Quo Vadis Information Retrieval Research

Kal Järvelin
University of Tampere

Agenda

- 1. Introduction
- 2. Some Concepts
- 3. The Laboratory Approach to IR
- 4. Mounting Critical Evidence
- 5. Cognitive Framework for Research on IIR
- 6. Concluding Discussion

1. Introduction

- IR theory?
  - SIGIR theory = Formal Retrieval Models
  - Not really empirical theories to be confirmed or refuted
  - Are there other types of theories?
- What theories are we trying to construct?

Motivation

- The ultimate goal of information retrieval is support humans to better access information in order to carry out their task.

Agenda

- ✓ 1. Introduction
- 2. Some Concepts
- 3. The Laboratory Approach to IR
- 4. Mounting Critical Evidence
- 5. Cognitive Framework for Research on IIR
- 6. Conclusions

2. Concepts : Frameworks, Models

- Frameworks in research
  - Essential objects to study
  - The relationships of objects
  - The changes in the objects / relationships that affect the functioning of the system
  - Promising goals and methods of research
- The concept model
  - A precise (often formal) representation of objects and relationships (or processes) within a framework
  - Modeling may also in principle encompass human actors and organizations

Hypotheses, Laws, Theories

- Variables
  - represent objects etc.
  - are used in hypotheses, laws ...
- Hypotheses
  - state verifiable facts / relationships whose truth is unknown.
- Scientific laws
  - empirical laws express verified relationships between objects, properties or events
- Theories
  - systematic collections of theoretical and empirical laws

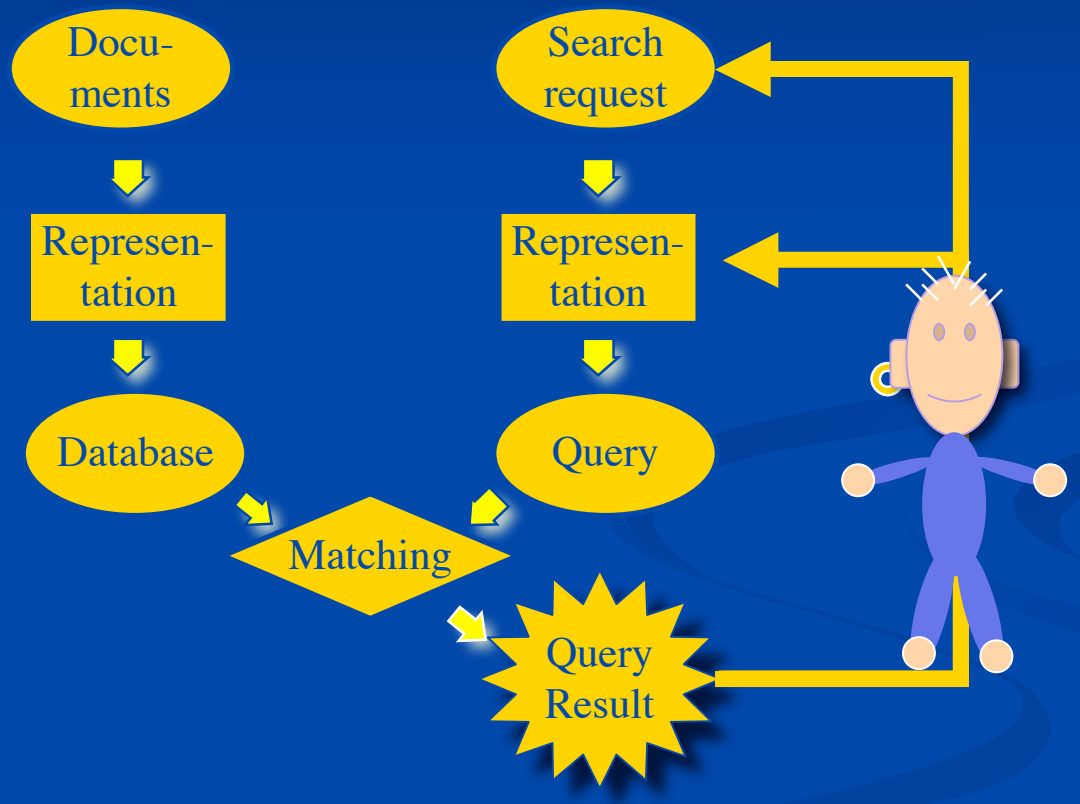
Variables

- Types of variables in study designs:
  - dependent variables – the variation of which is explained
  - independent variables – the ones systematically varied in order to see the responses in the dependent ones
  - controlled variables – the ones fixed to prevent uncontrolled variation in the results
  - hidden variables – all other variables

Agenda

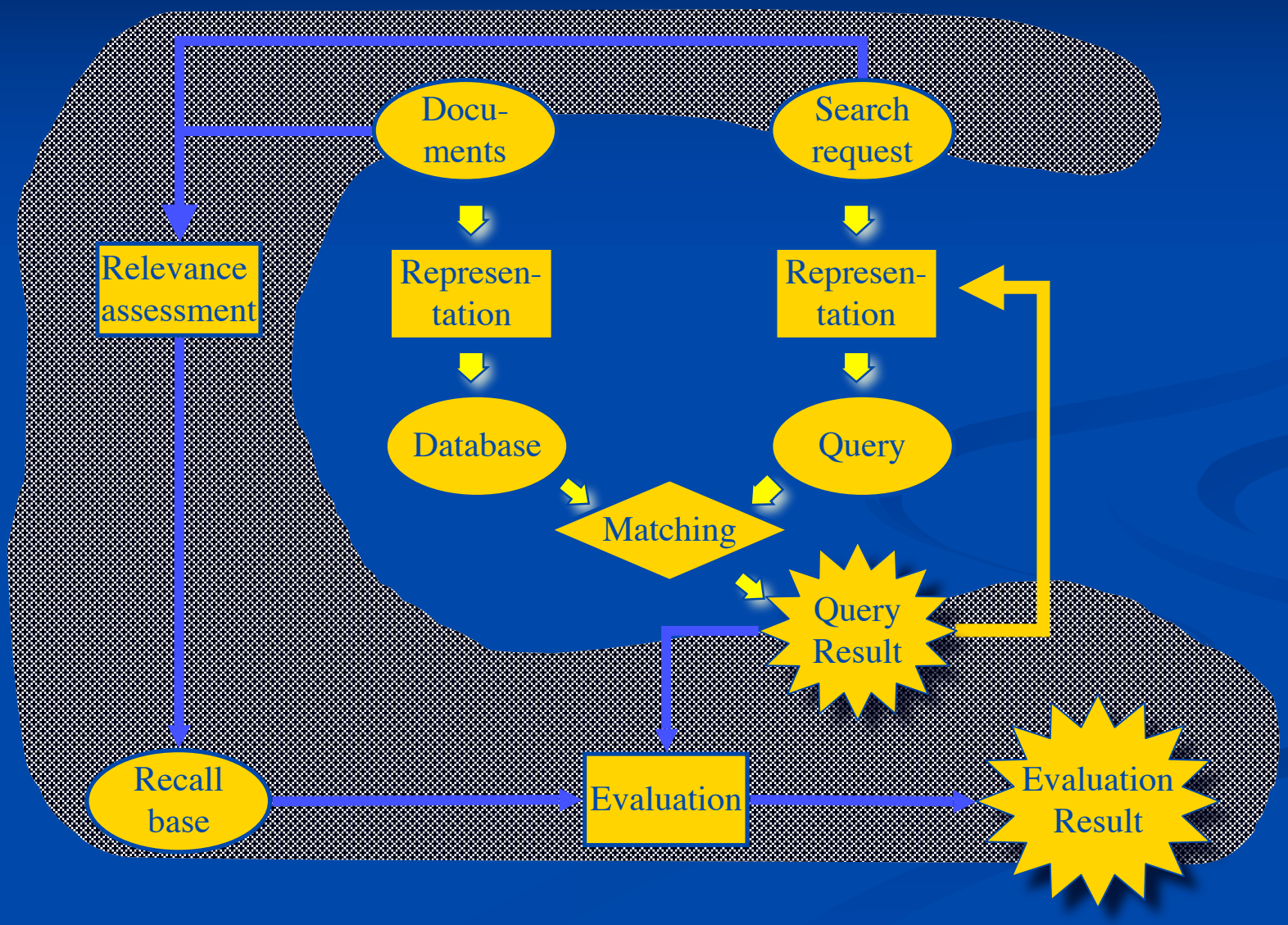
- ✓ 1. Introduction
- ✓ 2. Some Concepts
- 3. The Laboratory Approach to IR, ltd.
  - The Framework and Model(s)
  - Variables, Hypotheses, Laws and Theories
- 4. Mounting Critical Evidence
- 5. Cognitive Framework for Research on IIR
- 6. Conclusions

3. The Laboratory Approach to IR

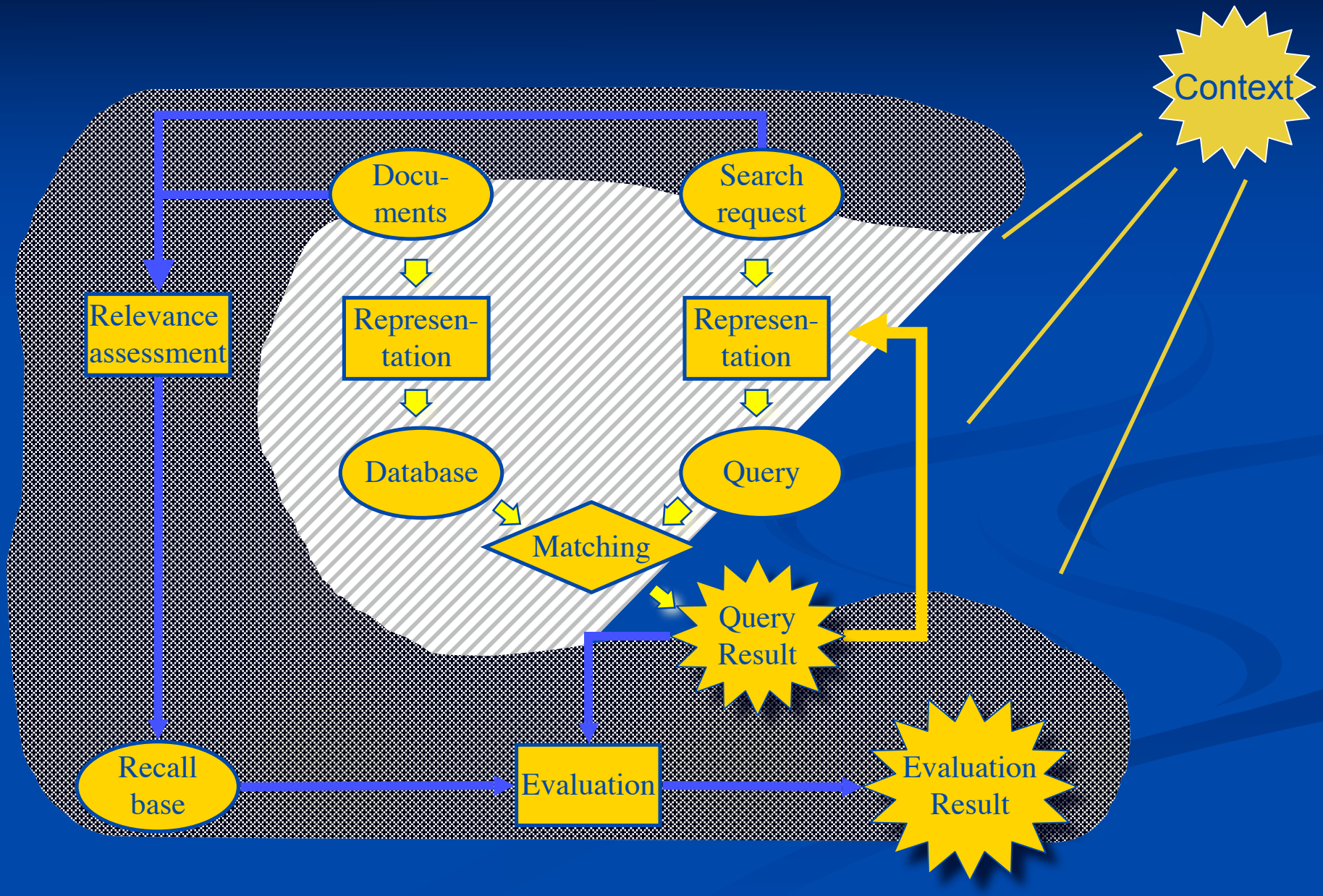


This is information retrieval, isn't it? But where is the lab?

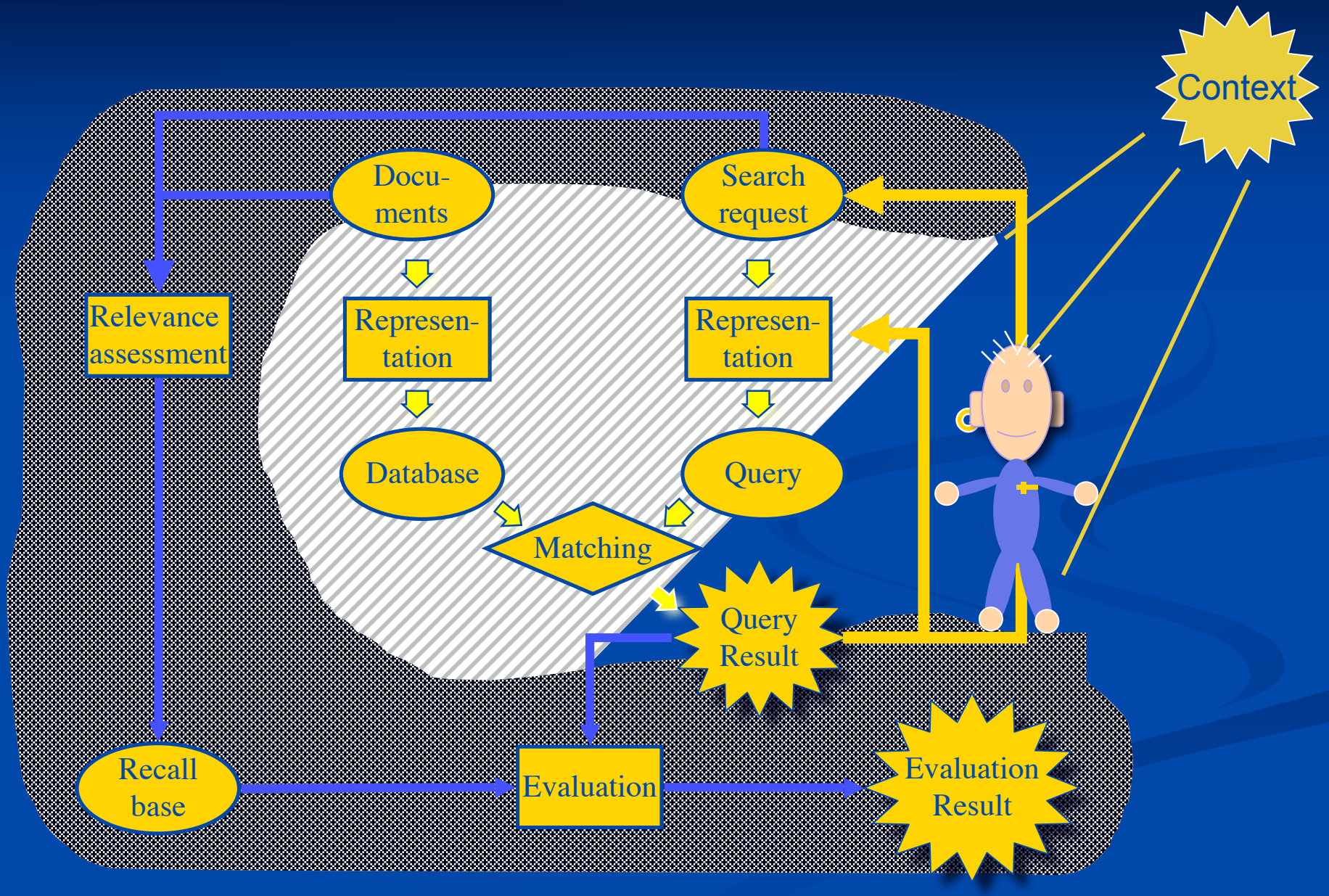
The Lab Included – and User Dropped



The Lab IR Cave

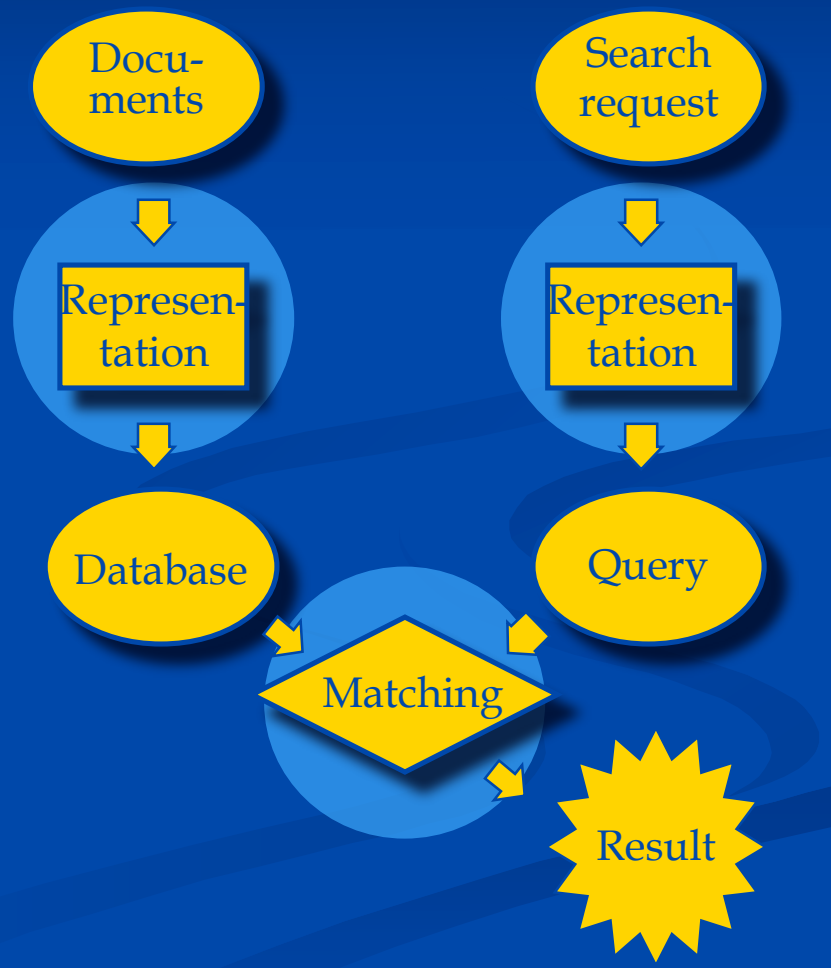


The Lab IR Cave, with a Visitor

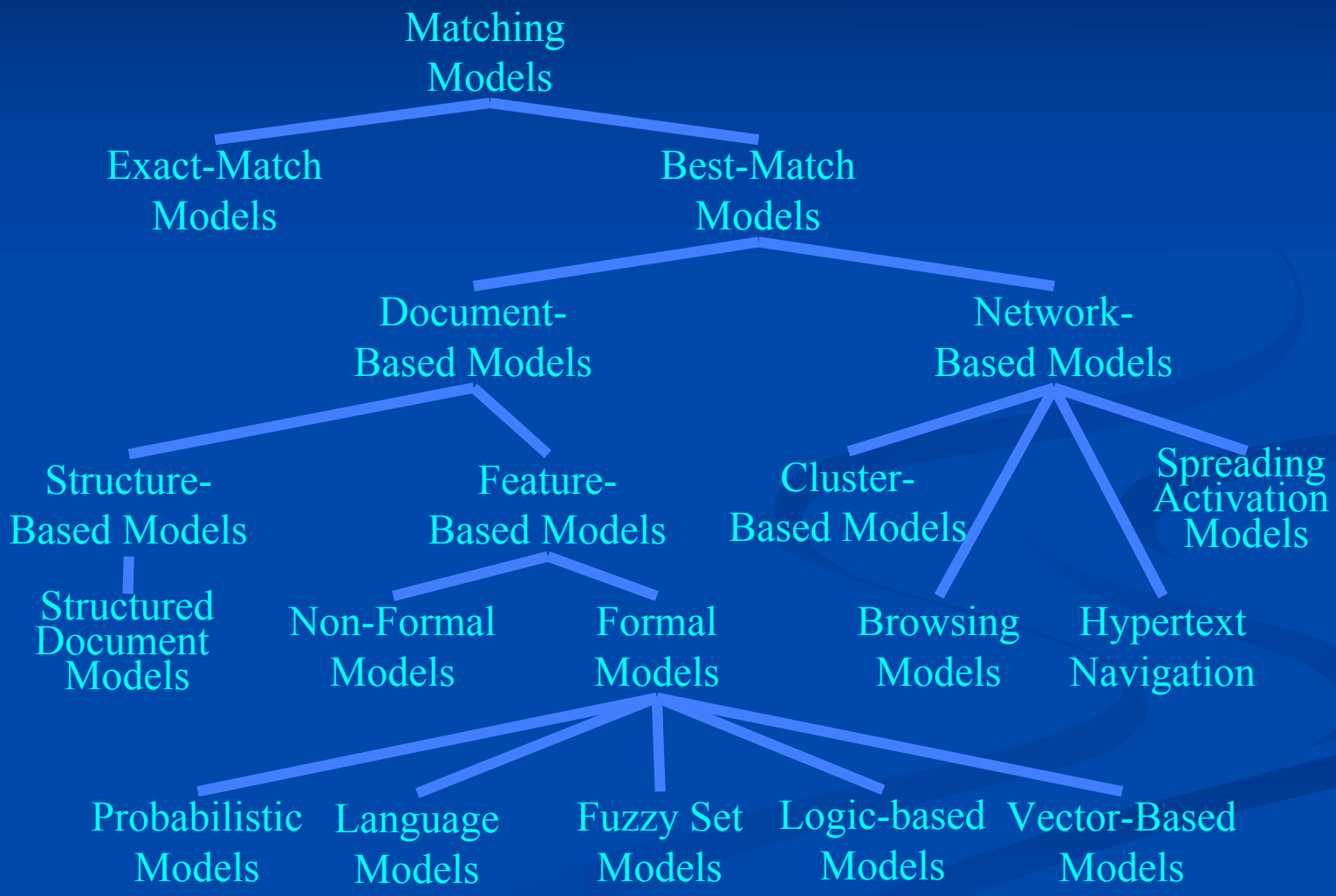


Lab IR: The Model(s)

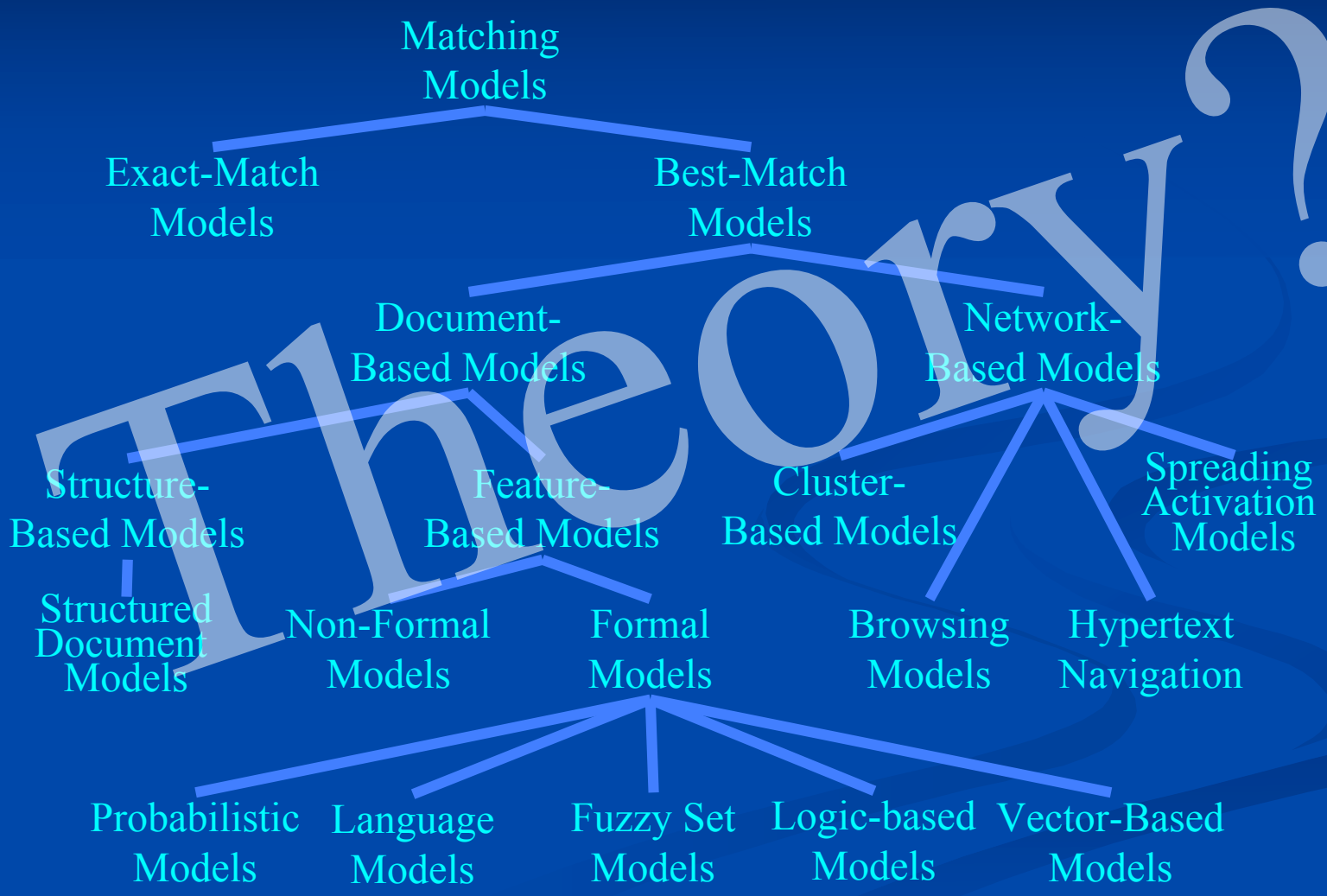
- Models in IR are retrieval models which specify
  - document and request representations, and
  - the matching algorithm for comparing these representations



Lab IR: Models

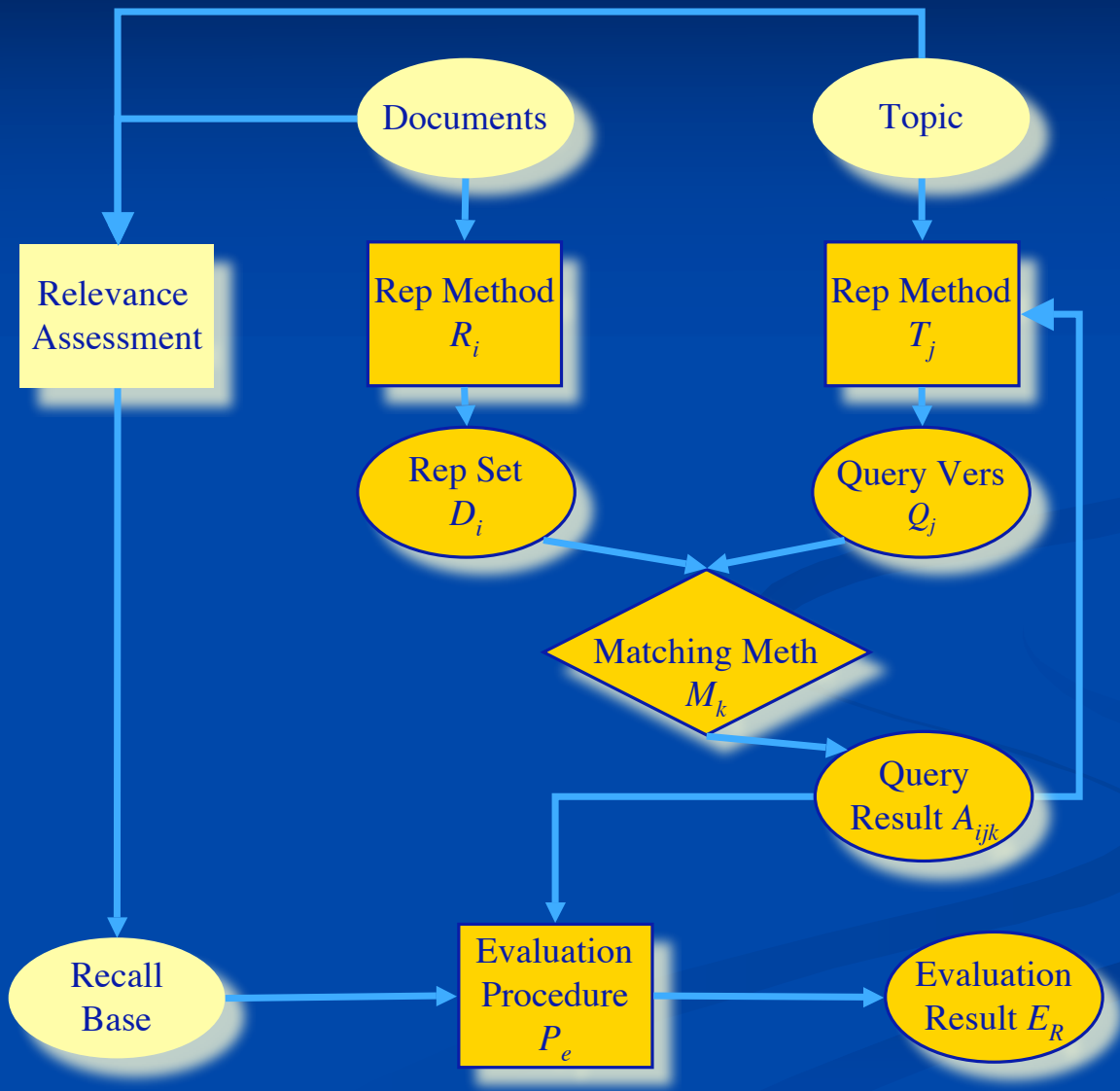


Lab IR: Models



Theory?

The Laboratory Setting: Variables



So What is Lab IR All About?

- Variables, Hypotheses, Laws, and Theories are about the explanation of IR effectiveness.
- The dependent variables typically are recall and precision or derived from them (e.g., MAP, nDCG).
- The explaining factors, the independent variables, are the use / non-use of selected techniques implementing retrieval models.
- The controlled variables are test collections, topics, assessments.
- There hardly are any critical hypotheses - cf. Physics

... All About?

- Strong standardization of the research designs facilitates comparison of results and has, admittedly, led to much progress in IR practice
- However, there strength and success of the approach may be a straightjacket.
- There is mounting evidence that the Lab IR approach may be ...

… Plato's Cave

The fire
Prisoners
shadows cast
on wall
Roadway where
puppeteers perform

LabIR: Framework Issues

- W Y D S I W Y D U
  - Tasks
  - Searchers
  - Relevance assessments
  - Interface functionalities
  - Search processes
- C N C L ?

CCBImages.com

Top Fuel - Which one is better for real life?

Agenda

- ✓ 1. Introduction
- ✓ 2. Some Concepts
- ✓ 3. The Laboratory Approach to IR
- 4. Mounting Critical Evidence on ...
  - information needs, relevance
  - system vs. user performance
  - sessions, systems integration
- 5. Cognitive Framework for Research on IIR
- 6. Conclusions

Information needs, good queries?

- Articulated needs assumed in IR
- Kuhlthau (1994) and Byström & Järvelin (1995) have shown that sometimes there are no articulated information needs preceding information access.
- Embarrassment and confusion often present - Belkin's ASK (1982)

Kuhlthau 1994

Stages	Initiation	Selection	Exploration	Formulation	Collection	Presentation
Feelings	Uncertainty	Optimism	Confusion, frustration, doubt	Clarity	Sense of direction, confidence	Relief, satisfaction or disappointment
Thoughts	Vague			Clearer	Increased interest	Focused
Actions	Seeking background information		Seeking relevant information		Seeking pertinent information	
Appropriate tasks	Recognize	Identify, investigate	Identify, investigate	Formulate	Gather	Complete

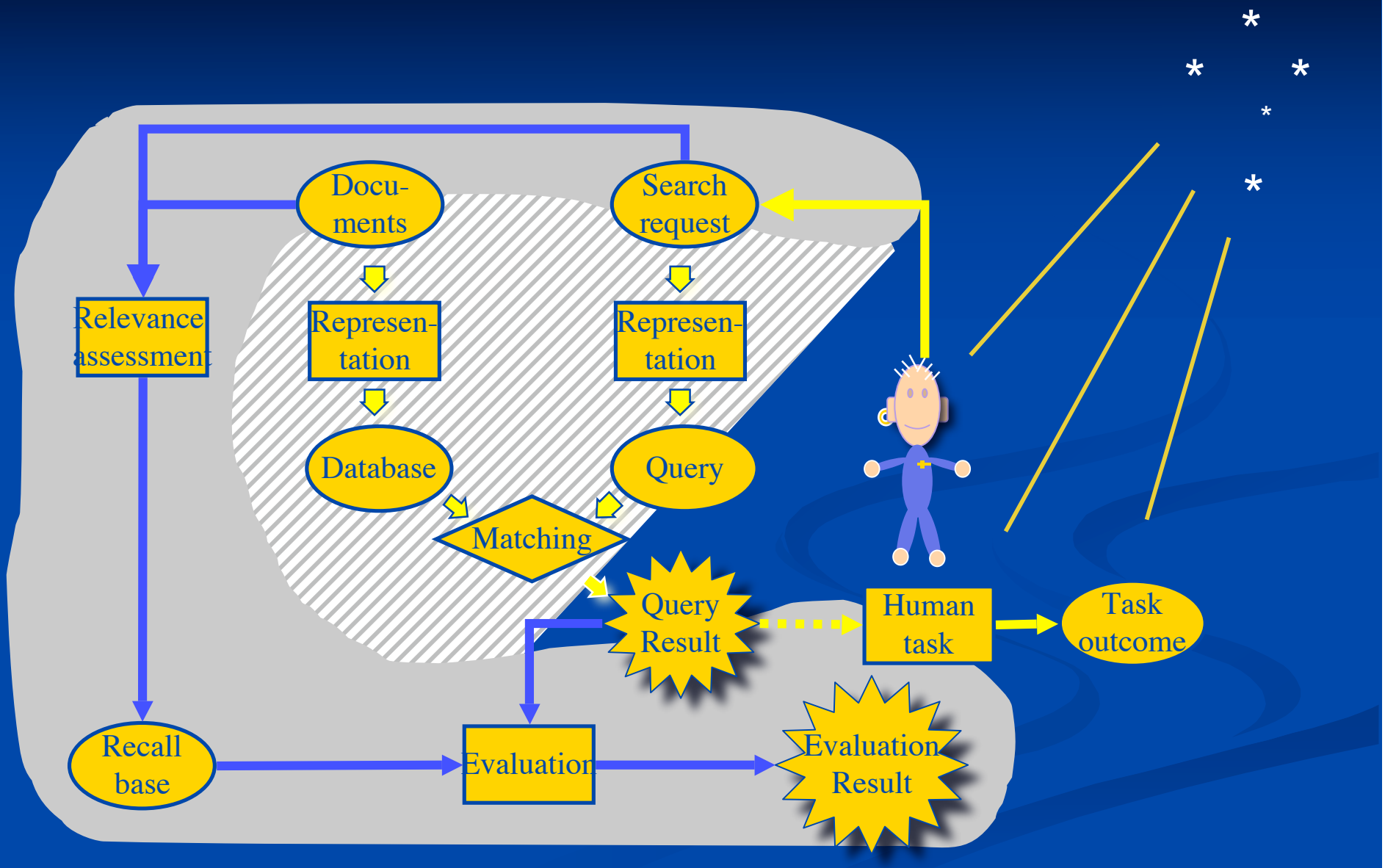
Information needs, ..., relevance?

- Relevance
  - multiple degrees, multi-dimensional, individual
- ... depends on problem stage, difficulty, and information construction.
- Clicks perhaps not reliable predictors of relevance.

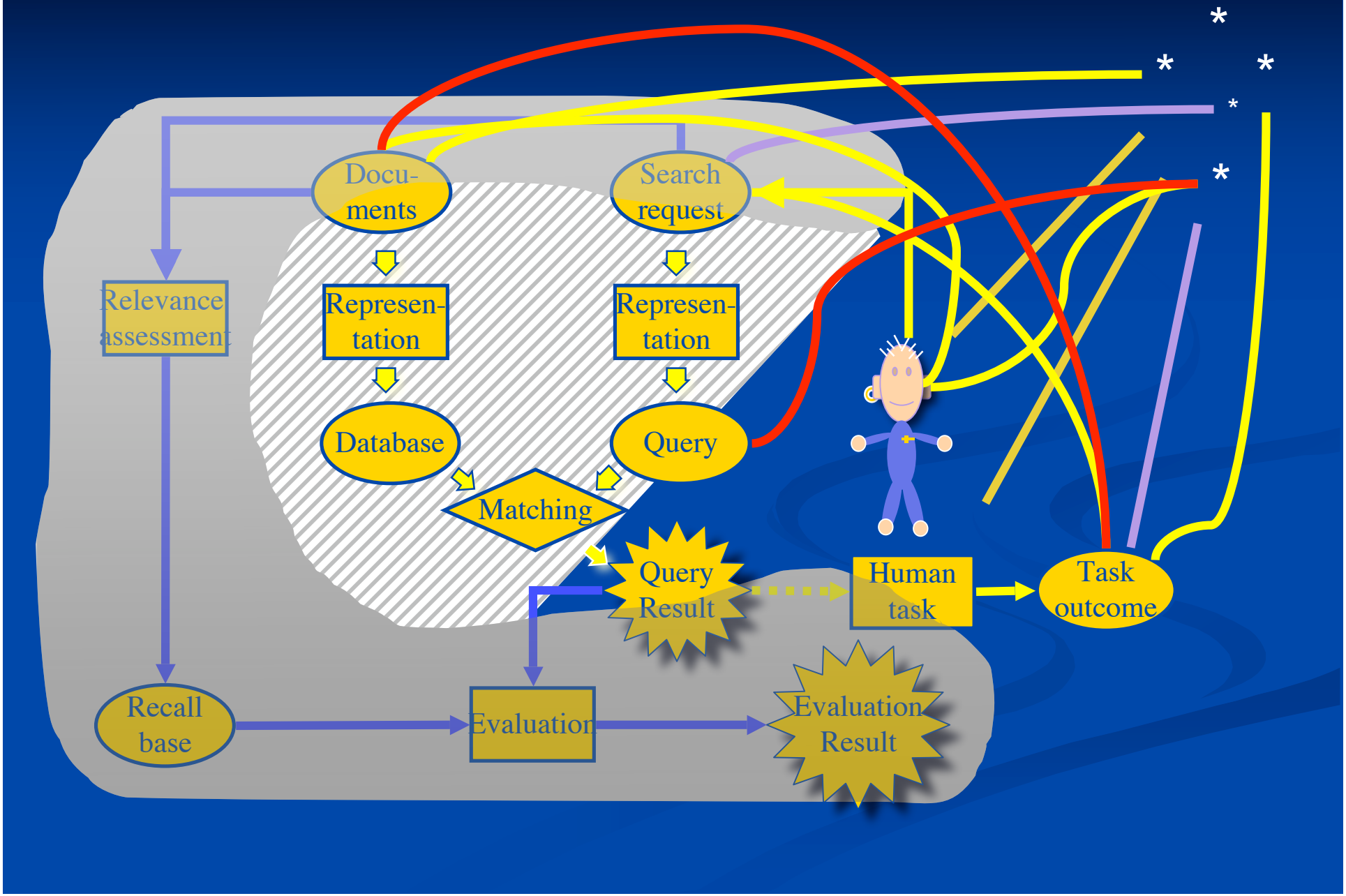
What about human performance?

- Allan, Carterette & Lewis (2005) searcher productivity in a passage-based QA task
- Turpin & Scholer (2006) studied user performance on simple web search tasks
- Smith & Kantor (2008) explored the relation between system performance and searcher behavior
- ... all insiders we can rely on

Exploring the outside...



... and all interactions ...



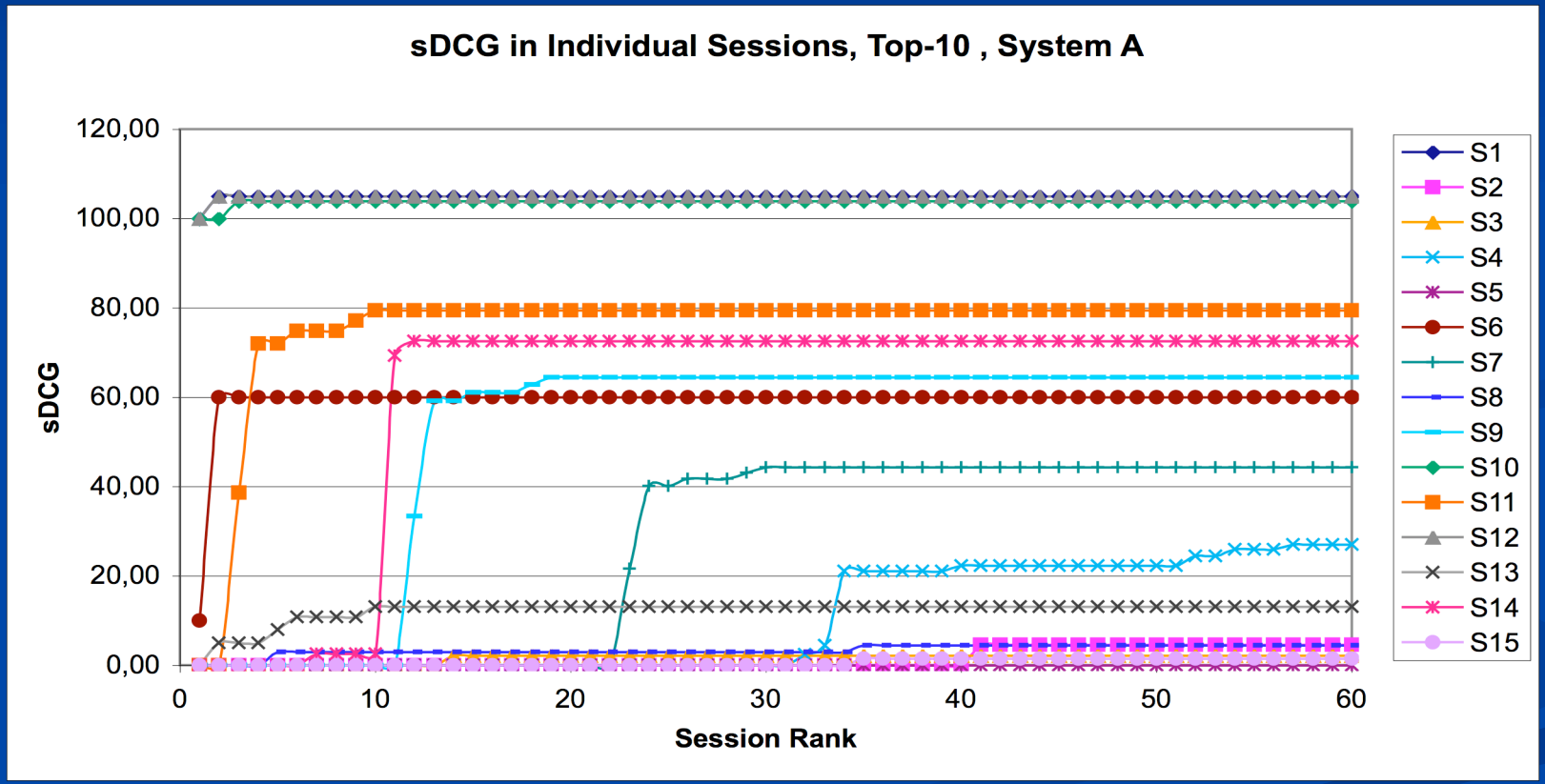
Single shot

- A hunter having one cartridge in his shotgun ...

Single shot vs. process

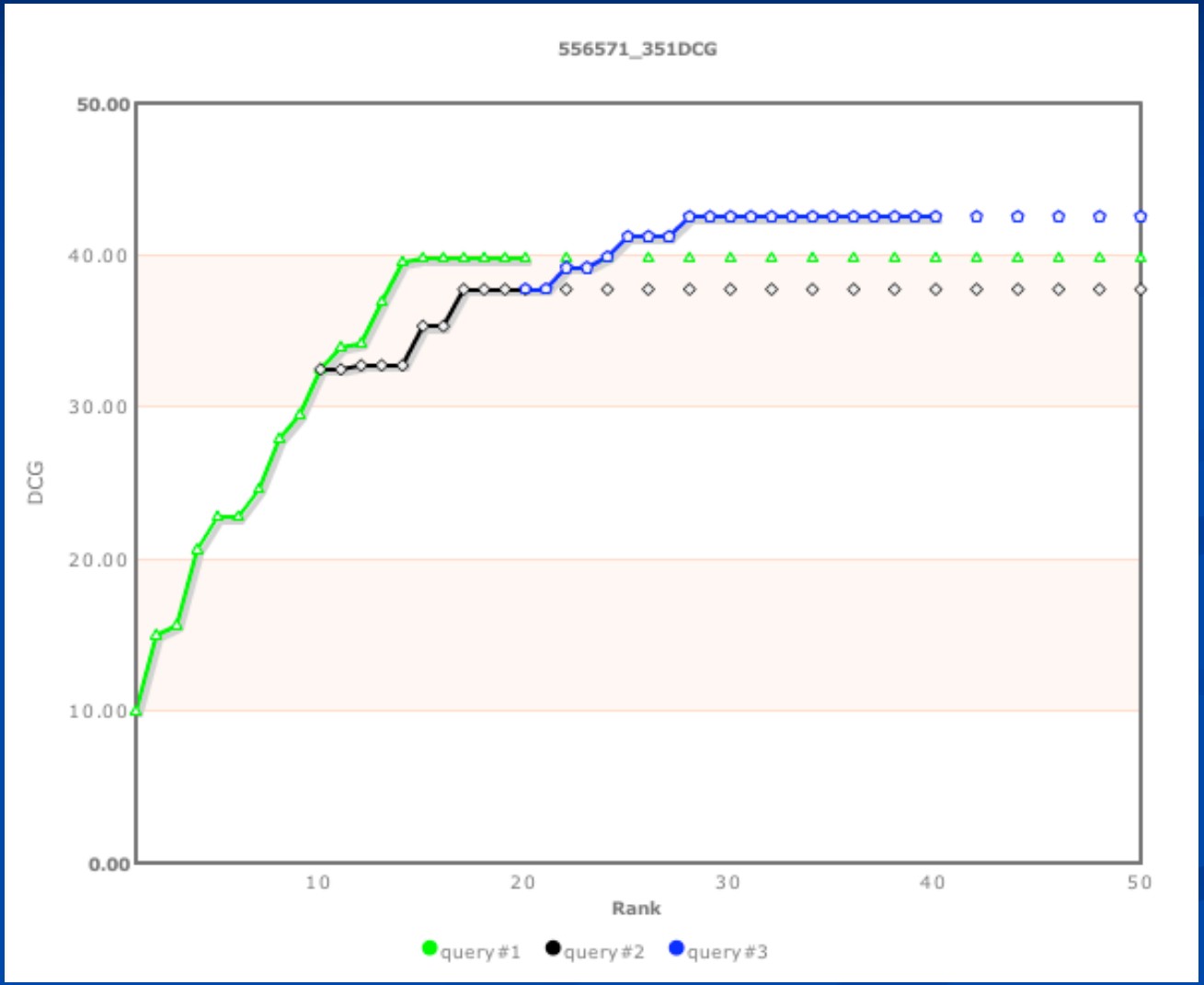
- Information retrieval processes have not been sufficiently described; therefore
  - they cannot be understood
  - they cannot be properly supported by IR techniques
  - they cannot be properly evaluated
- TREC IIR evaluation
- Session-based evaluation (Järvelin & al, 2008)

session-DCG, individual queries



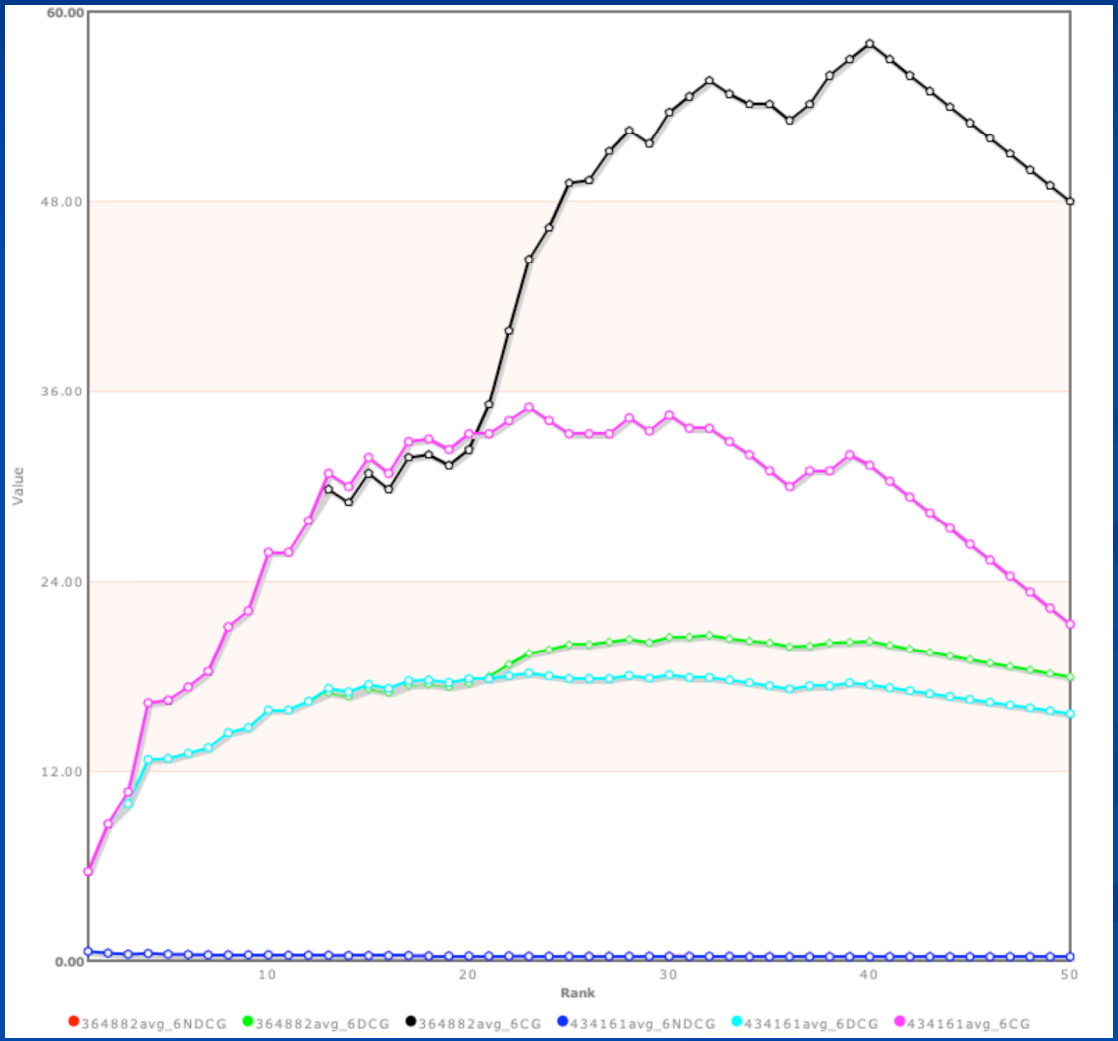
Individual sessions, Top-10, Systems A and B (b=2; bq=4; 0-1-10-100)

s(D)CG Feedback Simulation



2, 2, 0-1-5-10

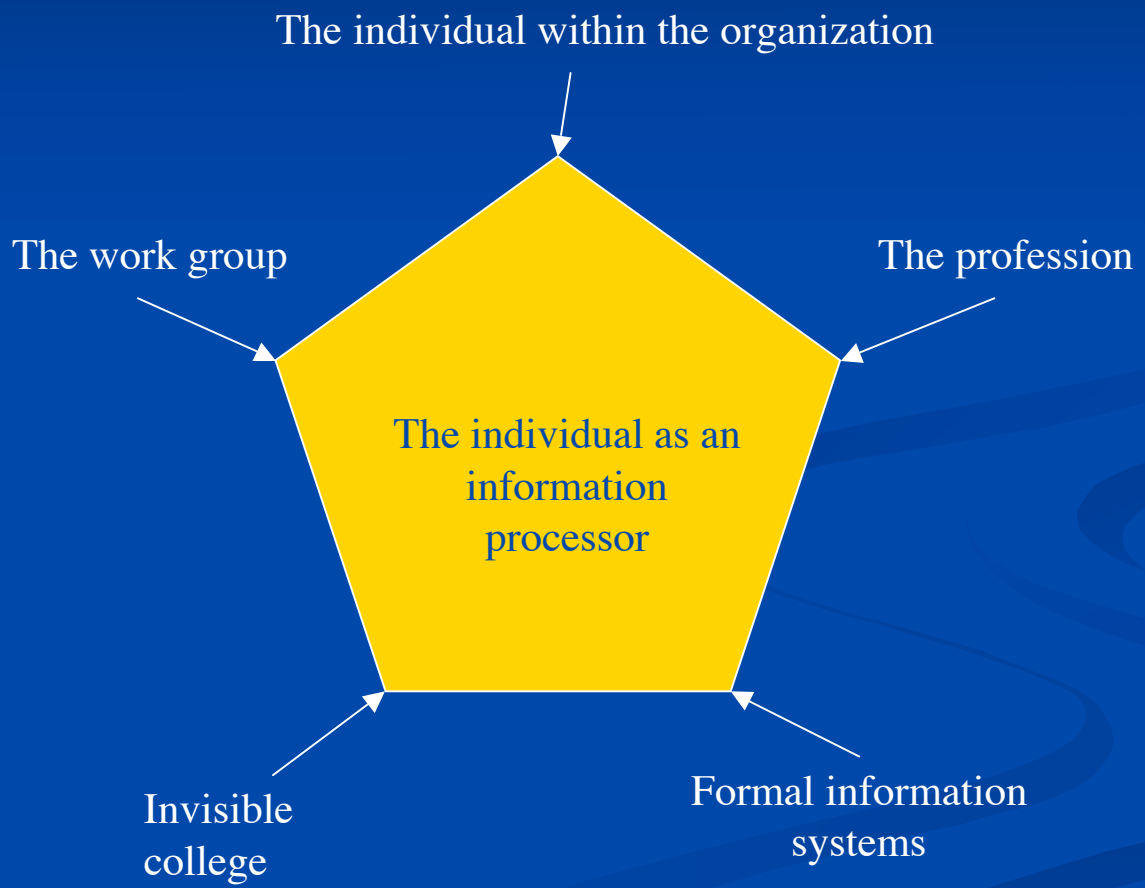
s(D)CG Feedback Simulation



IR in isolation?

- Information needs and seeking research in LIS in 1960's - 1980's - ARIST reviews
- Models vs. practice of research

Allen's IS Model



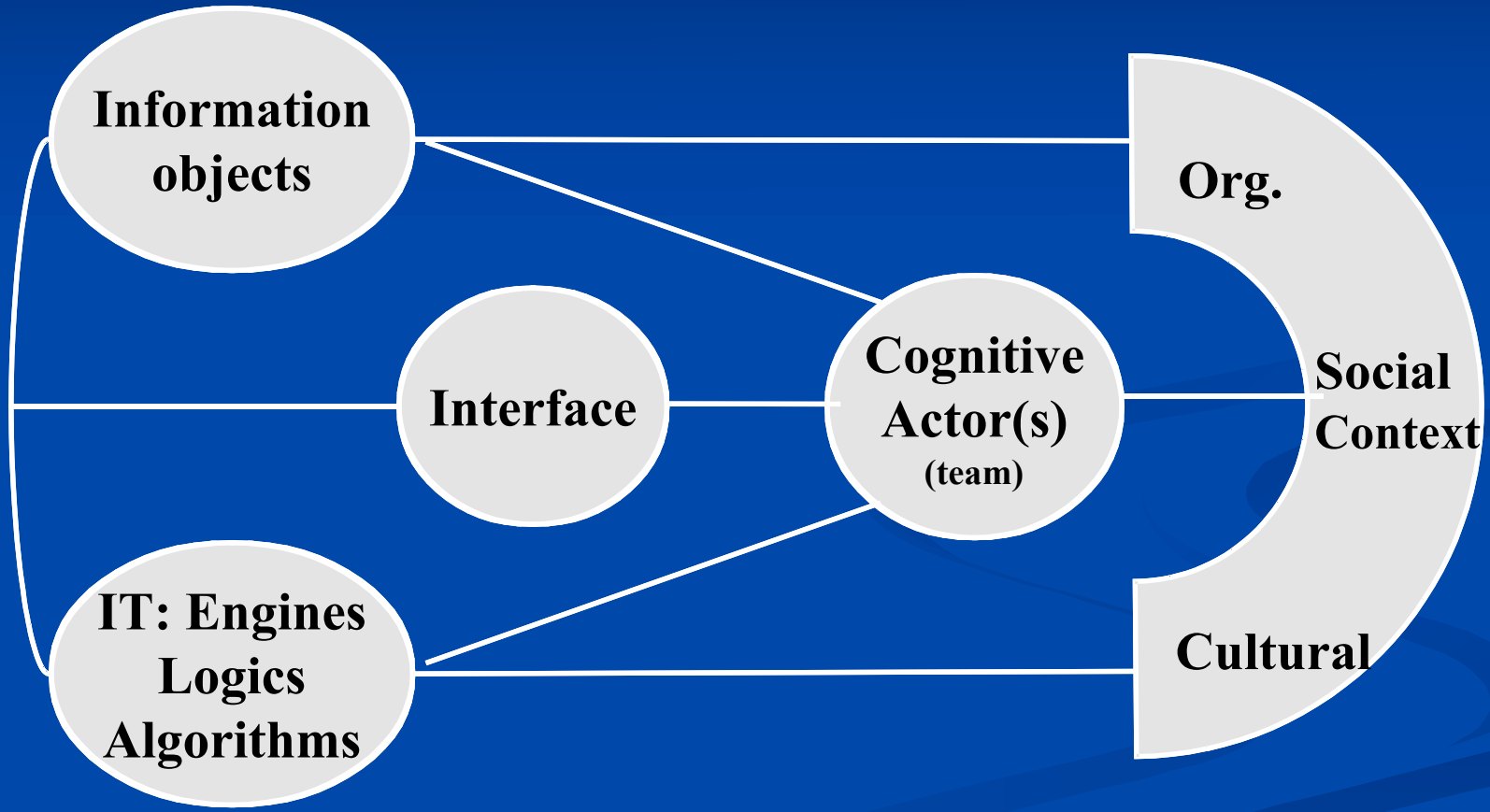
IR in isolation?

- IR rarely is the user's main task
  - maybe just a pain in the neck
- The information environment
- IR is not performed in isolation in practice
  - often multi-source, multi-tool information environment
  - often IR integrated into other tools

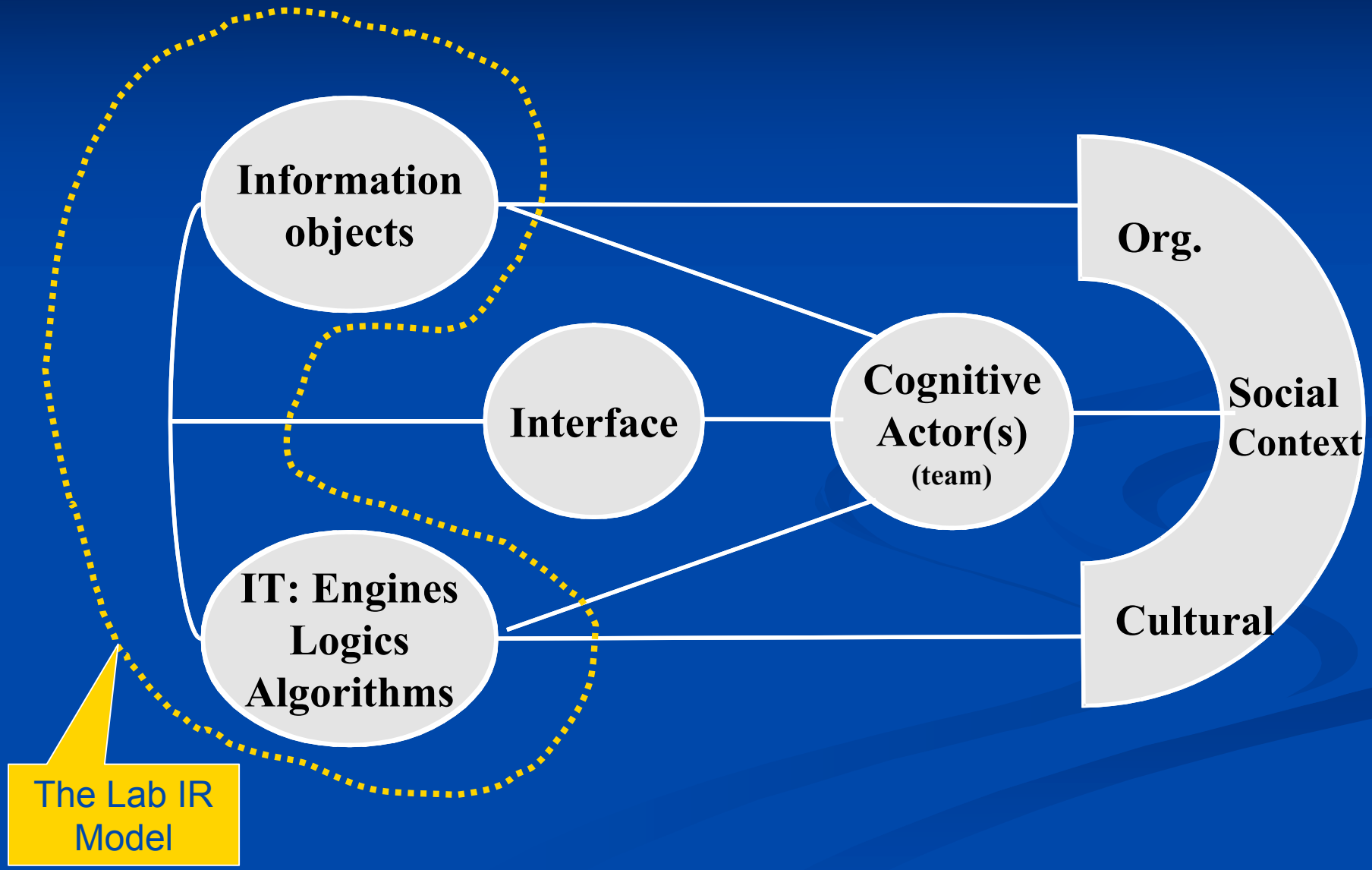
Agenda

- ✓ 1. Introduction
- ✓ 2. Some Concepts
- ✓ 3. The Laboratory Approach to IR
- ✓ 4. Mounting Critical Evidence
- ■ 5. Cognitive Framework for Research on IIR
- ■ 6. Concluding Discussion

A Cognitive Framework for IS&R



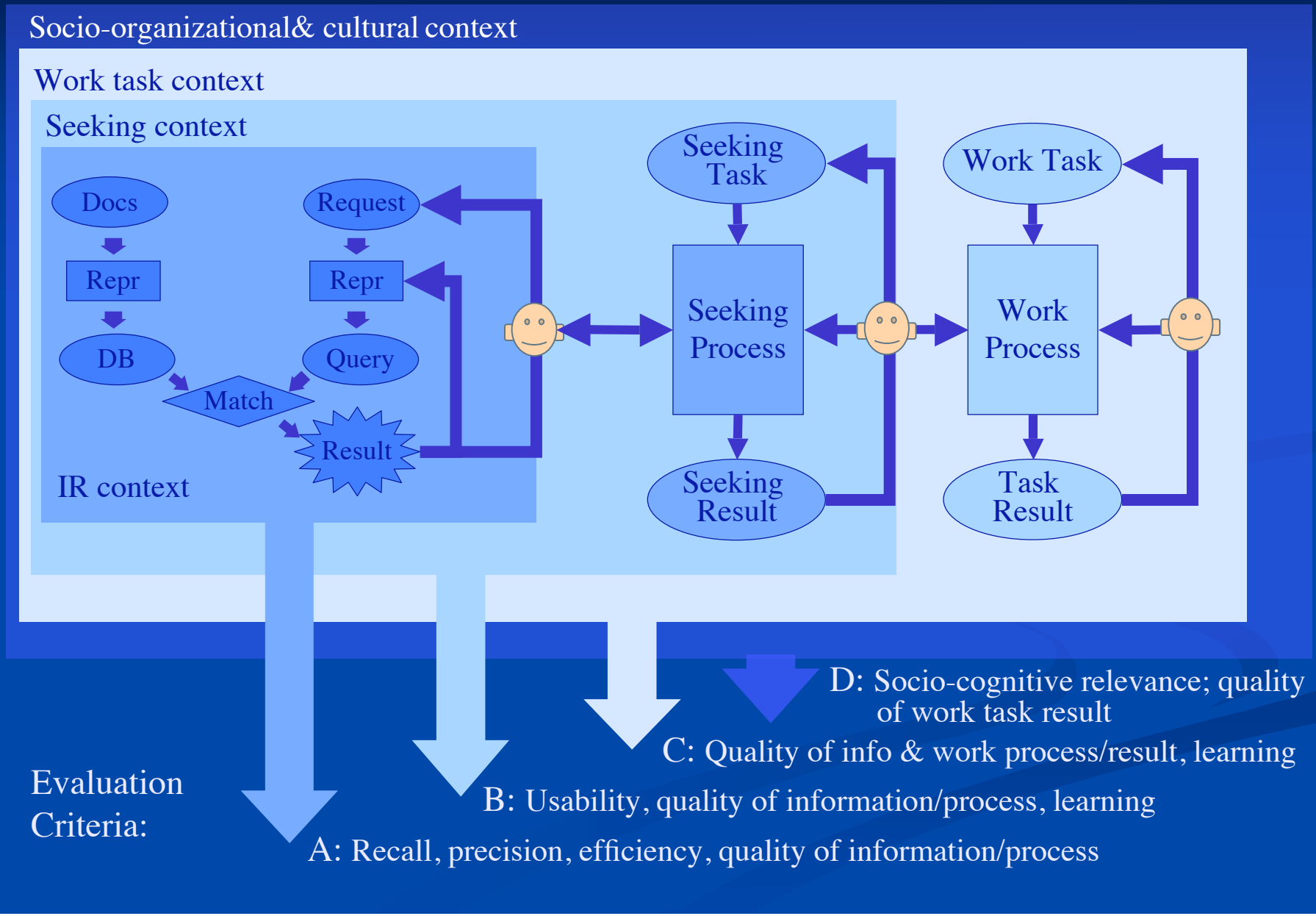
A Cognitive Framework for IS&R



The Applications of the Model

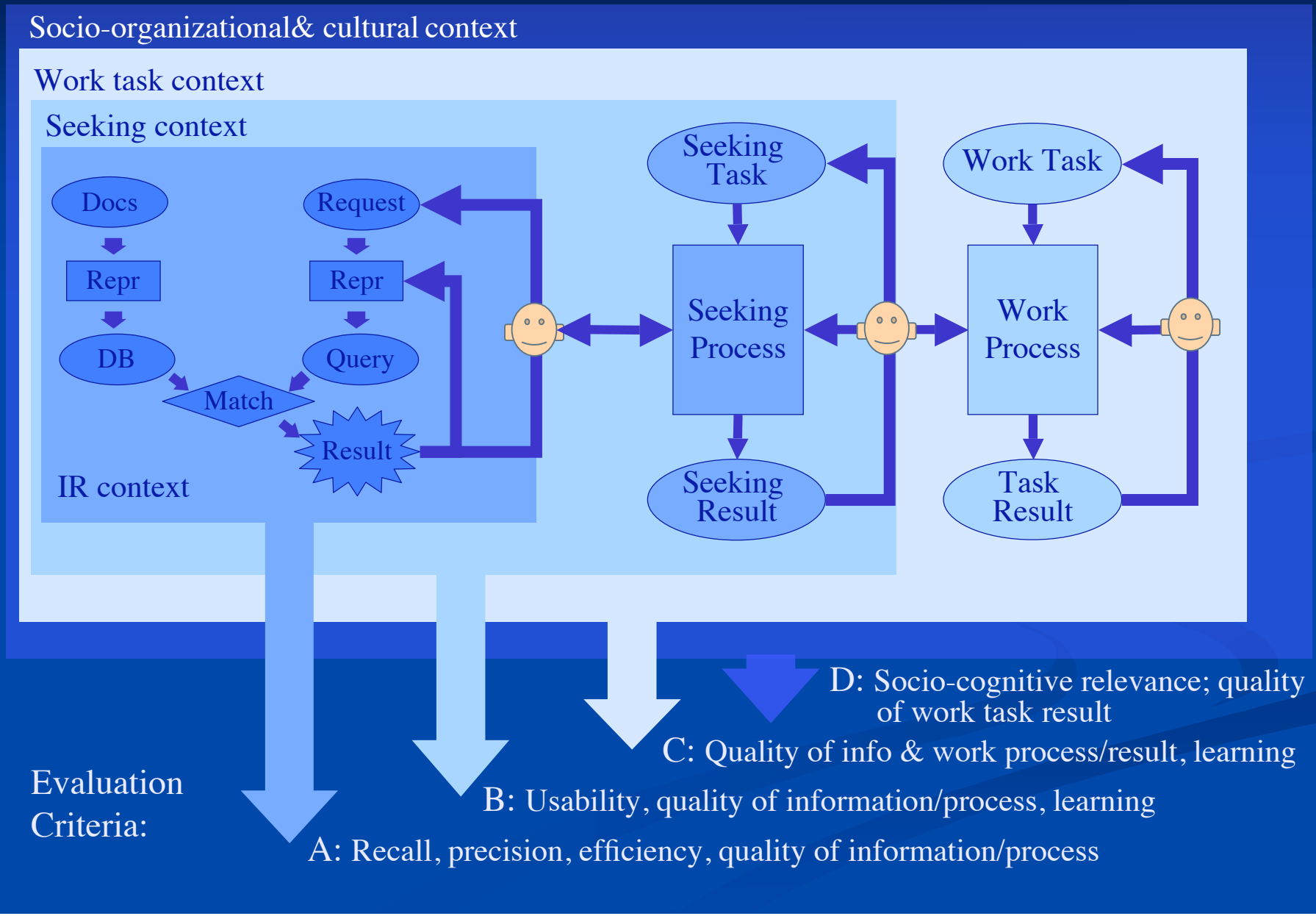
- Illustrating the roles of actors in a variety of cases of information behavior;
- Pointing to core components and information processes depending on (or influencing) such cases – i.e.,
- Pointing to kinds of context;
- Pointing out central variables involved in a variety of research designs – with a number of independent variables
- Pointing out new research questions and study designs

Cognitive Framework and Evaluation Criteria

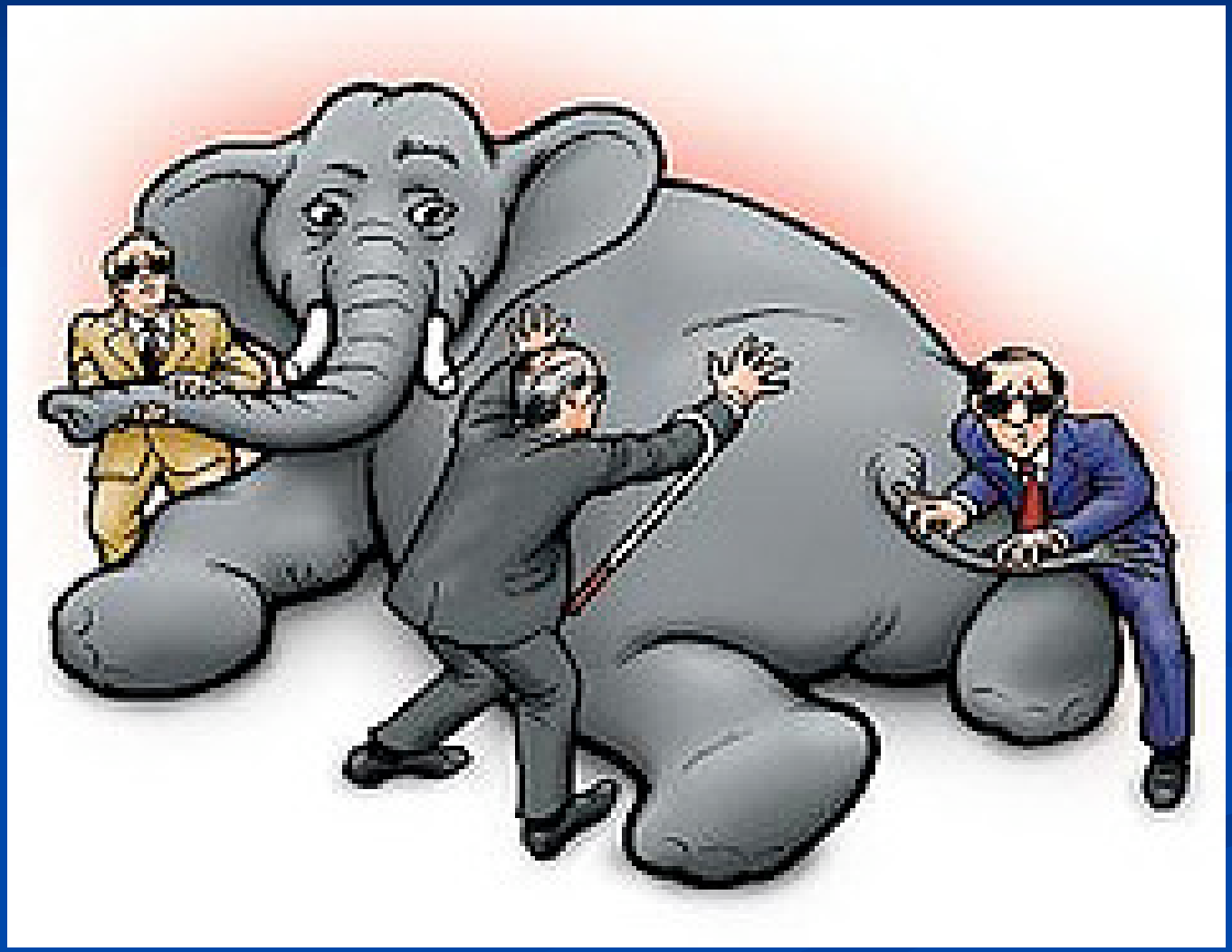


Lawn Mower?

Cognitive Framework and Evaluation Criteria



Developing understanding

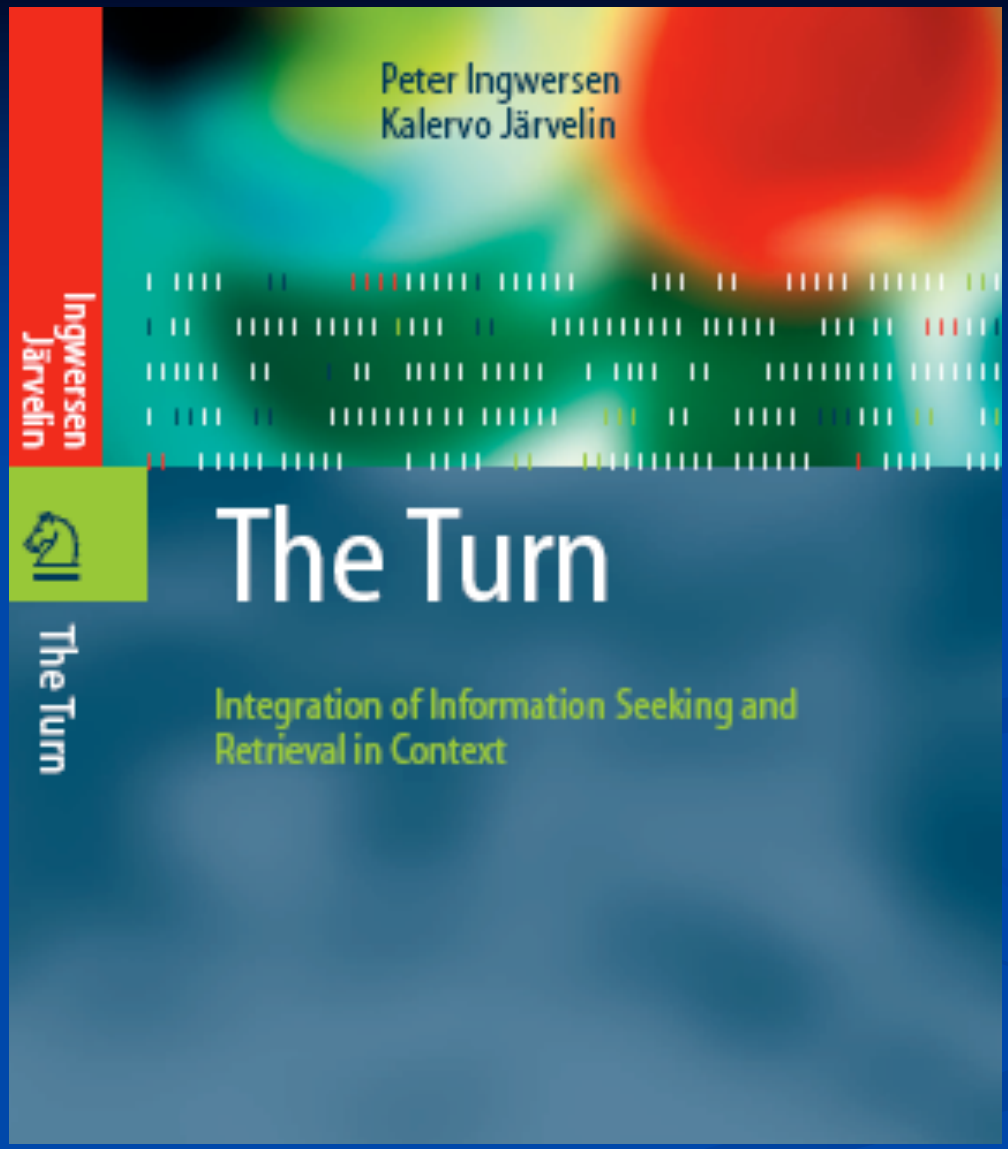


Agenda

- ✓ 1. Introduction
- ✓ 2. Some Concepts
- ✓ 3.The Laboratory Approach to IR
- ✓ 4. Mounting Critical Evidence
- ✓ 5. Cognitive Framework for Research on IIR
- ■ 6. Conclusions

The Cognitive Research Framework Suggests

- Analyses of retrieval and access in different types of collections
- Analyses of various actor types
- Analyses of various simulated task types for experimental control or real task types for understanding real situations
- Analyses of actor support in search processes



http://www.springeronline.com/**1-4020-3850-X**

Thank you!

Literature

Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research, 8*(3), paper no. 152 [available at: http://informationr.net/ir/8-3/paper152.html]. May 13, 2003.

Ellis, D. (1996). *Progress and problems in information retrieval*. London, UK: Library Association Publishing.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50. (Available: http://www.db.dk/pi)

Ingwersen, P. & Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context*. Dortrecht, NL: Springer. ISBN 1-4020-3850-X

Järvelin, K. (2007). An Analysis of Two Approaches in Information Retrieval: From Frameworks to Study Designs. *Journal of the American Society for Information Science and Technology* 58(7): 971-986. (Available: http://www.info.uta.fi/tutkimus/fire/archive/2007/KJ-IRtheory'06.pdf

Järvelin, K. & Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information Research, 10(1), paper 212*. (Available: http://informationr.net/ir/10-1/paper212.html

Robertson, S. E. (2000). On theoretical argument in information retrieval. SIGIR Forum, 34(1), 1-10. (Available: http://www.acm.org/sigs/sigir/forum/S2000/salton_lecture.pdf

Saracevic, T, (1997). Users lost: Reflections of the past, future and limits of information science. ACM Sigir Forum 31(2), 16-27.