#### **User Interface for Search**

Berlin Chen Department of Computer Science & Information Engineering National Taiwan Normal University

Reference:

Modern Information Retrieval, Chapter 2 & Teaching material

#### Introduction

- This chapter focuses on
  - The human users of search systems
  - The search user interface, i.e., the window through which search systems are seen
- The **user interface role** is to aid in the searchers' understanding and expression of their information need
- Further, the interface should help users
  - Formulate their queries
  - Select among available information sources
  - Understand search results
  - Keep track of the progress of their search

## How People Search

- User interaction with search interfaces differs depending on
  - The type of task
  - The domain expertise of the information seeker
  - The amount of time and effort available to invest in the process
- Marchionini makes a distinction between information lookup and exploratory search

#### Information lookup tasks

- Are akin to fact retrieval or question answering
- Can be satisfied by discrete pieces of information: numbers, dates, names, or Web sites
- Can work well for standard Web search interactions

# How People Search (cont.)

- Exploratory search is divided into learning and investigating tasks
- Learning search
  - Requires more than single query-response pairs
  - Requires the searcher to spend time
  - Scanning and reading multiple information items
  - Synthesizing content to form new understanding
- Investigating refers to a longer-term process which
  - Involves multiple iterations that take place over perhaps very long periods of time
  - May return results that are critically assessed before being integrated into personal and professional knowledge bases
  - May be concerned with finding a large proportion of the relevant information available

## How People Search (cont.)

- More broadly, information seeking can be seen as being part of a larger process referred to as sensemaking
  - Sensemaking is an iterative process of formulating a conceptual representation from a large collection
- Russell et al. observe that most of the effort in sensemaking goes towards the synthesis of a good representation
- Some sensemaking activities interweave search throughout, while others consist of doing a batch of search followed by a batch of analysis and synthesis

# How People Search (cont.)

- Examples of deep analysis tasks that require sensemaking (in addition to search)
  - The legal discovery process
  - Epidemiology (disease tracking)
  - Studying customer complaints to improve service
  - Obtaining business intelligence

# Classic vs. Dynamic Models of Information Seeking

- Classic notion of the information seeking process:
  - 1. problem identification
  - 2. articulation of information need(s)
  - 3. query formulation
  - 4. results evaluation

Assume that user's information need is static

- More recent models emphasize the dynamic nature of the search process
  - The users learn as they search
  - Their information needs adjust as they see retrieval results and other document surrogates
- This dynamic process is sometimes referred to as the berry picking model of search

# Classic vs. Dynamic Models of Information Seeking (cont.)

- The rapid response times of today's Web search engines allow searchers:
  - To look at the results that come back
  - To reformulate their query based on these results
- This kind of behavior is a commonly-observed strategy within the berry-picking approach
- Sometimes it is referred to as **orienteering**
- Jansen et al. made a analysis of search logs and found that the proportion of users who modified queries is 52%

# Classic vs. Dynamic Models of Information Seeking (cont.)

- Some seeking models cast the process in terms of strategies and how choices for next steps are made
  - In some cases, these models are meant to reflect conscious planning behavior by expert searchers
  - In others, the models are meant to capture the less planned, potentially more reactive behavior of a typical information seeker

#### Navigation vs. Search

- **Navigation**: the searcher looks at an information structure and browses among the available information
- This browsing strategy is preferred when the information structure is well-matched to the user's information need
  - It is mentally less taxing to recognize a piece of information than it is to recall it
  - It works well only so long as appropriate links are available
- If the links are not available, then the browsing experience might be frustrating

# Navigation vs. Search (cont.)

- Spool discusses an example of a user looking for a software driver for a particular laser printer
- Say the user first clicks on *printers*, then *laser printers*, then the following sequence of links:

HP laser printers HP laser printers model 9750 software for HP laser printers model 9750 software drivers for HP laser printers model 9750 software drivers for HP laser printers model 9750 for the Win98 operating system

• This kind of interaction is acceptable when each refinement makes sense for the task at hand

#### Search Process

- Numerous studies have been made of people engaged in the search process
- The results of these studies can help guide the design of search interfaces
- One common observation is that users often reformulate their queries with slight modifications
- Another is that searchers often search for information that they have previously accessed
  - The users' search strategies differ when searching over previously seen materials
- Researchers have developed search interfaces support both query history and re-visitation

## Search Process (cont.)

- Studies also show that it is difficult for people to determine whether or not a document is relevant to a topic
  - The less users know about a topic, the poorer judges they are about if a search result is relevant to that topic
- Other studies found that searchers tend to look at only the top-ranked retrieved results
- Further, they are biased towards thinking the top one or two results are better than those beneath them

#### Search Process (cont.)

- Studies also show that people are poor at estimating how much of the relevant material they have found
- Other studies have assessed the effects of knowledge of the search process itself
- These studies have observed that experts use different strategies than novices searchers
- For instance, Tabatabai et al. found that
  - Expert searchers were more patient than novices
  - This positive attitude led to better search outcomes

#### Search Interfaces Today: Getting Started

- How does an information seeking session begin in online information systems?
  - The most common way is to use a **Web search engine**
  - Another method is to select a Web site from a personal collection of already-visited sites
    - which are typically stored in a **browser's bookmark**
  - Online bookmark systems are popular among a smaller segment of users
    - Ex: Delicious.com
  - Web directories are also used as a common starting point, but have been largely replaced by search engines

# Query Specification

- The primary methods for a searcher to express their information need are
  - Either entering words into a **search entry** form
  - Selecting links from a **directory** or other information organization display
- For Web search engines, the query is specified in textual form
  - But in future, query specification via spoken commands will most likely become increasingly common, using mobile devices as the input medium
- Typically, Web queries today are very short consisting of one to three words

# Query Specification (cont.)

- Short queries reflect the standard usage scenario in which the user tests the waters:
  - If the results do not look relevant, then the user reformulates their query
  - If the results are promising, then the user navigates to the most relevant-looking Web site
- This search behavior is a demonstration of the orienteering strategy of Web search

# Query Specification (cont.)

- Before the Web, search systems regularly supported
  Boolean operators and command-based syntax
  - However, these are often difficult for most users to understand
- Jansen et al. conducted a study over a Web log with 1.5M queries, and found that
  - 2.1% of the queries contained Boolean operators
  - 7.6% contained other query syntax, primarily double-quotation marks for phrases
- White et al. examined interaction logs of nearly 600,000 users, and found that
  - 1.1% of the queries contained one or more operators
  - 8.7% of the users used an operator at any time

# Query Specification (cont.)

- Web ranking has gone through three major phases
- In the first phase, from approximately 1994–2000:
  - Since the Web was much smaller then, complex queries were less likely to yield relevant information
  - Further, pages retrieved not necessarily contained all query words
  - Information about query term proximity within the page was not used, nor was the information about relative importance of Web pages
- Around 1997, Google moved to conjunctive queries only
  - The other Web search engines followed, and conjunctive ranking became the norm
  - Google also added term proximity information and page importance scoring (PageRank)
  - As the Web grew, longer queries posed as phrases started to produce highly relevant results

# Query Specification Interfaces

- The standard interface for a textual query is a search box entry form
- Studies suggest a relationship between query length and the width of the entry form
  - Results found that either small forms discourage long queries or wide forms encourage longer queries

- Some entry forms are followed by a form that filters the query in some way
- For instance, at yelp.com, the user can refine the search by location using a second form



 Notice that the yelp.com form also shows the user's home location, if it has been specified previously

- Some interfaces show a list of query suggestions as the user types the query
  - This is referred to as auto-complete, auto-suggest, or dynamic query suggestions
  - Anick et al. found that users clicked on dynamic Yahoo suggestions one third of the time
- Often the suggestions shown are those whose prefix matches the characters typed so far
  - However, in some cases, suggestions are shown that only have interior letters matching
- Further, suggestions may be shown that are synonyms of the words typed so far

• Dynamic query suggestions, from Netflix.com

	bei	Search
Vinners	Being There	
-	Being John Malkovich	
	Being Julia	
	Beijing Bicycle	
	Beiderbecke Connection	
	The Unbearable Lightness of Being: Special Edition	
	The Importance of Being Earnest	
	The Beiderbecke Affair	
	The Business of Being Born	
	Devious Beings	

- The dynamic query suggestions can be derived from several sources, including:
  - The user's own query history
  - A set of metadata that a Web site's designer considers important
  - All of the text contained within a Web site

## Retrieval Results Display

- When displaying search results, either
  - The documents must be shown in full, or else
  - The searcher must be presented with some kind of representation of the content of those documents
- The document surrogate refers to the information that summarizes the document
  - This information is a key part of the success of the search interface
  - The design of document surrogates is an active area of research and experimentation
  - The quality of the surrogate can greatly effect the perceived relevance of the search results listing

Vertication of the second seco		▼ 😽 🗙 🚼 限水	
案(F) 編輯(E) 檢視(V) 我的最愛(A)	工具(T) 說明(H)		
我的最愛 😭			
限水 - Google 搜尋	🛐 ▼ 🗟 ▼ 🖃 🖶 ▼ 網頁(P)▼ 安全性(S)▼		
有物质 圖片 影片 地圖 新聞 翻	耀 Gmail 更多▼		
C			
Google	限水	搜尋	
0	約有 10,300,000 項結果 (搜尋時間:0.74 秒)	進階搜尋	
諸 🛃	「限水」的新聞搜尋結果		廣告
● 圖片			UB-FINE 免手觸節水系統 Q
■ 影片	2 小時前		省水節能深入您生活.腳踏式水龍頭 有效省水達65.5%.環保規格高標準.
■ 新聞	經濟部今天(16日)中午舉行抗旱會議,將敲定是否延後桃竹等地的第二階段 <mark>限水。如果延後後期的時間。</mark> 後原訂18、23日實施的二階段 <mark>限水</mark> 措施,全台工業用戶減供5%的用水壓力也可暫時解		月双省水達05.5%.1度保稅格高標準。 桃園縣蘆竹鄉吉林路137號
■ 即時	聯合新闻網 除。		www.ub-fine.com
			您可以在這裡登廣告 »
▼ 更多	六月底不第三階段 <mark>限水 桃園農灌可能移用</mark> ♀ 中時電子報-7 則相關文章		
搜尋所有網站	《經濟》老天幫忙,第二段階限水延後 🔍		
搜尋所有中文網頁	中網理財 - 422 則相關文章		
搜尋繁體中文網頁 外文網頁翻譯版	中部稍解渴第2階段限水延後-Yahool奇摩新聞 Q		
	2011年5月16日 中央社記者管瑞平苗栗縣15日電)行政院長吳敦義今天到苗栗縣鯉魚潭水庫視		
不限時間 最新內容	察中部地區水情,由於連日來明顯降雨,全台水庫皆有進帳,原訂23日啟動話栗及 tw.news.yahoo.com, 生活,氯象,中央社		
過去 24 小時	水庫稍稍解渴 <mark>限水</mark> 延至6月-Yahoo!奇摩新聞 Q		
過去2天	<u>水庫相相解為限水鉱至0月-101001台摩新面</u> 2011年5月14日 這幾天鋒面報到,讓全台水庫進帳3000萬噸,等於多了6天的用水		
過去1週	tw.news.yahoo.com> 財經>經貿財稅>華視		
過去1個月	水庫暫時"解渴"限水將延到6月-Yahoo!奇摩新聞 Q		
過去 1 年	2011年5月14日 大家都希望,下雨能下在對的地方,而這波鋒面目前已經讓全台		
自訂日期範圍	tw.news.yahoo.com) 生活) 氣象) 華視 - 頁庫存檔		
更多搜尋工具	■ 顯示更多來自 vahoo.com 的結果		

- In Web search, the **page title** is usually shown prominently, along with the URL and other metadata
- In search over information collections, metadata such as date published and author are often displayed
- Text **summary** (or **snippet**) containing text extracted from the document is also critical
- Currently, the standard results display is a vertical list of textual summaries
- This list is sometimes referred to as the SERP (Search Engine Results Page)

- In some cases the summaries are excerpts drawn from the full text that contain the query terms
- In other cases, specialized kinds of metadata are shown in addition to standard textual results
  - This technique is known as blended results or universal search
  - For example, a query on a term like "rainbow" may return sample images as one entry in the results listing



 A query on the name of a sports team (e.g., "rockets") might retrieve the latest game scores and a link to buy tickets

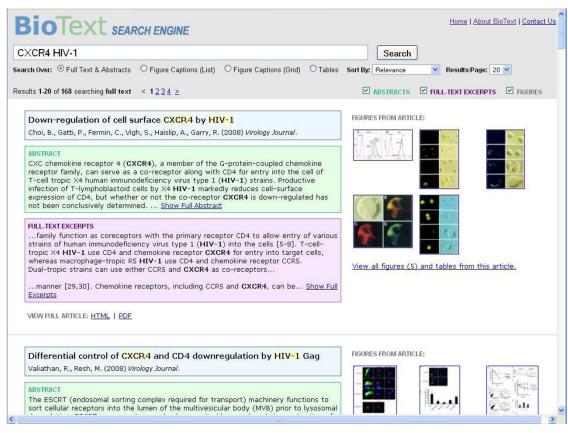
Web Images Maps News	Video Gmail more 🔻	
Google	kets	Search Advanced Search Preferences
Web Video News Blogs	Images	Results 1 - 10 of about 22,800,000 for rockets [definition].
NBA.com - Houston Rock Official site containing news, s schedules. www.nba.com/rockets/ - 7k - 0	cores, audio and video files,	player statistics, and
Scores and Schedule Tickets E-Brochure Players	<u>Rockets Power Dancers</u> <u>Stats</u> <u>Giveaway Nights</u> <u>Video Gallery</u>	
More results from nba.c	om »	
ROCKETS: 2008-09 ROO Rocket Power Dancers · Clutc Little Dippers · Recycle Item of www.nba.com/rockets/schedu	h the Bear · Red Rowdies · F the Month ROCKETS SCHI	EDULES & RESULTS
Rocket - Wikipedia, the fr A rocket or rocket vehicle is a reaction of the rocket to the eje en.wikipedia.org/wiki/Rocket -	missile, aircraft or other vehi ection of fast moving fluid	
Video results for rockets		
lakers vs rocke kobe bryant huo 7 min	<u>1e</u>	How To Make a Mentos Coke Rocket
www.youtube.co	om e	one.rewer.com

- Nielsen notes that in some cases the information need is satisfied directly in the search results listing
  - This makes the search engine an "answer engine"
- Displaying the query terms in the context in which they appear in the document:
  - Improves the user's ability to gauge the relevance of the results
  - It is sometimes referred to as **KWIC** keywords in context
  - It is also known as query-biased summaries, query-oriented summaries, or user-directed summaries

- The visual effect of query term highlighting can also improve usability of search results listings
  - Highlighting can be shown both in document surrogates in the retrieval results and in the retrieved documents
- Determining which text to place in the summary, and how much text to show, is a challenging problem
- Often the summaries contain all the query terms in close proximity to one another
- However, there is a **trade-off** between
  - Showing contiguous sentences, to aid in coherence in the result
  - Showing sentences that contain the query terms

- Some results suggest that it is better to show full sentences rather than cut them off
  - On the other hand, very long sentences are usually not desirable in the results listing
- Further, the kind of information to display should vary according to the intent of the query
  - Longer results are deemed better than shorter ones for certain types of information need
  - On the other hand, abbreviated listing is preferable for navigational queries
  - Similarly, requests for factual information can be satisfied with a concise results display

- Other kinds of document information can be usefully shown in the search results page
  - E.g., the page results below show figures extracted from journal articles alongside the search results



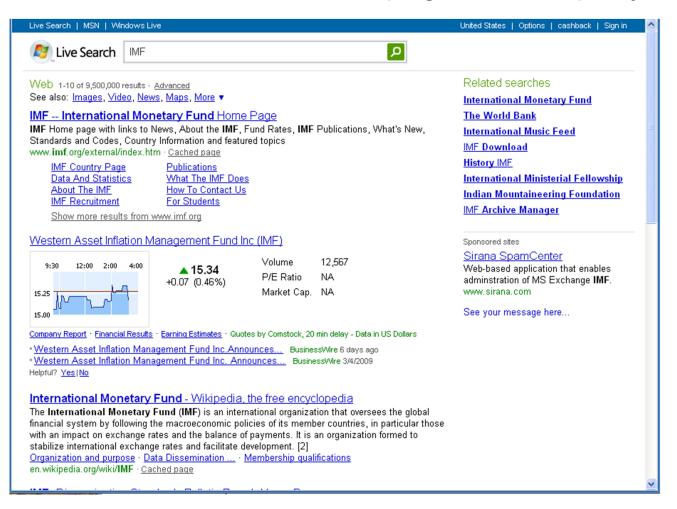
IR – Berlin Chen 33

# Query Reformulation

- There are tools to help users reformulate their query
  - One technique consists of showing terms related to the query or to the documents retrieved in response to the query
- A special case of this is spelling corrections or suggestions
  - Usually only one suggested alternative is shown: clicking on that alternative re-executes the query
  - Some years ago, the search results were shown using the purportedly incorrect spelling

#### Query Reformulation (cont.)

Microsoft Live's search results page for the query "IMF"



## Query Reformulation (cont.)

- Term expansion: search interfaces are increasingly employing related term suggestions
- Log studies suggest that term suggestions are a somewhat heavily-used feature in Web search
- Jansen et al. made a log study and found that 8% of queries were generated from term suggestions
- Anick et al. found that 6% of users who were exposed to term suggestions chose to click on them

## Query Reformulation (cont.)

- Some query term suggestions are based on the entire search session of the particular user
- Others are based on behavior of other users who have issued the same or similar queries in the past
  - One strategy is to show similar queries by other users
  - Another is to extract terms from documents that have been clicked on in the past by searchers who issued the same query

## Query Reformulation (cont.)

- Relevance feedback is another method whose goal is to aid in query reformulation
- The main idea is to have the user indicate which documents are relevant to their query
  - In some variations, users also indicate which terms extracted from those documents are relevant
- The system then computes a new query from this information and shows a new retrieval set

# Query Reformulation (cont.)

- Nonetheless, this method (i.e., relevance feedback) has not been found to be successful from a usability perspective
  - Because that, it does not appear in standard interfaces today
- This stems from several factors:
  - People are not particularly good at judging document relevance, especially for topics with which they are unfamiliar
  - The beneficial behavior of relevance feedback is inconsistent

# **Organizing Search Results**

- Organizing results into meaningful groups can help users understand the results and decide what to do next
- Popular methods for grouping search results: category systems and clustering
- Category system: meaningful labels organized in such a way as to reflect the concepts relevant to a domain
  - Good category systems have the characteristics of being coherent and relatively complete
  - Their structure is predictable and consistent across search results for an information collection

- The most commonly used category structures are flat, hierarchical, and faceted categories
- Flat categories are simply lists of topics or subjects
  - They can be used for grouping, filtering (narrowing), and sorting sets of documents in search interfaces
- Most Web sites organize their information into general categories
  - Selecting that category narrows the set of information shown accordingly

- Some experimental Web search engines automatically organize results into flat categories
  - Studies using this kind of design have received positive user responses (*Dumais et al.*, *Kules et al.*)
- However, it can difficult to find the right subset of categories to use for the vast content of the Web
- Rather, category systems seem to work better for more focused information collections

- In the early days of the Web, hierarchical directory systems such as Yahoo's were popular
  - Hierarchy can also be effective in the presentation of search results over a book or other small collection
- An alternative representation is the **faceted metadata** 
  - Unlike flat categories, faceted metadata allow the assignment of multiple categories to a single item
  - Each category corresponds to a different facet (dimension or feature type) of the collection of items

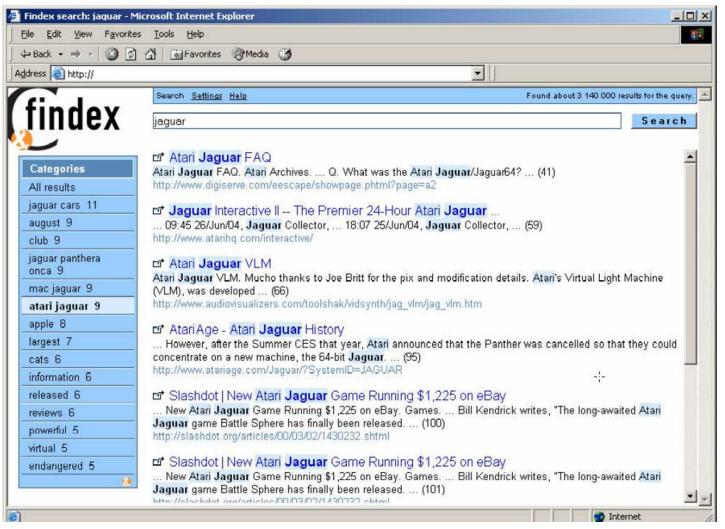
• Figure below shows a example of faceted navigation

Arts Search the Fine Arts Museums of San Francisc Museums, http://www.thinker.org	Powered by Flamenco Save Search History and Settings Return to Search New Search Logout
search O in current results ese categories:	These terms define your current search. Click the 💌 to remove a term. keyword "castle" 👻 LOCATION: Europe 💌 MEDIA: Print 💌
lithograph (21) mezzotint (14) woodcut (12)	197 items, grouped by MEDIA ( <u>view ungrouped items</u> ) <u>aquatint</u> (4)
oup results) Italy (14) Scotland (5) Spain (1) Switzerland (2) more	Caernavon Castle    Duntanborough Castle    Edinburgh Castle N    Untitled (landscap      18th - 19th century    1808    1801    Istic    Untitled (landscap
Musical Instruments (4) Vehicles (56) Weapons (27) Writing Tools (13)	drypoint (10)
Its) Dwelling (197) Part of Building (44) Road (21)	
up results) Mammals, Hoofed (43) Mammals, Other (39) Parts of Plants (4)	Lindesfame Castle      Stirling Castle, N      Castle Movie      Ilandscape with a        19th - 20th century      19th - 20th century      19th - 20th century      19th - 20th century
	hte Fine Arb Museums of San Franciso Museums, http://www.thinker.org © in current results ese categories: If thograph (21) mezzotint (14) woodcut (12) Oup results) Italy (14) Scotland (6) Spain (1) Switzerland (2) more Musical Instruments (4) Vehicles (56) Weapons (27) Writing Tools (13) Its) Dwelling (197) Part of Building (44) Road (21) Up results) Mammals, Hoofed (43) Mammals, Other (39)

IR – Berlin Chen 44

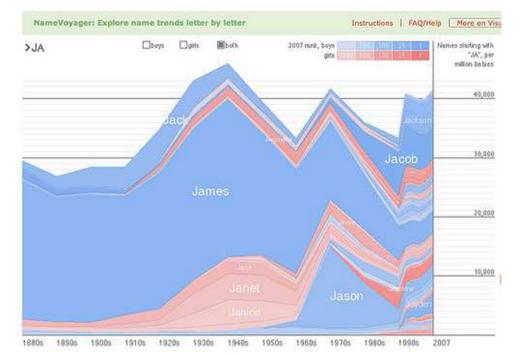
- **Clustering** refers to the grouping of items according to some measure of similarity
- It groups together documents that are similar to one another but different from the rest of the collection
  - Such as all the document written in Japanese that appear in a collection of primarily English articles
- The greatest advantage of clustering is that it is fully automatable
- The disadvantages of clustering include
  - An unpredictability in the form and quality of results
  - The difficulty of labeling the groups
  - The counter-intuitiveness of cluster sub-hierarchies

Output produced using Findex clustering



## Visualization in Search Interfaces

- Experimentation with visualization for search has been primarily applied in the following ways:
  - Visualizing Boolean syntax
  - Visualizing query terms within retrieval results
  - Visualizing relationships among words and documents
  - Visualization for text mining



# Design and Evaluation

- User interface design: a field of Human-Computer Interaction (HCI)
- This field studies how people think about, respond to, and use technology
- User-centered design: a set of practices developed to facilitate the design of interfaces
- The design process begins by determining what the intended users' goals are
- Then, the interface is devised to help people achieve those goals by completing a series of tasks

# Design and Evaluation (cont.)

- Goals in the domain of information access can range quite widely
- The design of interfaces is an iterative process, in which the goals and tasks are elucidated via **user research**
- Evaluating a user interface is often different from evaluating a ranking algorithm or a crawling technique
  - The quality of a user interface is determined by how people respond to it
  - If a person has a choice between two systems, they will use the one they prefer
  - The reasons for preference may be determined by a host of factors: speed, familiarity, aesthetics, preferred features, or perceived ranking accuracy