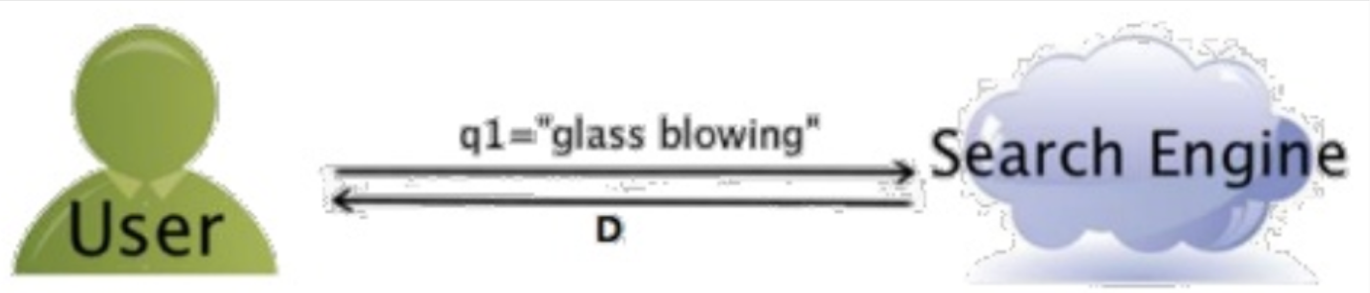
Interactive Information Retrieval

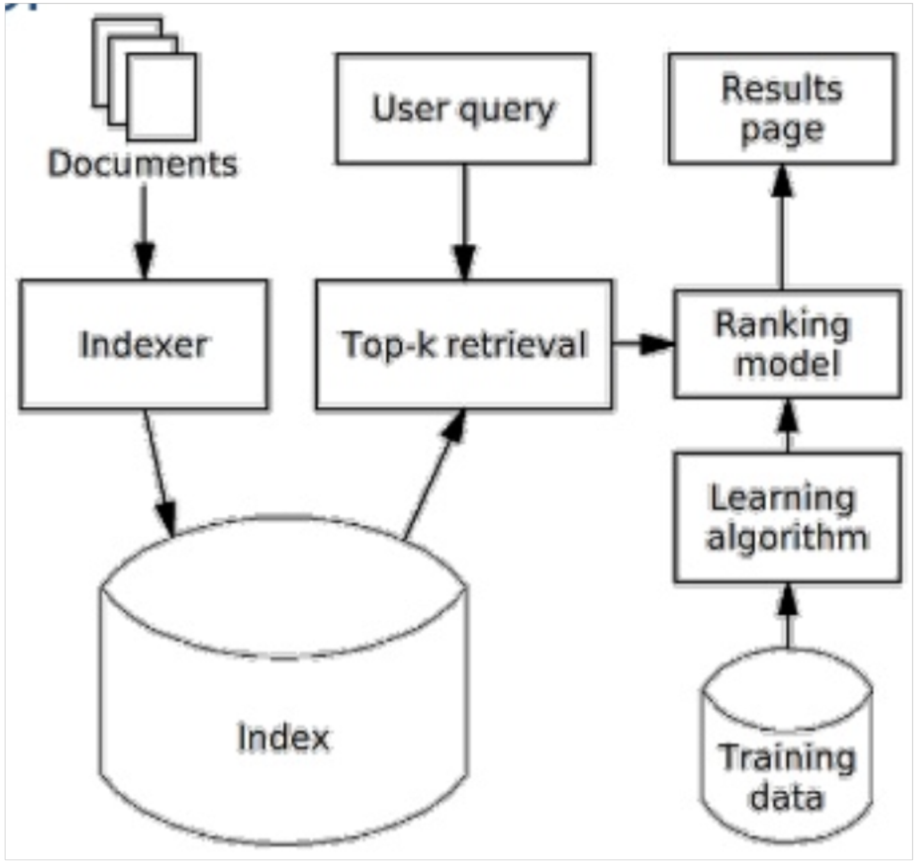
Dorota Glowack

glowacka@cs.helsinki.fi

Static Information Retrieval

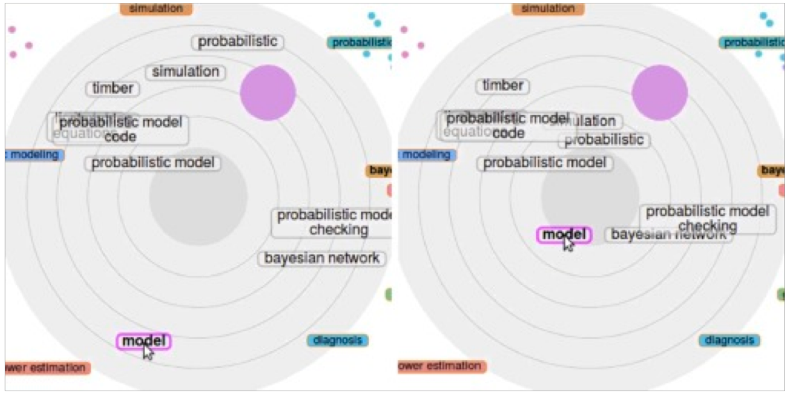
- System does not learn directly from the user
- Parameters updated periodically



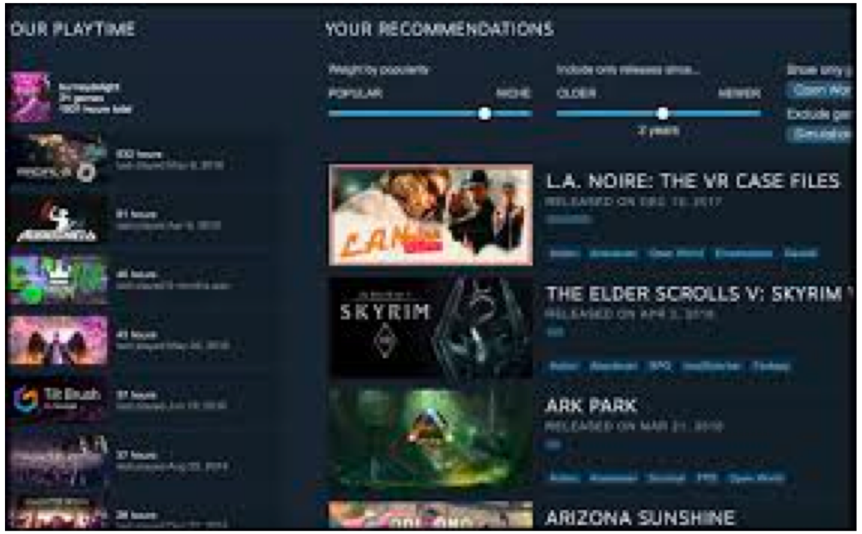


- Aim is to maximize precision and recall
- User's information need is static

Interactive Recommendation

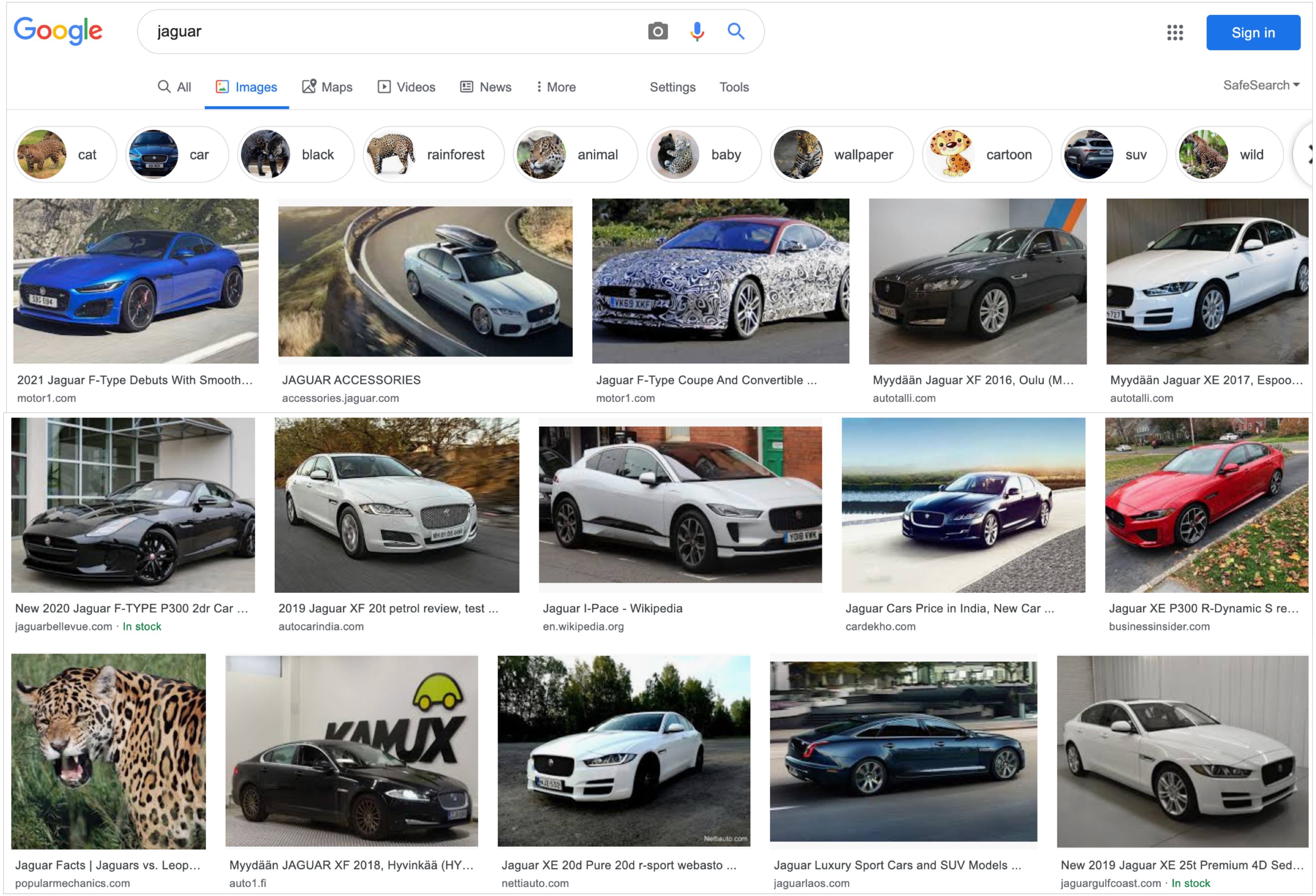


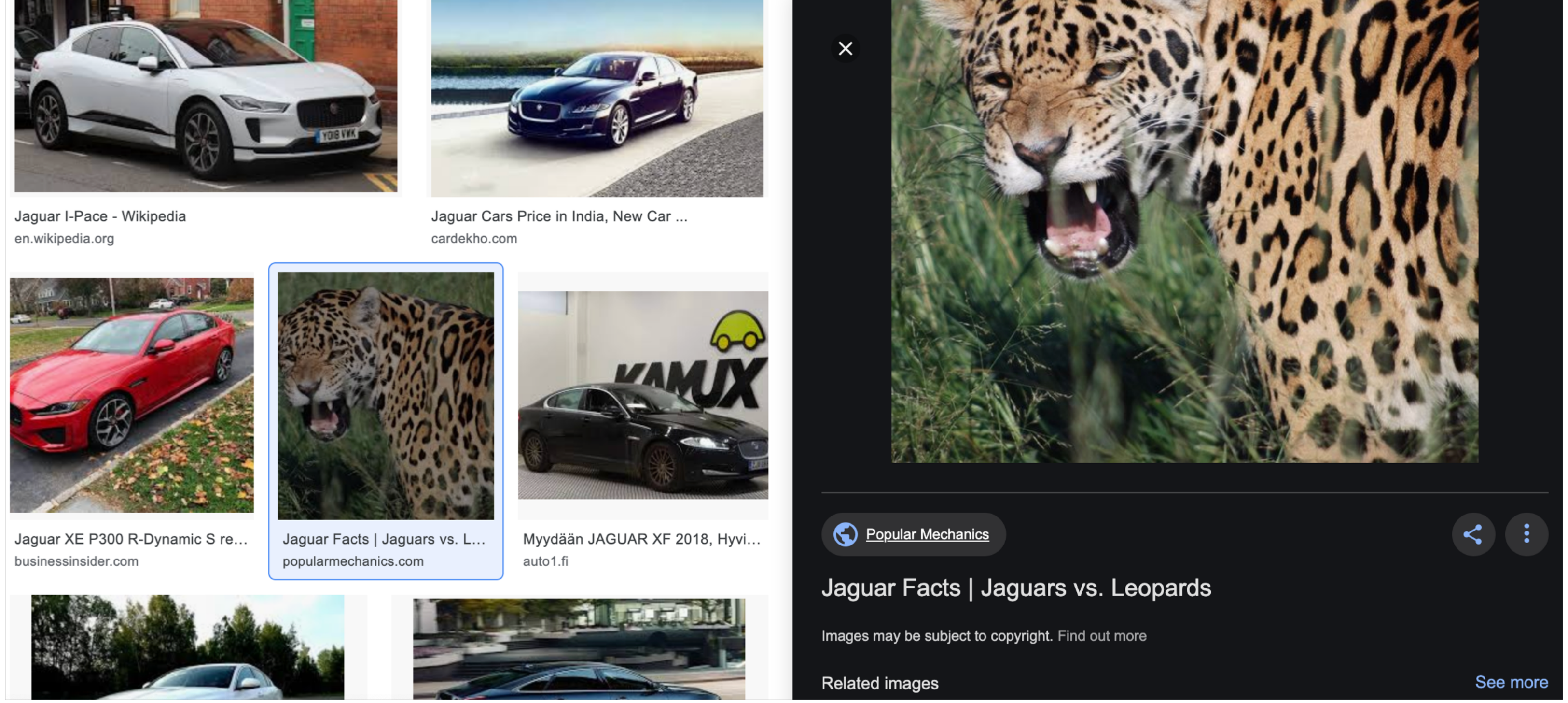
SciNet scientific literature recommendation

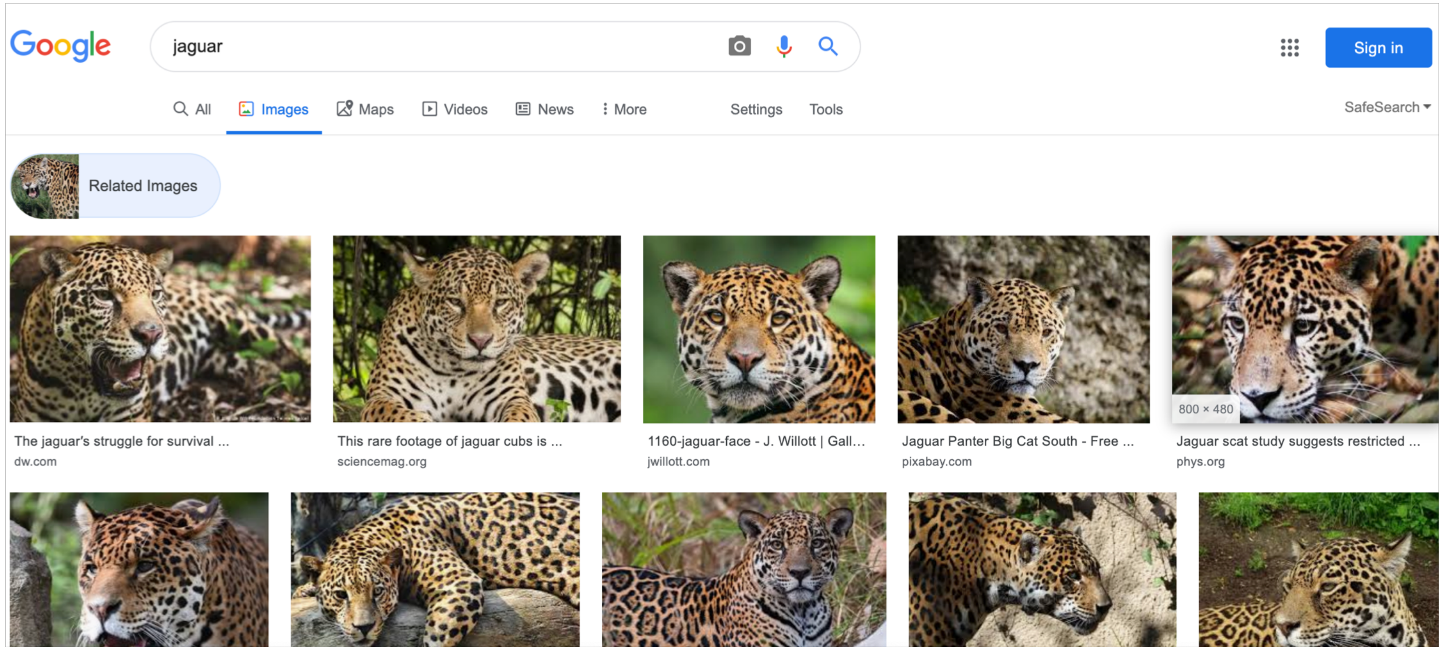


Steam Interactive Games Recommendation

- System learns from user feedback in an interactive setting (while the user is engaged with the system)
- Recommendations improve gradually within the interaction session (and beyond)


Google
jaguar
Sign in
All
Images
Maps
Videos
News
More
Settings
Tools
SafeSearch
cat
car
black
rainforest
animal
baby
wallpaper
cartoon
suv
wild
2021 Jaguar F-Type Debuts With Smooth...
motor1.com
JAGUAR ACCESSORIES
accessories.jaguar.com
Jaguar F-Type Coupe And Convertible ...
motor1.com
Myydään Jaguar XF 2016, Oulu (M...
autotalli.com
Myydään Jaguar XE 2017, Espoo...
autotalli.com
New 2020 Jaguar F-TYPE P300 2dr Car ...
jaguarbellevue.com · In stock
2019 Jaguar XF 20t petrol review, test ...
autocarindia.com
Jaguar I-Pace - Wikipedia
en.wikipedia.org
Jaguar Cars Price in India, New Car ...
cardekho.com
Jaguar XE P300 R-Dynamic S re...
businessinsider.com
Jaguar Facts | Jaguars vs. Leop...
popularmechanics.com
Myydään JAGUAR XF 2018, Hyvinkää (HY...
auto1.fi
Jaguar XE 20d Pure 20d r-sport webasto ...
nettiauto.com
Jaguar Luxury Sport Cars and SUV Models ...
jaguarlaos.com
New 2019 Jaguar XE 25t Premium 4D Sed...
jaguargulfcoast.com · In stock


Jaguar I-Pace - Wikipedia
en.wikipedia.org
Jaguar Cars Price in India, New Car ...
cardekho.com
Jaguar XE P300 R-Dynamic S re...
businessinsider.com
Jaguar Facts | Jaguars vs. L...
popularmechanics.com
Myydään JAGUAR XF 2018, Hyvi...
auto1.fi
Popular Mechanics
Jaguar Facts | Jaguars vs. Leopards
Images may be subject to copyright. Find out more
Related images
See more


Google
jaguar
Sign in
All
Images
Maps
Videos
News
More
Settings
Tools
SafeSearch
Related Images
The jaguar's struggle for survival ...
dw.com
This rare footage of jaguar cubs is ...
sciencemag.org
1160-jaguar-face - J. Willott | Gall...
jwillott.com
Jaguar Panter Big Cat South - Free ...
pixabay.com
800 × 480
Jaguar scat study suggests restricted ...
phys.org

Human centred approach to Information Retrieval

- Cognitive viewpoint of information retrieval:
  - more focus on user needs and how they search for information
  - matching user side and interaction
- Interactive IR
  - more than simply developing interfaces for searching (Shneiderman, Byrd, & Croft, 1998)
  - knowledge of people's search behavior and search context, including the environmental factors that influence behavior (Fidel & Pejtersen, 2004)
- HCIR (Human Computer Information Retrieval)
  - HCI and IR come from different traditions:
    - HCI places more emphasis on usability
    - IR emphasizes system effectiveness

Interactive Information Retrieval

- Goal: study ***user interaction*** with a search system to learn about the user's ***search intent*** and when they encounter relevant documents
- Taking account of user and their **search context** can improve understanding of the ***search process*** and the user's intent
- A search system that ”knows” this information can improve its performance in retrieving documents that ***satisfy user's needs***
- Awareness of demands imposed on user's ***cognitive processing*** and ***levels of user's knowledge*** can contribute to improvements in system performance.

Interactive Information Retrieval Problems

- IIR research addresses three major problem areas:
  1. Understanding information seeking needs and behaviors;
  2. Developing retrieval systems that respond to information needs and support information seeking behaviors and interactions;
  3. Developing methods and measures to study and evaluate behaviors, interactions and systems.
- Issues
  - Information seeking behavior related information needs and query intent
    - Models of the Information Seeking Process
  - Design of Search User Interfaces and presentation of search results
  - How to evaluate IR systems / Search Quality

Information Seeking Models

- Represent how people search for information in specific environments and how they interact with IRs and/or traditional sources to satisfy information needs
- Models vary based on what researchers investigate:
  - type of user, e.g. novices vs. advanced researchers
  - search environment, e.g. web search vs. online library
  - types of documents/information, e.g. specilised vs. non-specilised literature
  - user’s search goal, e.g. looking for a specific documents vs. general browsing
- Commonality across user information seeking

Models of Information Seeking Behaviour

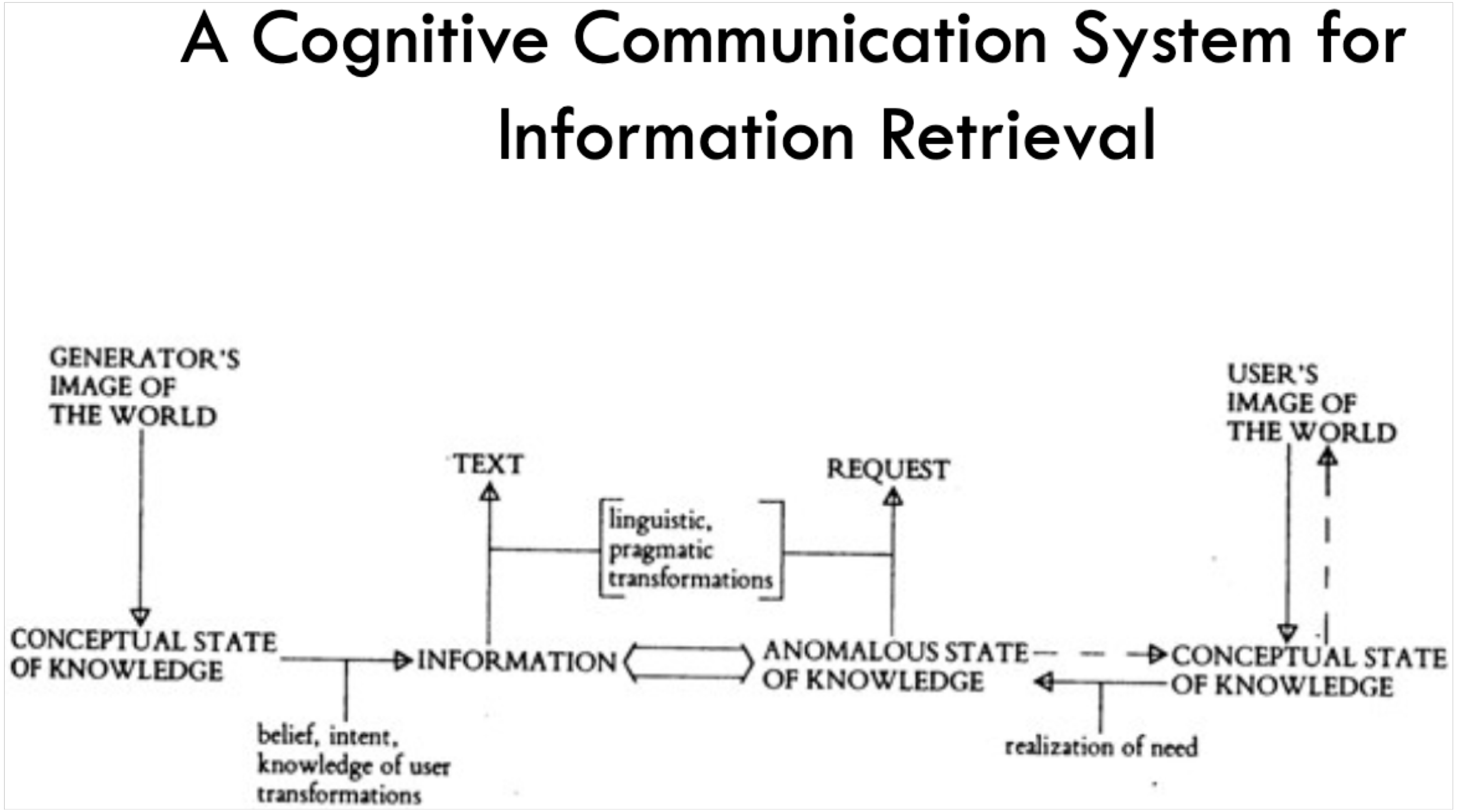
- Belkin’s Anomolous State of Knowledge (ASK)
- Bates’ berrypicking – acts in searching
- Dervin’s sense-making theory – gap, bridge
- Ellis’ Information Seeking Process
- Marchionin’s Information Seeking Model
- Kuhlthau’s information search process
- Ingwersen’s cognitive model
- Wilson’s information-seeking behaviour model
- Saracevic’s model of stratified interaction

Belkin’s Model (1984)

- User’s information need: Anomalous State of Knowledge (ASK)
- Knowledge gap (anomaly) and the need to solve it
- Difficult for user to specify information needs
- Interview to elicit problem statements to determine the user’s ASK
- After ASK determined, formulate query in system’s language
- ASK Definition:

  A recognition by an individual that his/her model of some aspect of the external world and of her/her position in it with respect to some particular situation is insufficient and knowledge is needed to reduce uncertainty

Belkin’s Model



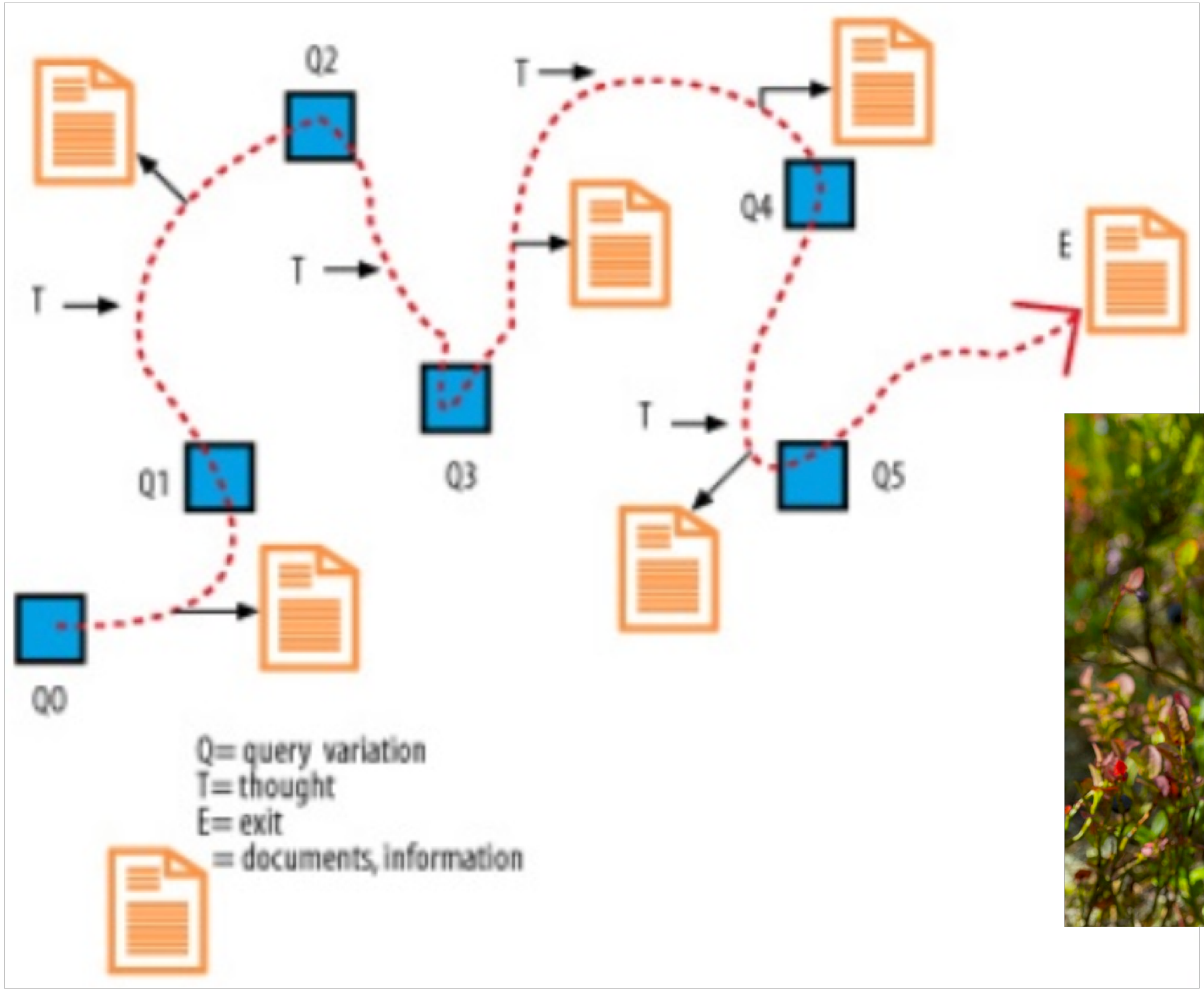
Contributions of ASK

- Reinforced the certainty of the user's needs

- Recognized the iterative nature of information retrieval
  - users return to the IR system repeatedly to satisfy their information needs

- Move towards system design that is user- rather than system-centered (people rather than documents)

Bates' Berrypicking – dynamic not static



Berrypicking model

- Static IR model:
  - The information need remains the same throughout the search session
  - Goal is to produce a perfect set of relevant documents with respect to the query
- Berrypicking model:
  - The query is continually shifting
  - Users may move through a variety of sources
  - New information may yield new ideas and new directions
  - The value of search is in the bits and pieces picked up along the way

Dervin's sense-making theory

SITUATIONS: The time-space contexts at which sense is constructed.

GAPS: The gaps seen as needing bridging, translated in most studies as "information needs" or the questions people have as they construct sense and move through time-space.

USES: The uses to which the individual puts newly created sense, translated in most studies as information helps and hurts.

questioning that can reveal the nature of a problematic situation, the extent to which information serves to bridge the gap of uncertainty, confusion, or whatever, and the nature of the outcomes from the use of information.

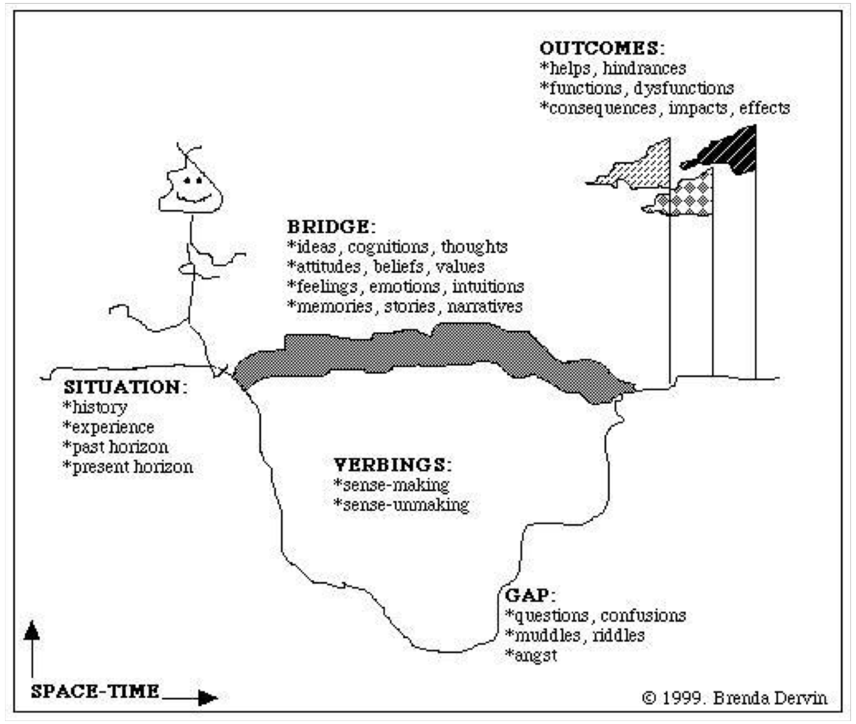
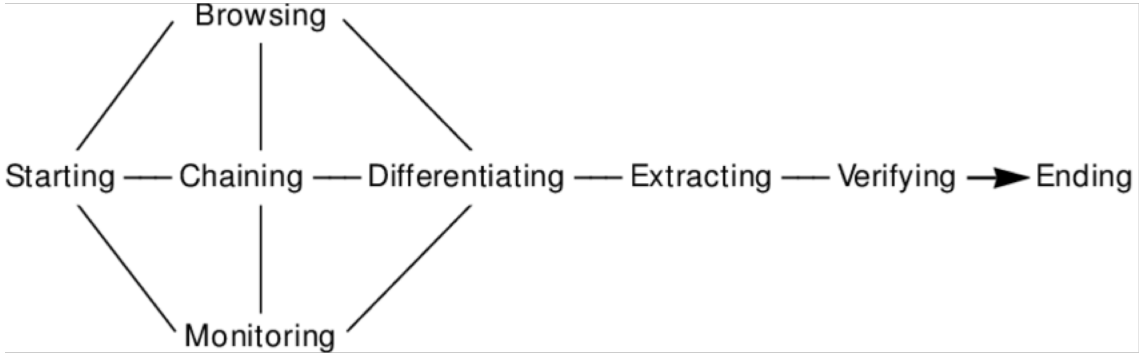


Figure 3: Dervin's 'sense-making'

Ellis' Behavioural Model

David Ellis describes 8 information seeking patterns of social scientists, physical scientists, and engineers in using hypertext (e.g., the Web).



- Starting (Surveying)
- Chaining
- Monitoring
- Browsing
- Differentiating (Distinguishing)
- Filtering
- Extracting
- Verifying
- Ending

Ellis' Behavioural Model (1989)

- ***Starting***: Looking for information in a new area on on a new topic.
- ***Chaining***: Searching by following citation connections between materials.
- ***Differentiating***: Selecting information sources based on their orientation and intended audience.
- ***Monitoring***: The continuous monitoring of developments in a field of study.
- ***Extracting***: Going through a particular source selectively identifying relevant material from that source

Marchionini’s Model

- Problem solving approach to understanding information seeking process in the electronic environment
- Eight processes that may work in parallel:
  - Problem recognition
  - Problem definition
  - Selection of system/source
  - Problem articulation (query formulation)
  - Search execution
  - Examination of results
  - Extraction of desired information
  - Reflection, Iteration, and Stopping of search process

Kuhlthau's Model (1991)

- Six stages in the information search process incorporated in three realms:
  - affective – feelings
  - cognitive - thoughts
  - physical - actions

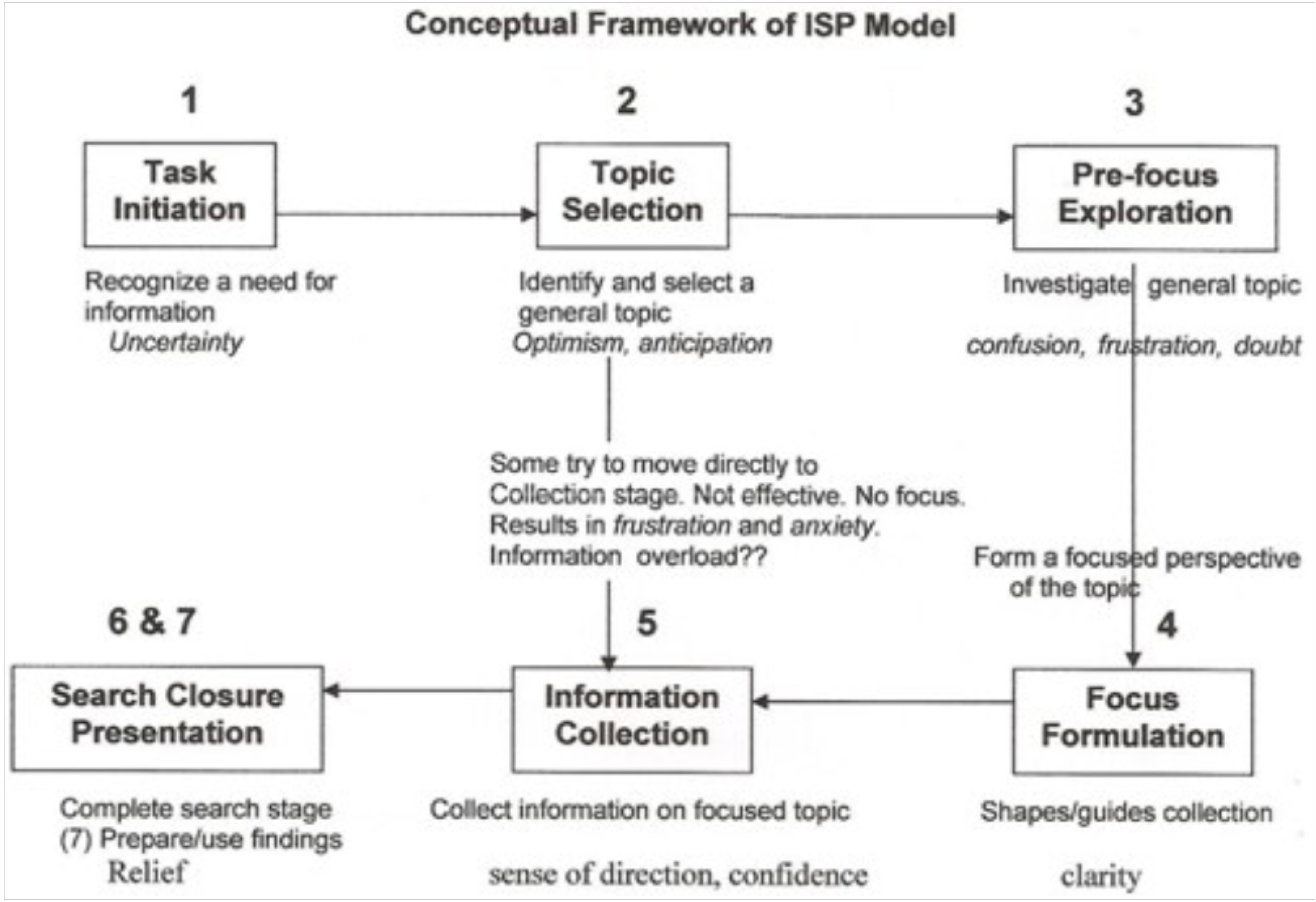
- **initiation** - the first awareness of a lack of knowledge or understanding
- **selection** - identifying the general topic of the approach to be pursued
- **exploration** - investigating information on the general topic, to improve orientation sufficiently to form a focus for resolving the problem
- **formulation** - forming a focus for the information encountered
- **collection** - extends and supports the focus and selects information pertaining specifically to the focus
- **presentation** - completing the search and preparing to present or otherwise use the findings

Kuklthau's Information Search Process

Tasks	Initiation	Selection	Exploration	Formulation	Collection	Presentation
Feelings (affective)	uncertainty	optimism	confusion/ frustration/doubt	clarity	sense of direction /confidence	satisfaction or disappointment
Thoughts (cognitive)	vague →		→	focused →	increased interest →	→
Actions (physical)	seeking relevant information exploring →		→	→	seeking pertinent information documenting	

Figure 5.1. Model of the Information Search Process (ISP).

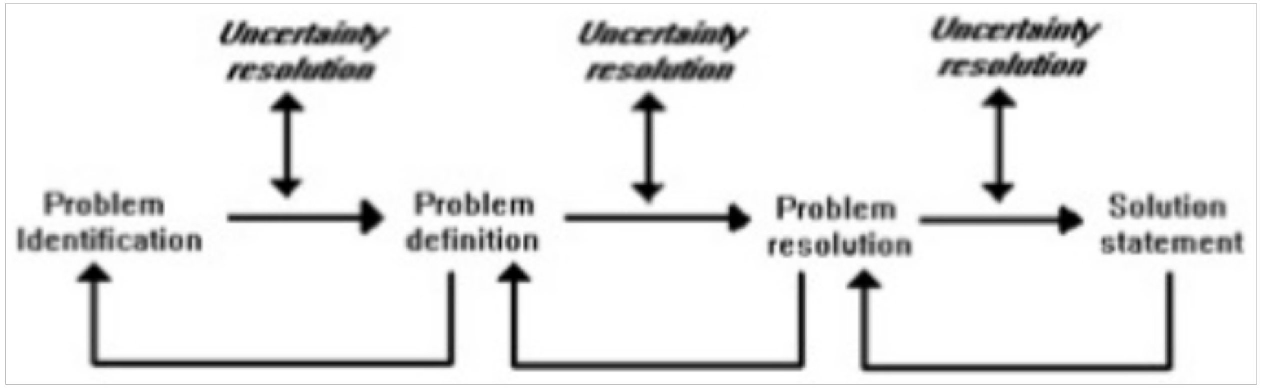
Kuhlthau's Information Search Process



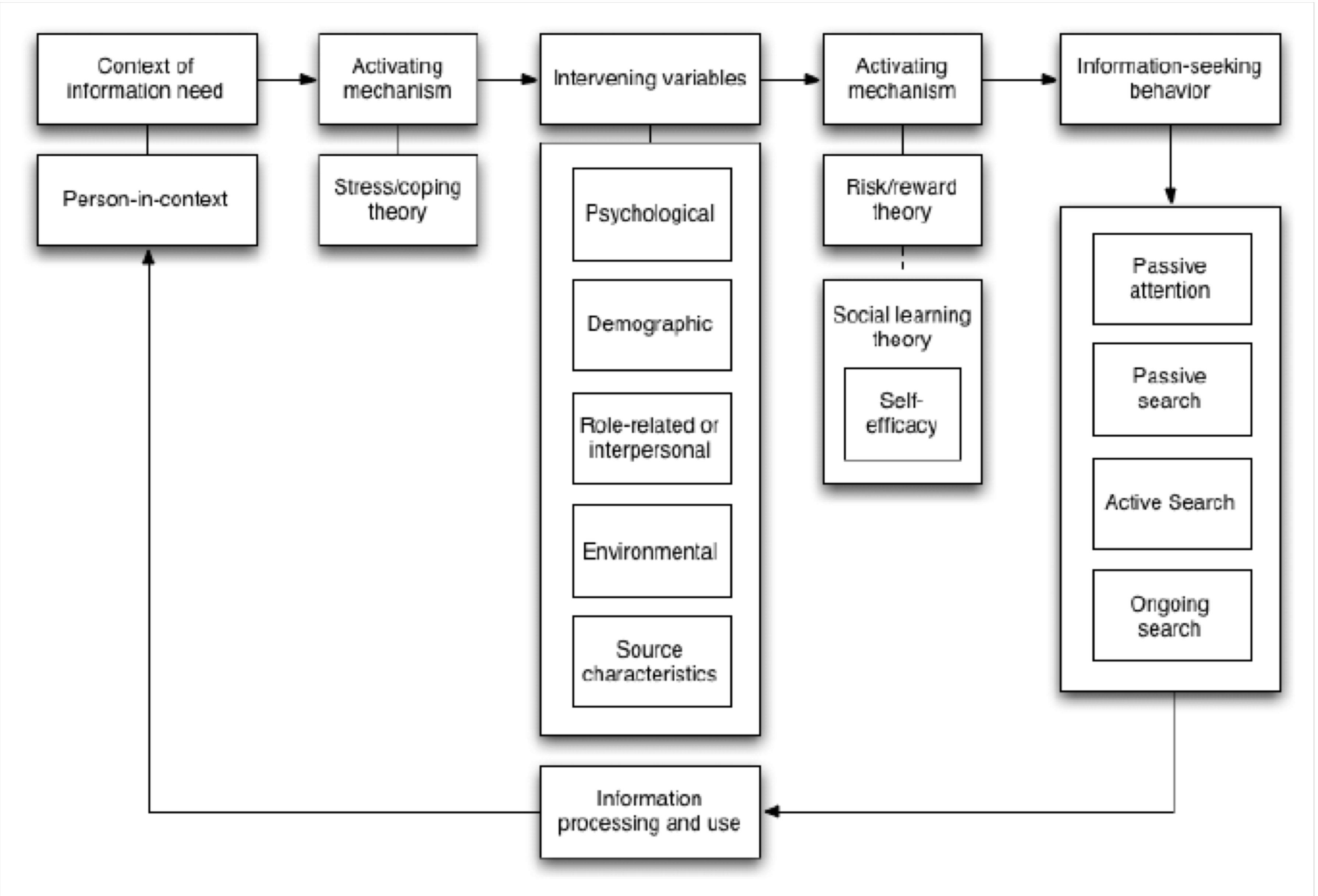
Wilson's Problem Solving Model

- Goal-directed towards problem solving
- Based on a survey of research in the health field
- Users move from *uncertainty* to *certainty* through the problem-resolution process

- Stages:
  - Problem identification
  - Problem definition
  - Problem resolution
  - Solution statement

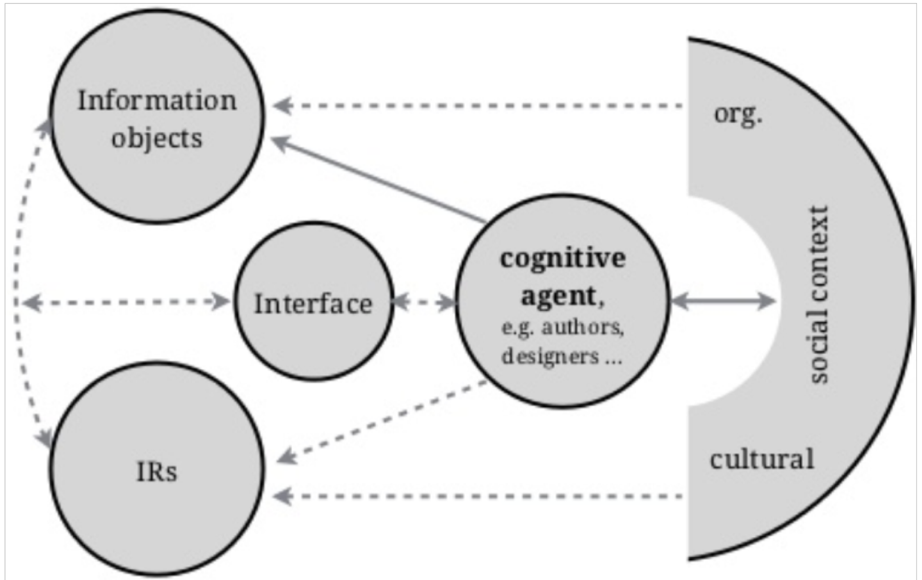


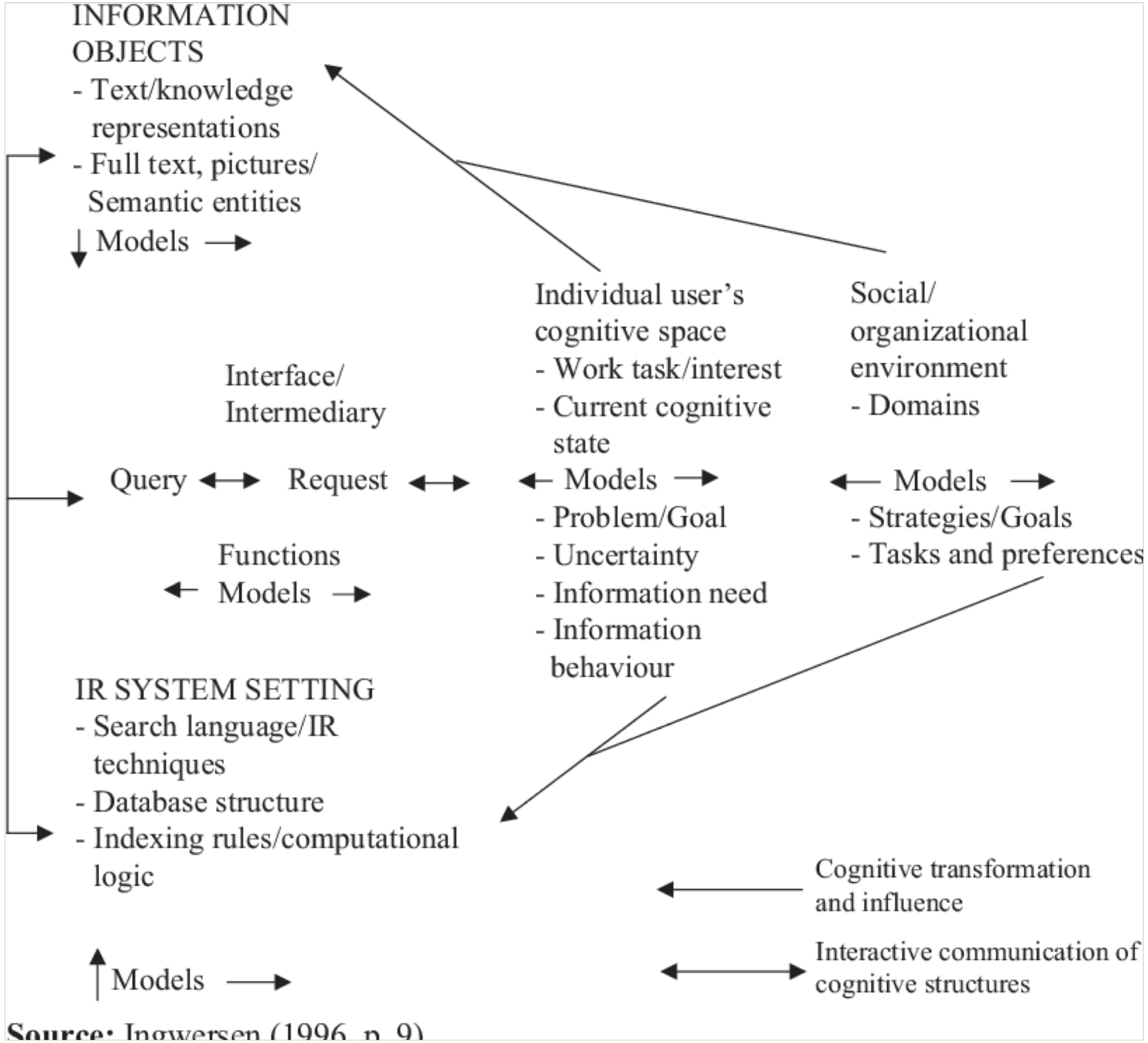
Wilson's General Model of Information Behaviour (1997)



Ingwersen Cognitive IR Model (1996)

- Ingwersen analyzes cognitive information retrieval.
- He focuses on the interaction of mental models.
- He shows the short and the long-term change of these models as well as their drivers.





Saracevic’s model of Stratified Interaction

- Stratified interaction model developed within an overall framework of an **acquisition-cognition-application model** of information use.
- The levels or strata are simplified to three:

1. **surface**, or the level of interaction between the user and the system interface;
2. **cognition**, or the level of interaction with the texts or their representation
3. **situation**, or the **context** that provides the initial problem at hand.

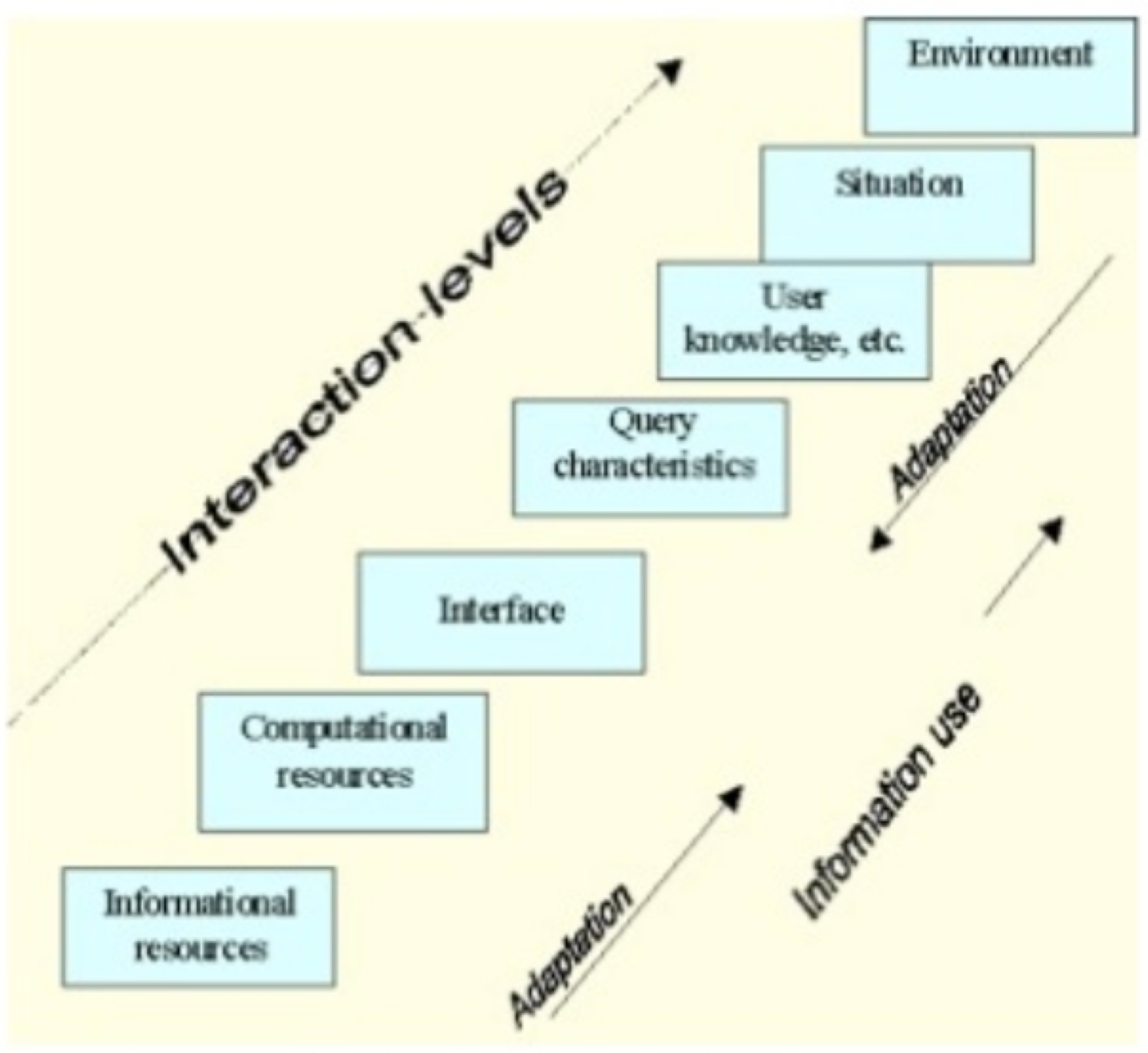


Figure 9: Saracevic's model of stratified

Information Foraging Theory (IFT)

- Information Foraging Theory (IFT)
  - Pirolli and Card – Xerox PARC
  - “an approach to the analysis of human activities involving information access technologies”
  - Derives from **optimal foraging theory** in biology and anthropology
  - Analyzes adaptive value of food-foraging strategies

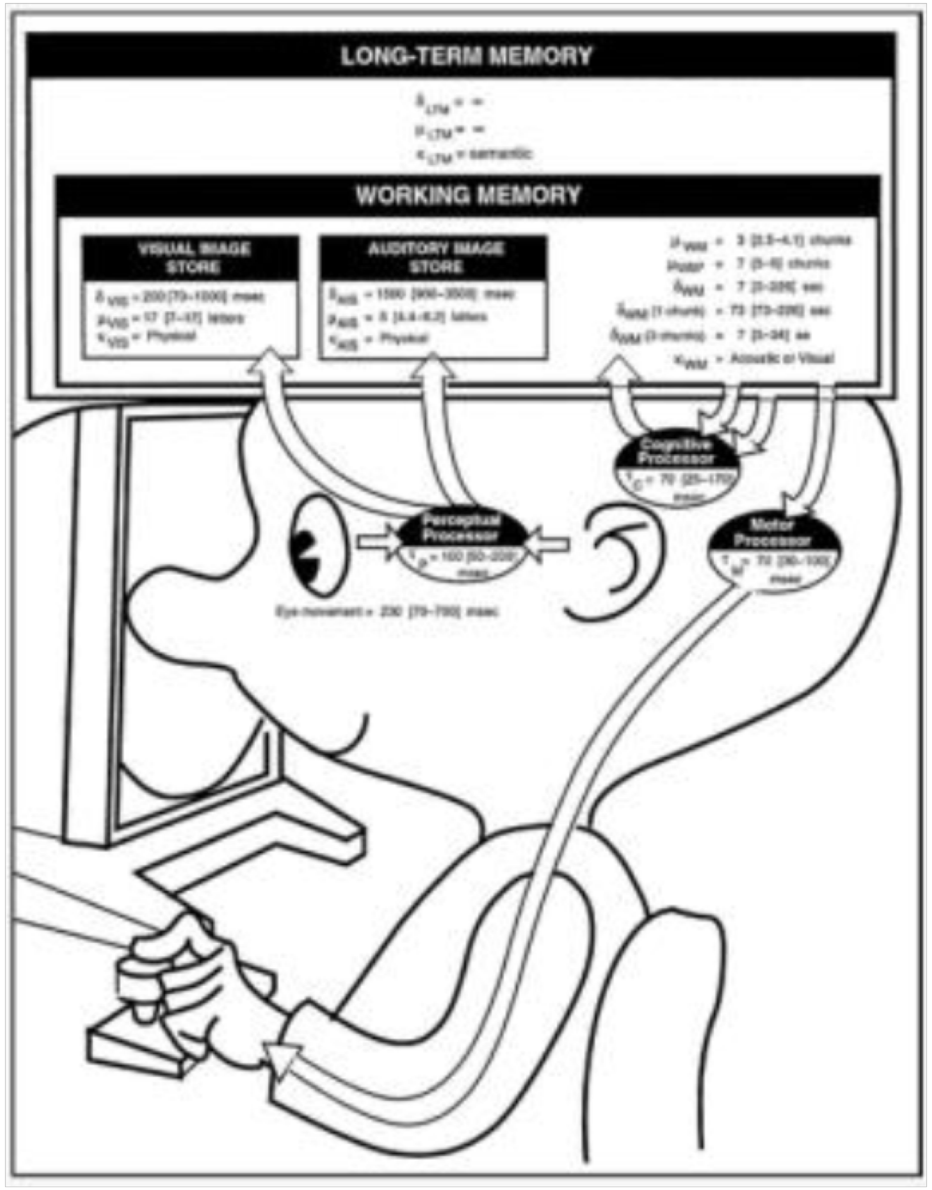
- Analyzes trade-offs in value of **information gained** against the **costs of performing activity** in human-computer interaction tasks
  - And needs models and analysis techniques to determine value added by information access, manipulation, and presentation techniques

- Real information system design problem is not **how to collect more information**, but **how to optimize user’s time**
  - Increase relevant information gained per unit time expended

- IFT provides a relatively “formal” (quantitative) account

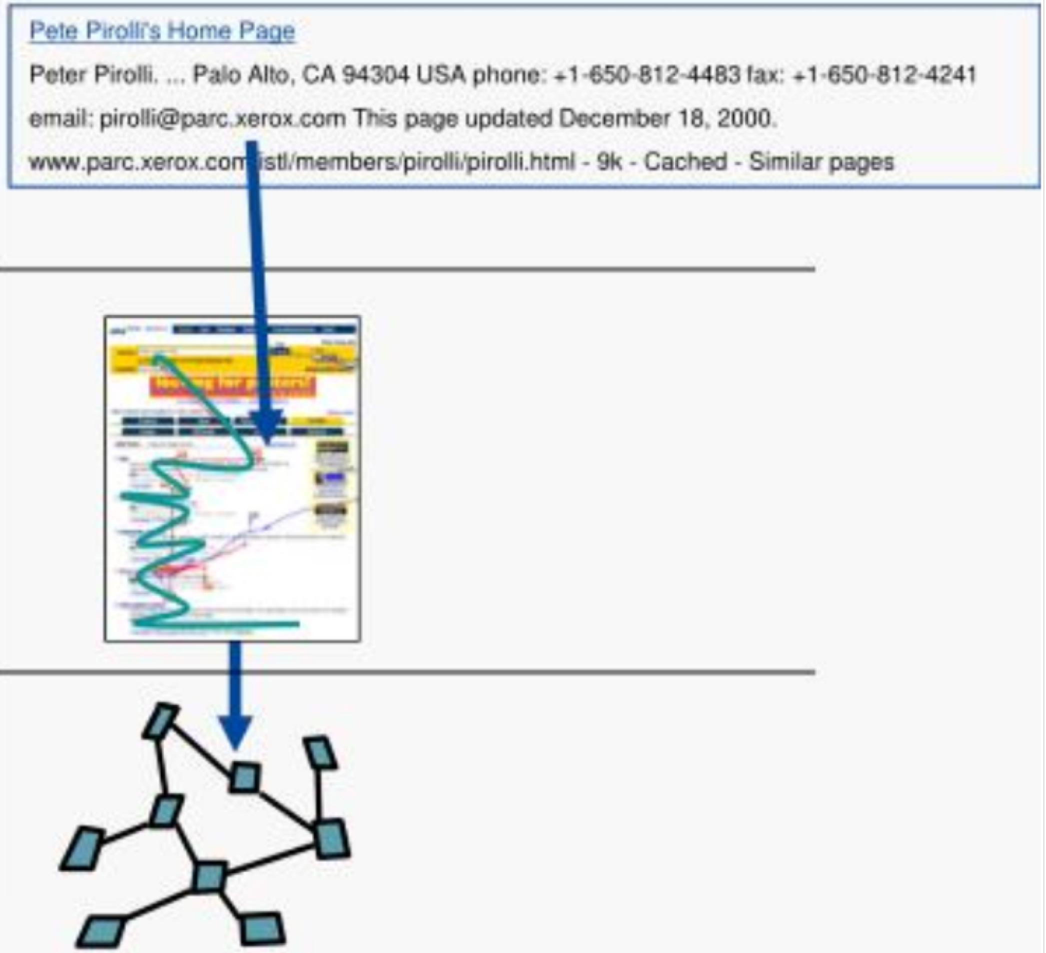
IFT timescales

- Considers “adaptiveness of human-system designs in the context of the information ecologies in which tasks are performed”
    - Ecology, as system; here -- information
- **Timescales** of information seeking and sense making activities:
    - **Cognitive** band (~100 ms – 10 s)
    - **Rational** band (minutes to hours)
    - **Social** band (days to months)



Timescale of Analysis

Time scale (s)	Psychological domain	User Interface Domain
.100-1	• Visual attention • Perceptual judgment	
1-100	• Visual search • Motor behavior	
10-1000	• Problem solving • Decision making	



IFT – An Ecological Perspective

- Time scales of information seeking and sense making activities
  - Cognitive band (~100 ms – 10 s)
  - Rational band (minutes to hours)
  - Social band (days to months)

- As time scale increases, less regard for how internal processing accomplishes linking of actions to goals

- Assumes behavior governed by “rational principles and shaped by constraints and affordances of the task environment”

- An **ecological perspective**, i.e., that behavior is “adaptive” in that it accomplishes some goal

Optimal Foraging Theory - Biology

- Developed in biology for understanding opportunities and forces of adaptation
  - the theory helps in understanding existing human adaptations for gaining and making sense of information
  - aid in task analysis for creating new interactive information system designs

- Optimality models include:
  - **Decision** assumptions
    - Which of the problems faced by an agent are to be analyzed, e.g., whether to pursue a particular type of information (or prey) when encountered, how long to spend
  - **Currency** assumptions
    - How choices are to be evaluated, e.g., information value (food value)
  - **Constraint** assumptions
    - Limit and define relationships among decision and currency variables, e..g., from task structure, interface technology, user knowledge

Information Foraging Theory

- Information foraging is usually a task embedded in context of some other task
  - Value and cost structure defined in relation to the embedding task
  - Value of external information may be in improvements to outcomes of embedding task
- Usually, embedding task is some ill-structured problem
  - Additional knowledge is needed to better define goals, available actions, heuristics, etc.
  - E.g., choosing a graduate school, developing business strategy
- Though use optimality model, not imply human behavior is classically rational
  - I.e., have perfect information and infinite computational resources
  - Rather, humans exhibit bounded rationality, or make choices based on satisficing

IFT – Information Patch Model

- Information patch model – from optimal foraging theory

- Rate of currency intake, $R = U / (T_s + T_h)$
  - U = net amount of currency (value, e.g., food, information) gained
  - T_s = time spent searching
  - T_h = time spent exploiting

- Net currency gain, $U = U_f - C_f$
  - U_f = overall currency intake (gross amount foraged)
  - C_f = currency expended in foraging

- Average rate of currency intake $u = U_f / \lambda T_s$
  - If assume information workers/foragers/consumers encounter information as **linear** function of time (will revisit this)
  - Total n items encountered = λT_s, where λ is rate of encounter with items
  - (will use next slide)

IFT – Information Patch Model

- Average cost of handling items (1st total/rate, the average) :

$$\bar{h} = \frac{T_h}{\lambda T_s}, \text{ and so } U_f = \bar{u}\lambda T_s \text{ and } T_h = \bar{h}\lambda T_s$$

- Let s = search cost per unit time, then total cost of search = sT_s

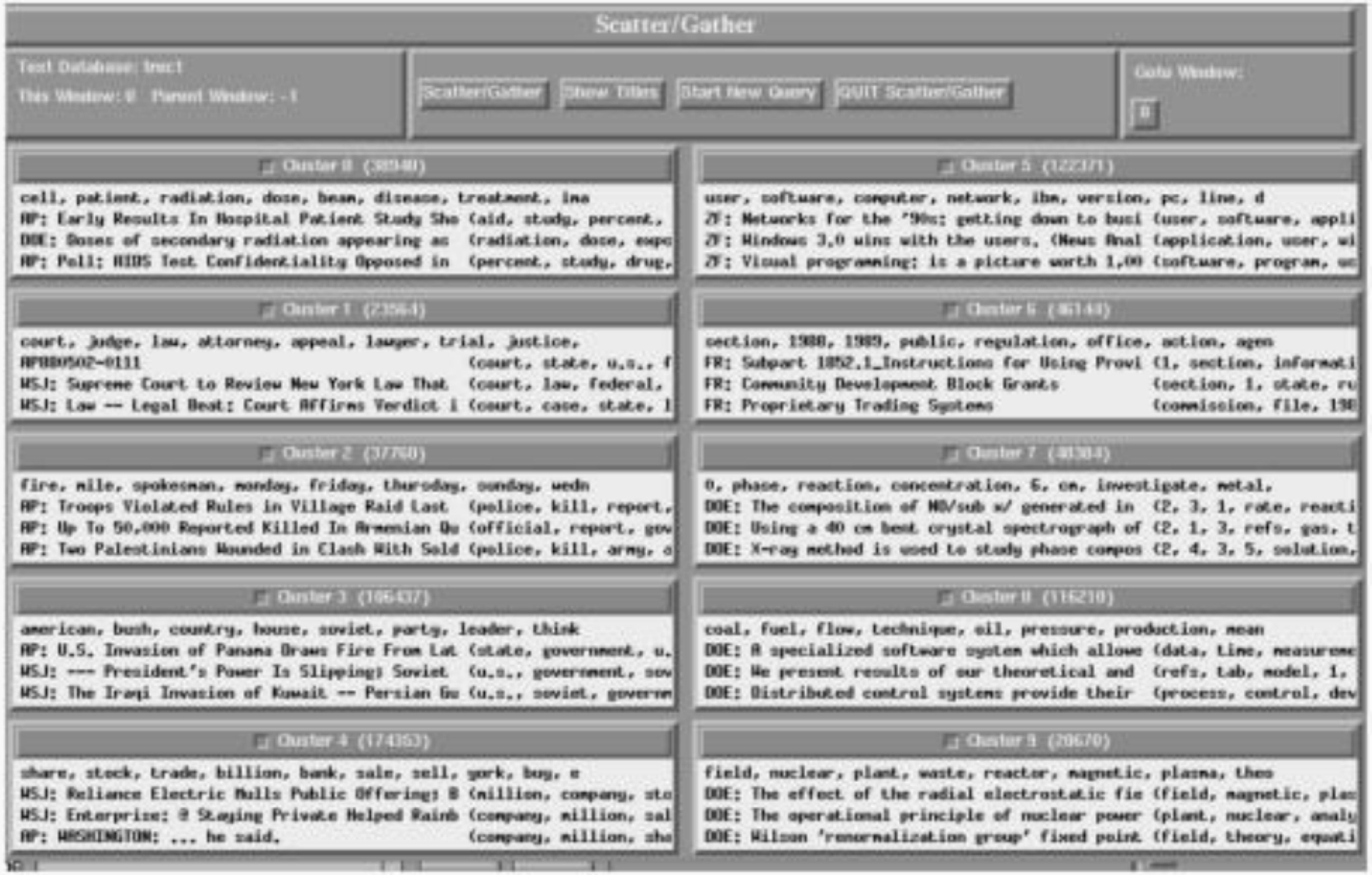
- Then, substituting in equation for R, rate of currency intake:

$$R = \frac{\bar{u}\lambda T_S - sT_S}{T_S + \bar{h}\lambda T_S} = \frac{\lambda\bar{u} - s}{1 + \lambda\bar{h}}$$

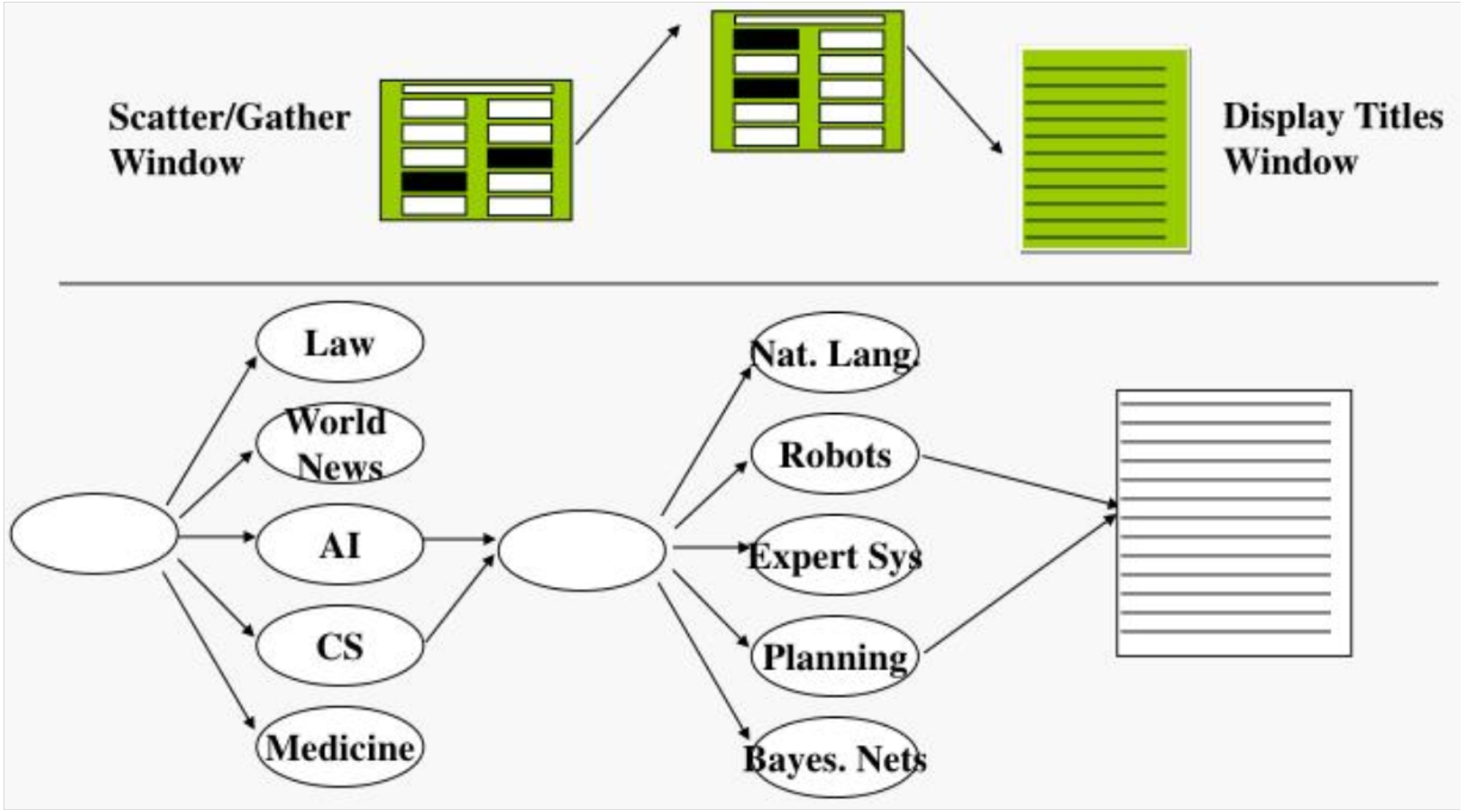
- So, can express R in terms of
  - Average rate of currency intake, u
  - Search cost per unit time, s
  - Cost of handling items, h

An Example: Scatter Gather

- Hierarchical clustering of document
- Users see “overview” of document clusters
- Allows user to navigate through clusters and overviews



Scatther/Gather Task



Optimal Foraging Time in a Patch

- $g_i(t)$, cumulative gain function
  - Amt of information gained in time t
  - $g_A(t)$ = *random order of encounter*
    - *Increase in information equal for all elements*
    - *Hence, constant slope*
  - $g_B(t)$ *and* $g_c(t)$ = *ordered by relevancy*
    - *"Relevant" items, those with higher information content, encountered earlier*
    - *Hence, highest rate of information increase earlier, and rate decreases*
- λ_p, rate of encounter with relevant items
- *x*-axis, travel time between patches
- R_B and R_C = rate of return
- t_c and t_b optimal foraging time
  - Foraging longer in the "patch" not optimal

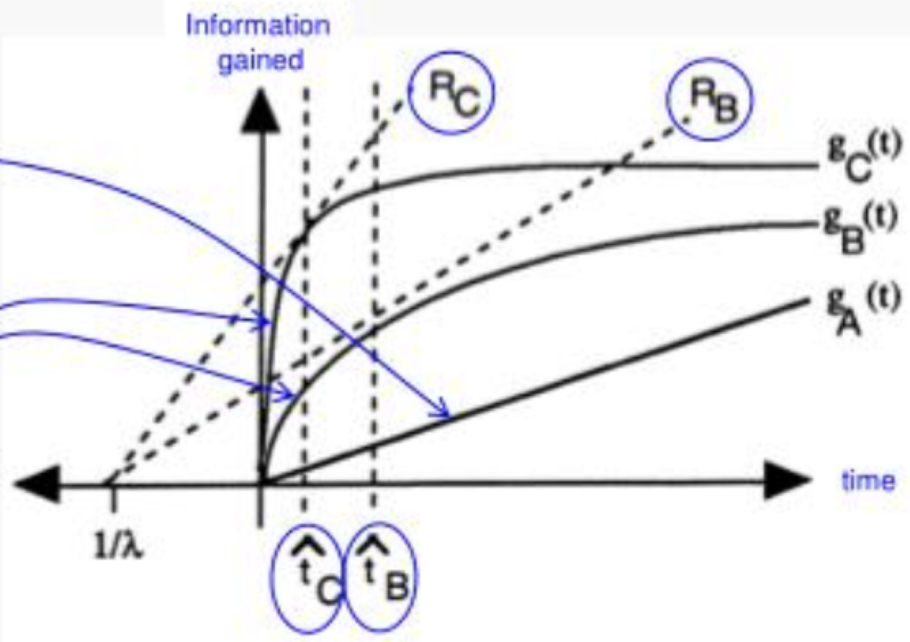


Figure 2. Optimal foraging in information cluster-patches. $g_A(t)$ is the current information-gain function, and $g_B(t)$ and $g_C(t)$ illustrate the effects of ordering by relevancy.

IFT – Cost of Knowledge

- Foraging Efficiency
  - Animals minimize energy expenditure to get required gain in sustenance
  - Humans minimize effort to get necessary gain in information
- Foraging for food has much in common with seeking information
  - Like edible plants in wild, useful information items often grouped together, but separated by long distances in an "information wasteland"
- Information "scent"
  - Like scent of food, information in current environment that will assist in finding more information clusters
- Activities analyzed according to value gained and the cost incurred
  - **Resource costs**, e.g. expenditures of time and cognitive effort incurred
  - **Opportunity costs**
    - Benefits that could be gained in engaging in other activities
    - "Cost of lost opportunity", e.g., if not gaining information about algorithms (or messing with registration system), could be gaining information about software design

IFT - Conclusions

- Information processing systems evolve so as to maximize the gain of valuable information per unit cost
  - Sensory systems (vision, hearing)
  - Information access (card catalogs, offices)

$$\textit{maximize}\left(\frac{\text{information value}}{\text{cost of interaction}}\right)$$