more coherent or less coherent result sets more often lead to term removals and term additions?
- Do term removals and term additions increase or decrease the coherence of the result sets?

Data		Avg Sim		Avg Sim Diff			
		A	R	A	R		
News	all	0.56	>>	0.52	-0.016	<<	0.034
	2 terms	0.56	>>	0.52	-0.012	<<	0.034
	>=2 terms	0.56	>>	0.52	-0.013	<<	0.034
iCLEF	all	0.32	>>	0.29	-0.012	<<	0.025
	2 terms	0.34	>>	0.27	-0.029	<<	-0.015
	>=2 terms	0.35	>>	0.29	-0.033	<<	0.025
Web	all	0.28	>>	0.27	0.002	<<	0.010
	2 terms	0.29	>>	0.25	-0.000	>>	-0.042
	>=2 terms	0.30	>>	0.27	-0.014	<<	0.010

>>/<<: significantly larger/smaller with p-value <0.01 (Wilcoxon ranksum/signed-rank test).

# User, system, and interaction: current research in the community

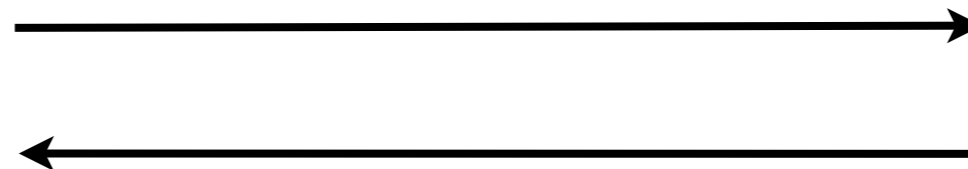
Diverse types of efforts have been made to understand users

- models for user behavior patterns
- studies to understand users' search strategies
- conceptual models describing users' cognitive activities

with respect to specific system setup



User updates

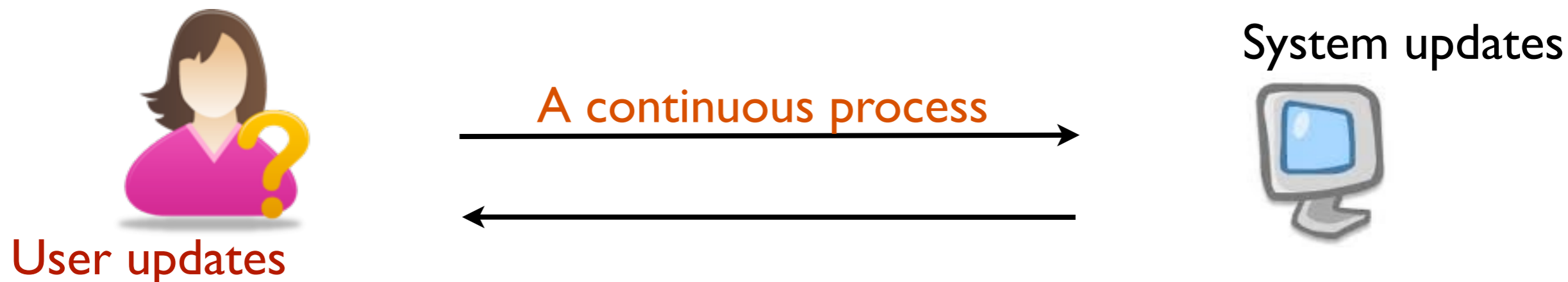


System updates



Diverse types of retrieval models are developed, each optimized with respect to specific assumptions about users/tasks

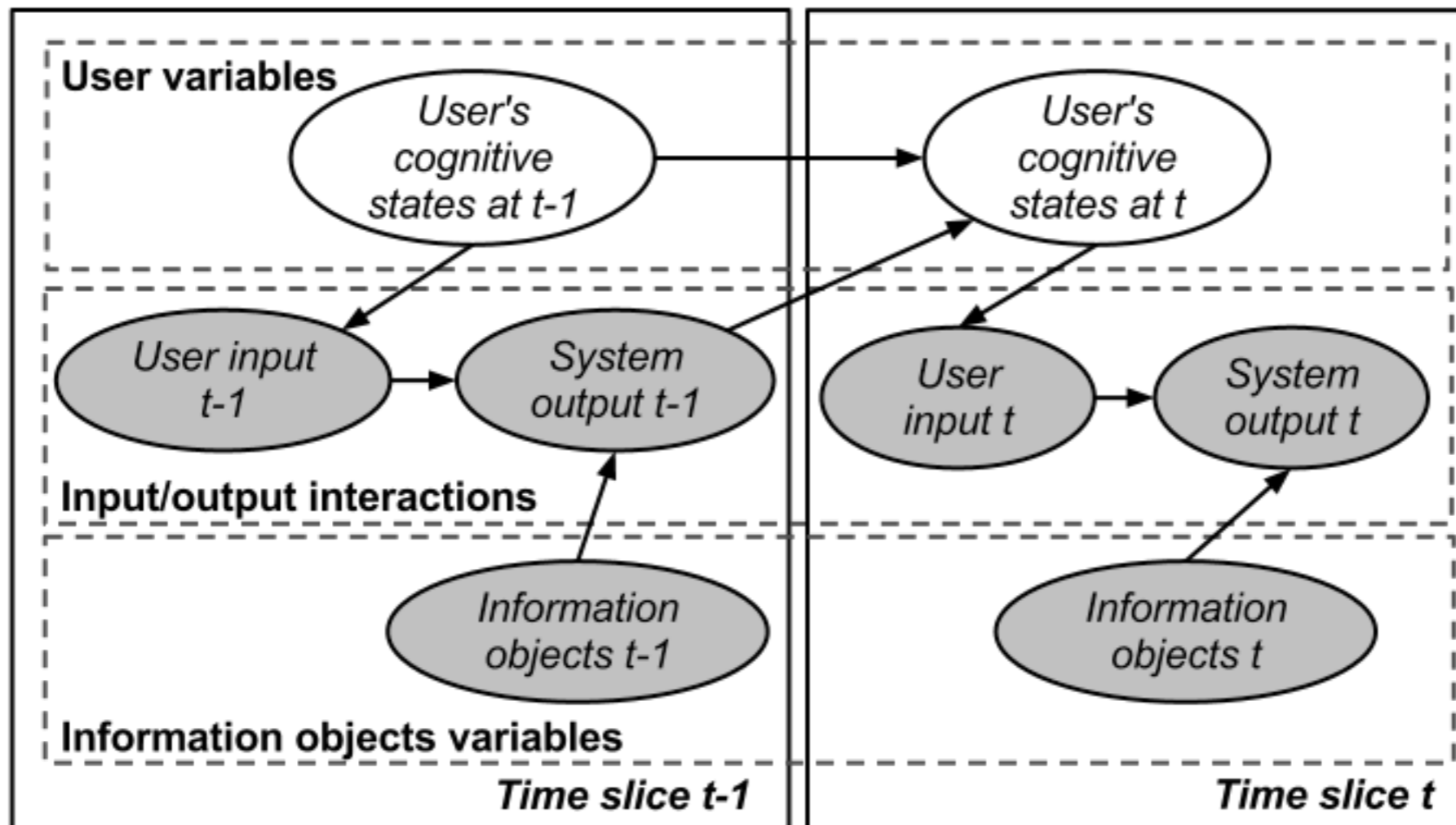
# User, system, and interaction: to move forward



- Which assumption is valid and **when?** (thus which retrieval model is appropriate)
  - Users information needs are not static during the interaction iterations
- How can we put together various system- and user-end studies
  - that are specific to certain assumptions
  - that consider different sets of variables

# User, system, and interaction: to move forward

- Can we formally model the system-user dependencies, and the evolvement of user information needs over time?



# (Interactive) Information Retrieval for Complex Search Tasks

- The problem
- User, system, and their interactions
- Two examples of system- and user-end studies
- Moving forward

Thank you and questions?

