The Information Need (IN)

- Searching is motivated by a problematic situation
- **Gap** in user knowledge between what they know and what they want to know is the *information need*
  - ASK- Anomalous State of Knowledge *(Nicholas Belkin)*
- The IN **is not static**, and develops during the search session as the user **learns** from interaction

- The transformation of a user's information need into a query is known as *query formulation process*
  - One of the most challenging activities in information seeking: amplified if the information need is **vague** or the user knowledge about the collection is *poor*
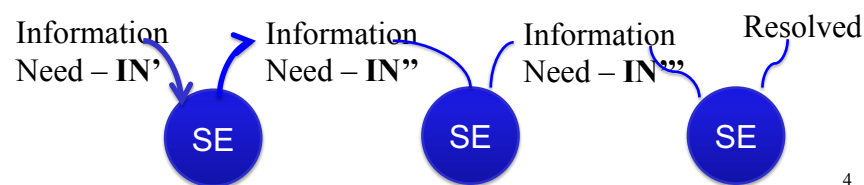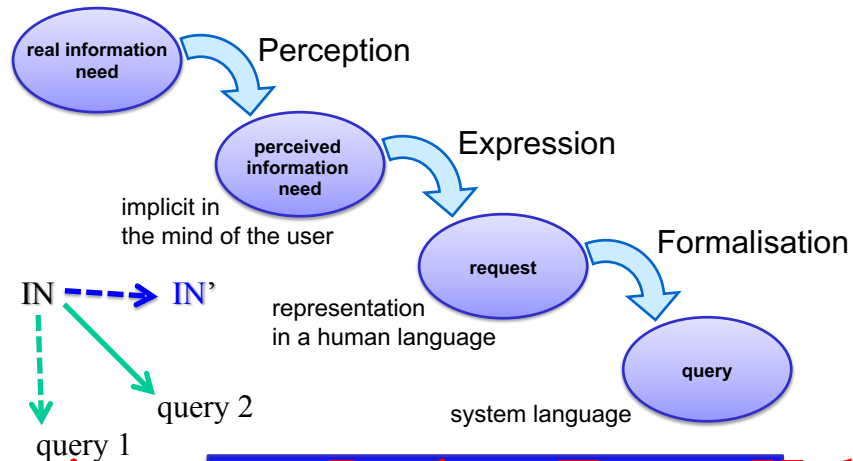
Information Need is Dynamic

- It evolves as we interact with a system
- We **learn** from our search **interaction**
- Our understanding possibly improves over a search session
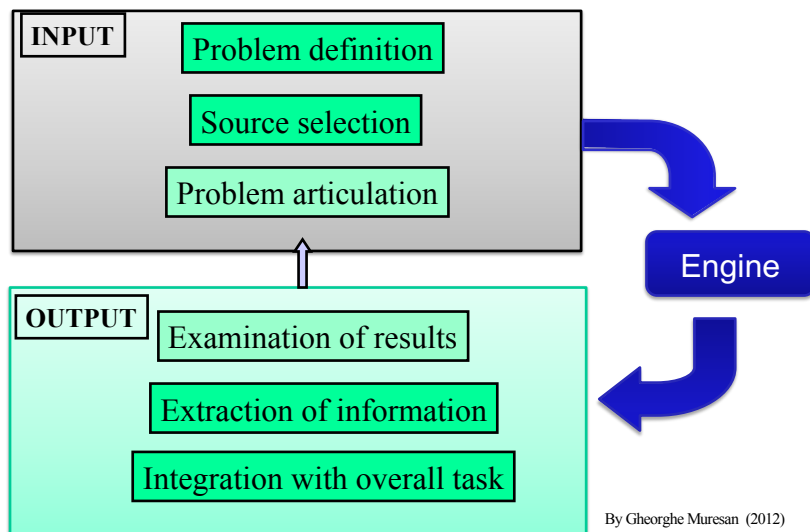- Our information need changes

Information Need – **IN'**   Information Need – **IN"**   Information Need – **IN'''**   Resolved

SE   SE   SE

4

Information Need Transformation

real information need

Perception

perceived information need

implicit in the mind of the user

Expression

IN - - - > IN'

request

representation in a human language

Formalisation

query 1

query 2

query

system language

The information need itself can change

Role of User Interfaces in IR

INPUT

Problem definition

Source selection

Problem articulation

Engine

OUTPUT

Examination of results

Extraction of information

Integration with overall task

By Gheorghe Muresan (2012)

Effective interaction is crucial for successful experience

6

Role of Search User Interfaces

- Effective search interfaces should allow users to:
  - Recognise that they really need to complete a task
  - Express their needs (queries) easily and accurately
  - Understand the structure of a search corpus
  - Judge the relevance of retrieved documents easily and accurately

> **Search User Interfaces should allow the users to operate search and manage a search task as efficiently as possible**

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Supporting Query Input

- The role of search interfaces is to allow users to **express** their information need as easily and as completely as possible
  - Keywords
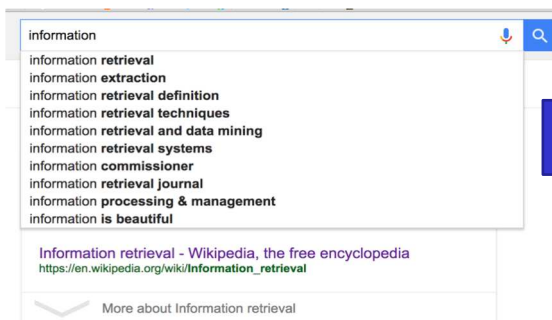 - Query by Example
  - Behavioural Signals

Simple Query Box

Web | Images | Groups | Directory

- Advanced Search
- Preferences
- Language Tools

Google Search | I'm Feeling Lucky

Simple. Easy. Everyone can use it. ASK-friendly.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

But … (Side Effect)

- 75% of Web queries are less than 3 words. Jansen and Spink (2006)

information

information **retrieval**
information **extraction**
information **retrieval** definition
information **retrieval** techniques
information **retrieval** and data mining
information **retrieval** systems
information **commissioner**
information **retrieval** journal
information **processing & management**
information **is beautiful**

Query Auto-completion/ Suggestion. Google Inc. (2008)

Information retrieval - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Information_retrieval
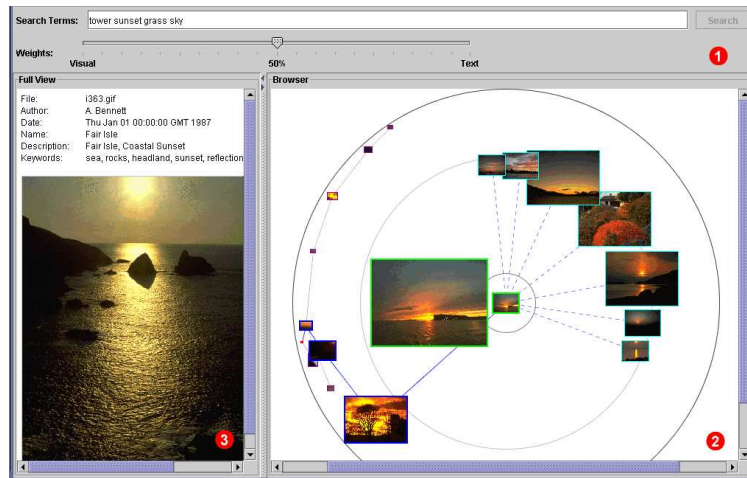
More about Information retrieval

Other Ideas

- Longer Box: People submit longer queries when a search box is longer (Belkin et al. 2003)
- More Boxes: Context box that asks what you know about the search topic (Kelly et al. 2005)

Balance between additional effort and performance gain is a challenge

10

Ostensive Interaction

Campbell and van Rijsbergen (1996)
Urban et al. (2006)

Showing or pointing to an example of what you want to tell, as opposed to verbally describing it

Other Examples of Natural User Interfaces

- **Voice input**: A user speaks to a device to express their information need (e.g. Siri, Google Voice Search)

- **Eye input**: A user's eye movement is tracked and used as an indication of interest (like a mouse cursor)

  Garkavijs et al. (2012)

- **Facial expression**: A user's facial expression is recorded and analysed to gauge their satisfaction (e.g. identify their favourite part in a video clip)

  Joho et al. (2009)

12

What is Eye-tracking?

Eye tracking device

View of subject's pupil on monitor; used for calibration
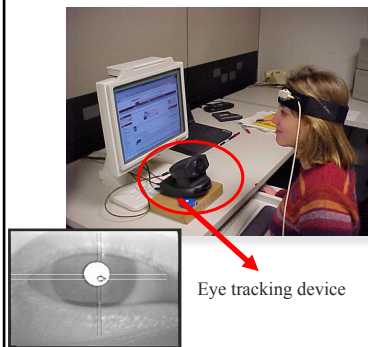
Figure: Cornell HCI eye-tracking configuration

- Device to detect and record where and what people **look at**

- Multiple applications: reading, usability, visual search, in both physical and virtual contexts

Granka et al. (2004)

Why use Eye-tracking for Information Retrieval?

- Understand how searchers browse online search results
- Suggest ideas for enhanced interface design
- More accurate interpretation of implicit feedback (e.g. clickthrough data)
- More targeted metrics for evaluating retrieval performance
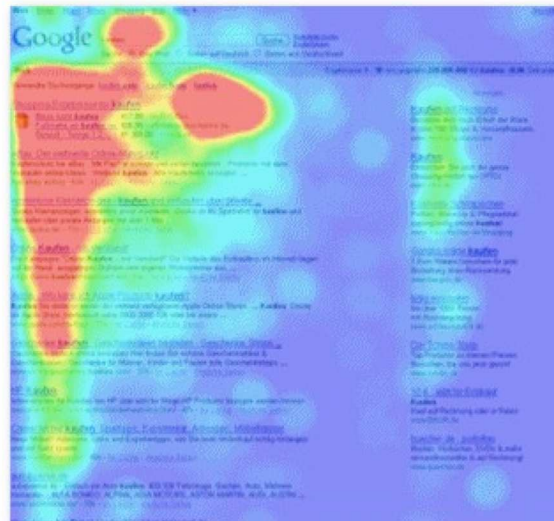
Granka et al. (2004)

14

Sample Eye-tracking Output

Granka et al. (2004) - 15
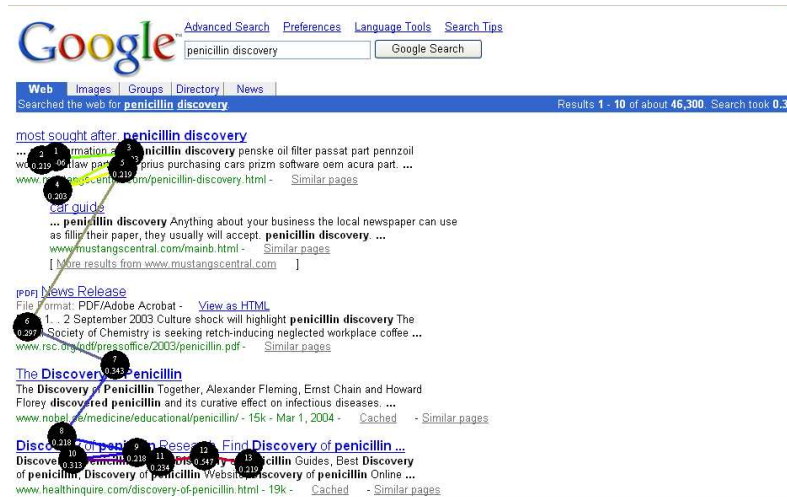
Eye-tracking Heatmap over Many Users

https://www.branded3.com/blog/seo-and-eye-tracking/

16

Eyetracking of Google Instant

Standard

Instant

17
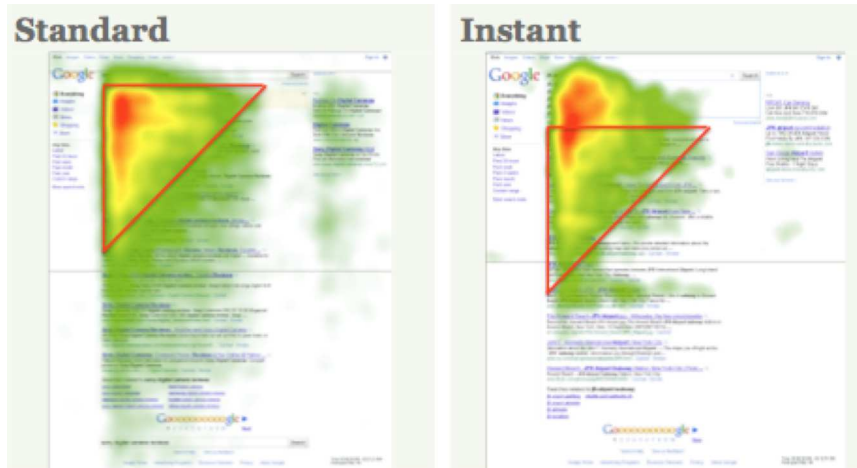
Supporting Results Examination

- The search interfaces should allow users to obtain relevant information directly, or to select documents that lead to relevant information
  - Explicit & Implicit Relevance Feedback
  - Results in Context
  - Faceted Search
  - Diversification and Aggregation

18

Relevance Feedback in Operation

Your search returned 282,000 documents

Documents in descending order (most relevant first)

☑ = more documents like these please

☐ 1. Email Us
☐ 2. Sammi`s Best Soy
☐ 3. Automatic Soy Milk Maker
☐ 4. making tofu from soy milk
☐ 5. soy milk recipe
☐ 6. Soy Milk
☐ 7. Soy bean milk recipes
☐ 8. Soy Milk Machine
☐ 9. Soy, Soy Proteins,Soy Milk, Isoflavones, Meno...
☐ 10. Soy Milk Is Safe! That Is What the Formula In...
☐ ▶ 14. Homemade Soy Milk

[Next>>]

Search the web for...

soy milk

Submit   Reset

Document Summary

Title: Homemade **Soy Milk**

- Homemade **Soy**
- Welcome to one of the most delicious, nutritious, and inexpensive beverage recipes around! Since I've started making my own **soy milk**, I drink it instead of soda and I find it much more satisfying.
- My homemade **soy milk** costs about 50 cents per gallon storebought **soy milk** is about $1.99 per quart , and no animals were exploited in the process! Another thing about **soy milk** is that it doesn't contain hormones or pesticides, and I don't think anyone wants to give those to their families.
- 2 cups organic **soy** beansbottled or filtered water 1 ½ Tablespoons vanilla 1 ½ C. Turbinado sugar or white ⅛ teaspoon lecithin granules optional Soak: **soy** beans overnight or about 10 hours in enough water to cover, and then some they will expand greatly

Number of links: 8
Fast object:
Document size: 0 K

Top Ranking Sentences

104 sentences in total (top 25 shown)

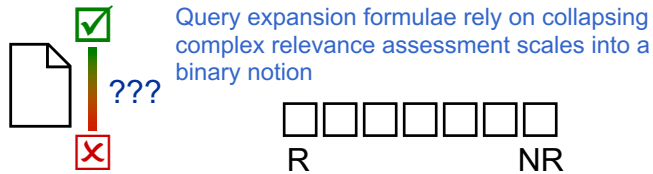|	
	in terms of disease prevention, even as only part of a non vegetarian
10	**Soy Milk** vs Rice **Milk** vs Nut **Milk** From Beccs April 30, 1999 In your opinion, which tastes better:
11	Enter your email address above to receive the next issue of The **Soy** Daily! ● Find **soy** while dining out, promote **soy** choices everywhere! Be one of The **Soy** Explorers! .
	Nutritional Analysis of Soybeans, Dr.

White et al. (2004)

Explicit Feedback

19

Problems with (Explicit) Relevance Feedback (1)

- Relevance Feedback systems are recall-dependent
  – A small number of retrieved relevant docs can adversely affect the derivation of a new query

- Exploration-Exploitation trade-off
  – The balance between users visiting documents to assess relevance because *they want to* and because *they have to*

- Visiting documents to assess relevance is a **tedious, cumbersome** and **time-consuming** process

Relevance Feedback is a cognitively overloaded activity

20

Problems with Relevance Feedback (2)

- Two problems that are somewhat linked:
  - Treats relevance as a binary notion
  - Does not handle multi-topic or partially relevant documents

☑

??? Query expansion formulae rely on collapsing complex relevance assessment scales into a binary notion

☒

☐☐☐☐☐☐☐
R   NR

- Relevance is an abstract, intuitive concept that cannot be adequately expressed with 'yes/no'

Recall that user information need is vague and document representation is uncertain

21

Implicit Relevance Feedback

- Implicit Relevance Feedback grew out of the reluctance of users to mark documents as relevant by clicking checkboxes next to document titles
  - Some searches are precision-oriented
  - Explicit feedback is **tedious**

- Can we gather feedback without requiring the user to do anything?

- **Idea**: gather feedback from observed user behavior
  - Attempting to determine what is relevant based on user **interaction** with a search system and returned documents

22

Observable Behavior

Minimum Scope

	Segment	Object	Class
Examine	View Listen	Select	
Retain	Print	Bookmark Save Purchase Delete	Subscribe
Reference	Copy / paste Quote	Forward Reply Link Cite	
Annotate	Mark up	Rate Publish	Organize

*(rotated label on left: **Behavior Category**)*

23

Implicit Relevance Feedback in Search Engines

- Users no longer have to click checkboxes and browse to each document to assess it
- Typically, Web search engines **log users' interaction**
  - Clickthrough activity; Doc viewing time; Scrollbar activity, Mouse clicks; Search **session** activity; etc.
- Combine and use (mine) these to predict the user's intention and information need
- Implicit Relevance Feedback is less accurate than explicit RF
  - Confounding variables like *presentation bias*
  - Like much in Computing Science it is a trade-off!
  - But more useful than pseudo-relevance feedback (aka Blind RF), which contains no evidence of user judgments

24

Traditional vs. Session-based Retrieval

Traditional ("1-query")

Query="IR applications"

Retrieval System ← **Document Collection**

"IR" can mean either "information retrieval" or "infrared"

Results:
D1 (infrared)
D2 (infrared)
D3 (retrieval)
D4 (infrared)
D5 (retrieval)

Session-based

Query="IR applications"

Previous query= "retrieval systems" ...

Retrieval System →

Frequency in viewed docs:

Infrared: 0
Retrieval: 5
...

Results:
D3 (retrieval)
D5 (retrieval)

Uses more contextual information
Gives more accurate results

Sriram et al. (2004)

TileBars – Hearst (1995)

http://people.ischool.berkeley.edu/~hearst/research/tilebars.html

TileBars

TileBars: Term Distribution in Information Access

Min Overlap Span
0   2   4   6   8   10

Search   Redisplay   Reset   Exit

Min Hits   Min Distribution (%)

Term Set 1: patient medicine medical   0 2 4 6 8 10   0 10 20 30 40 50
Term Set 2: test scan cure diagnosis   0 2 4 6 8 10   0 10 20 30 40 50
Term Set 3: software program   0 2 4 6 8 10   0 10 20 30 40 50

Documents Within Contraints

Total: 9

"Paragon ports its 2–D imaging software to DEC's AVS. (Paragon Imaging In

"DOD health system costs break $1.1 billion barrier. (Composite Health Care

"Computer graphics in medicine: from pictures to analysis.&O (Industry Profi

"LabVIEW 2.0 still just over the horizon. (National Instruments LabVIEW 2.0

"Hospital bridges islands of data. (includes related articles on growth of a me

"VA automation means faster admissions. (US Department of Veterans Affair

"Better ADP could cut VA delays 40%, officials say. (automatic data process

"Army tests prototype battlefield information system.&O "

"The electric cadaver. (computerized anatomy lessons and digital dissection

Three Term sets

Large rectangle indicates a document

Click on a tile to see the contents of the document.

Term frequency and distribution information is important for determining relevance.

Results in context: Show user the relationship between the words in the query and the documents retrieved.

26

Diversity in Web Search

Queries are often underspecified — Santos et al. (2015)

Labels on results: *movie*, *city*, *coach*, *movie*

Aggregated Search Results

Labels on results: **Structured Data**, **News**, **Homepage**, **Wikipedia**, **Video**, **Twitter**

Emphasis on mixing different media types — Lalmas al. (2010)

Aggregated Search General Architecture

Diaz et al. (2010)
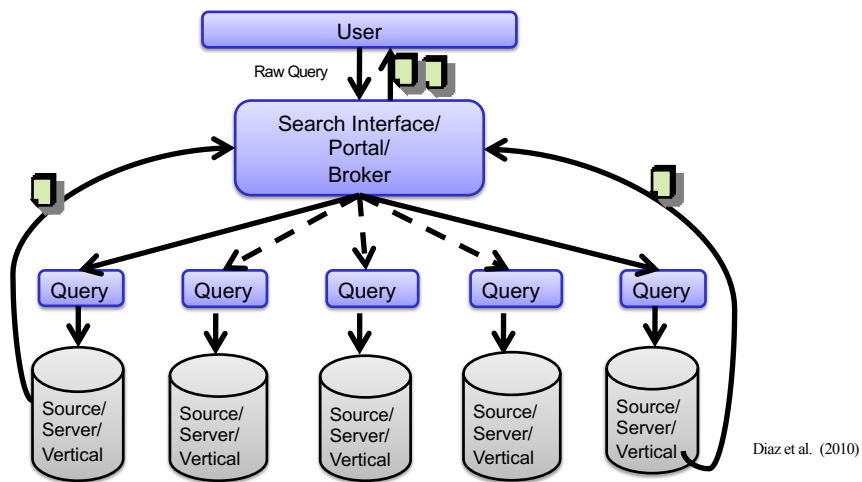
Present the user with a summary of search results from one or more resources

Faceted Search Interface

http://flamenco.berkeley.edu/demos.html
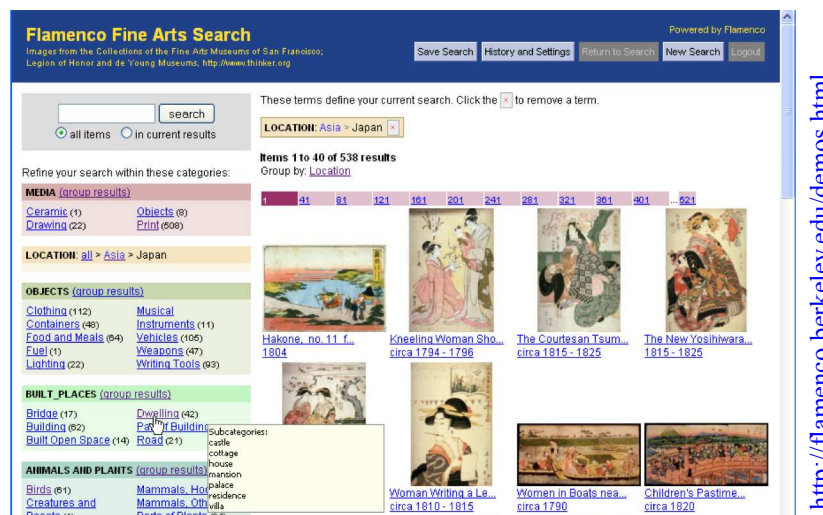
Browsing-oriented interface supporting both query formulation and results examination Tunkelang (2009)

30

Faceted Search in Action
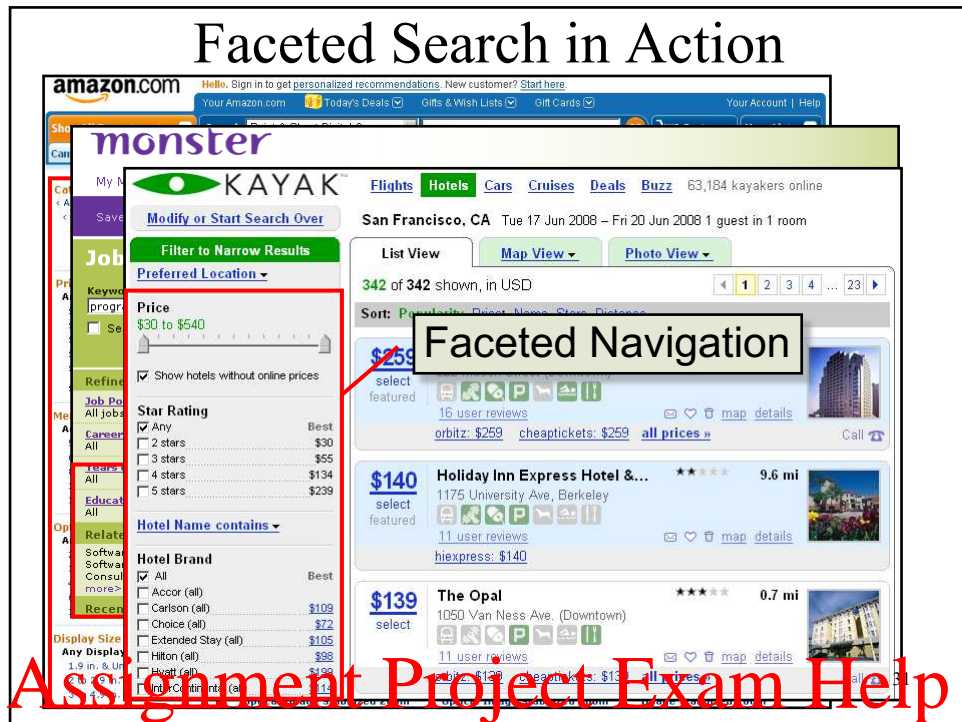
Saracevic's Relevance Model

- **Algorithmic (System) Relevance**: A search engine says this document is relevant

- **Topical Relevance**: This document is about the topic of the information need (query)

- **Cognitive Relevance**: This document has information that is new to me or has a suitable level of difficulty to me

- **Situational Relevance**: This document is suitable for my task (e.g. entertainment, report writing)

- **Affective (Motivational) Relevance**: I like the design or writing style of this document

Saracevic (1997)

32

Summary

- Information Retrieval is intrinsically an iterative process

- Effective search interfaces allow users to operate search and manage a search task as efficiently as possible

- A user-oriented evaluation of the system is necessary to gauge the overall performance of an IR system in helping users complete their search tasks

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

33