
Internet as an Information & Communication Utility: It's Getting Personal

Wei-Min Lu
CTO
Anchora

IEEE ComSoc-SCV
September 9, 2009



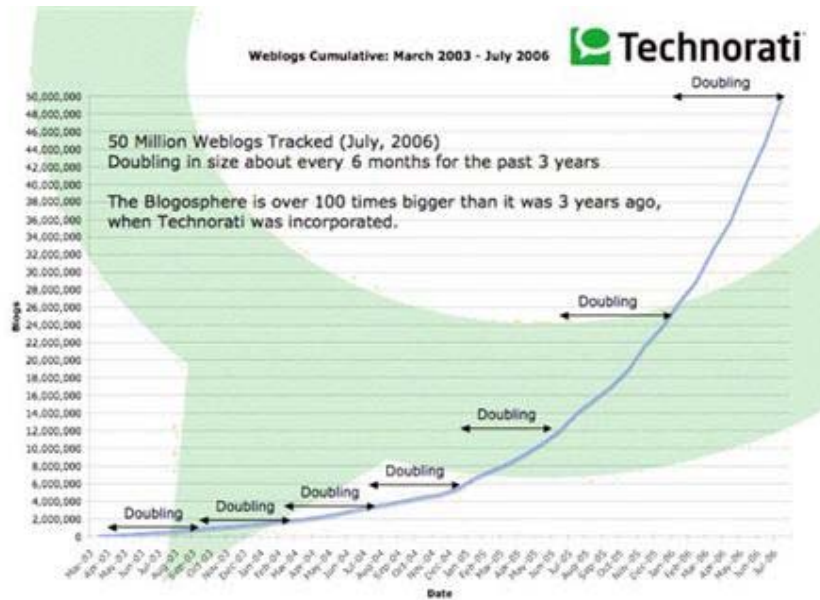
weimlu@gmail.com ♦ (408)-921-0672

Page 1
09/09/09

Overview

- Background
 - From Email to Twitter and Beyond
 - Simon's Law: Information and Attention
 - The Need for Personalization
- Personalized Information Discovery
 - Web Search: Challenges
 - Personalized Search
- Toward a Solution to Information Personalization
 - Information Search Is only the First Step...
 - Internet as a Personal Information & Communication Utility

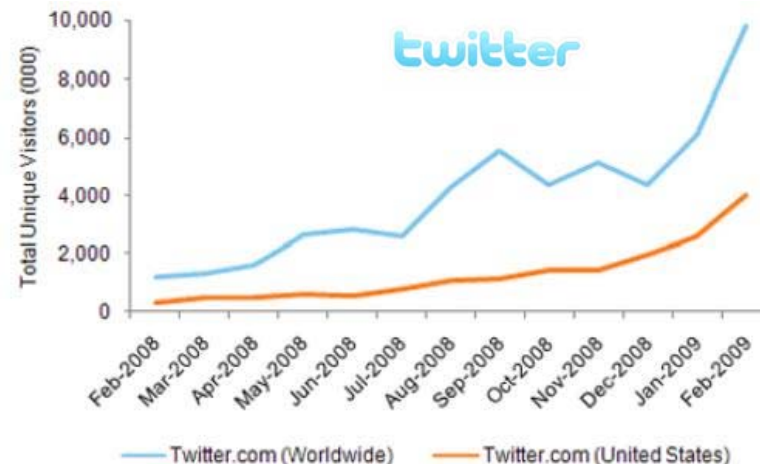
Overabundance and Exponential Growth of Internet Information



The type, scope, & amount of Web information has changed radically:

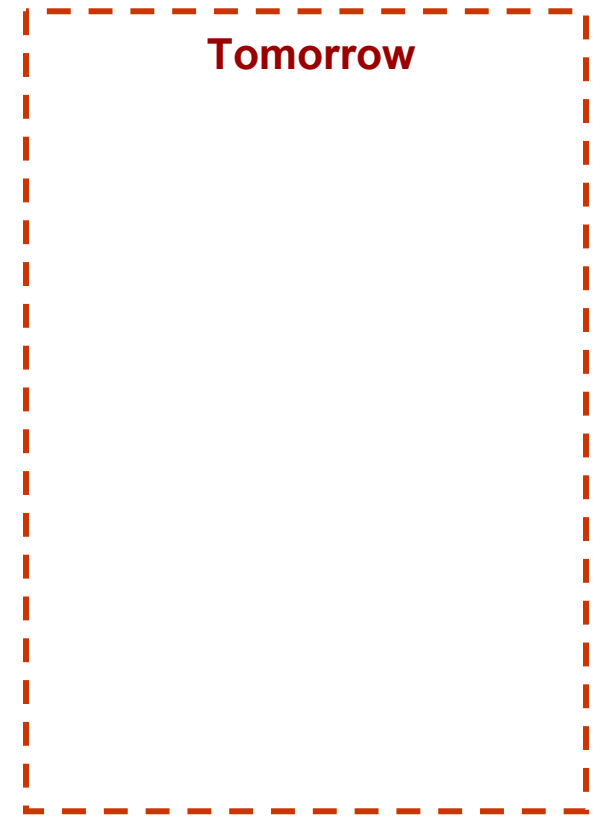
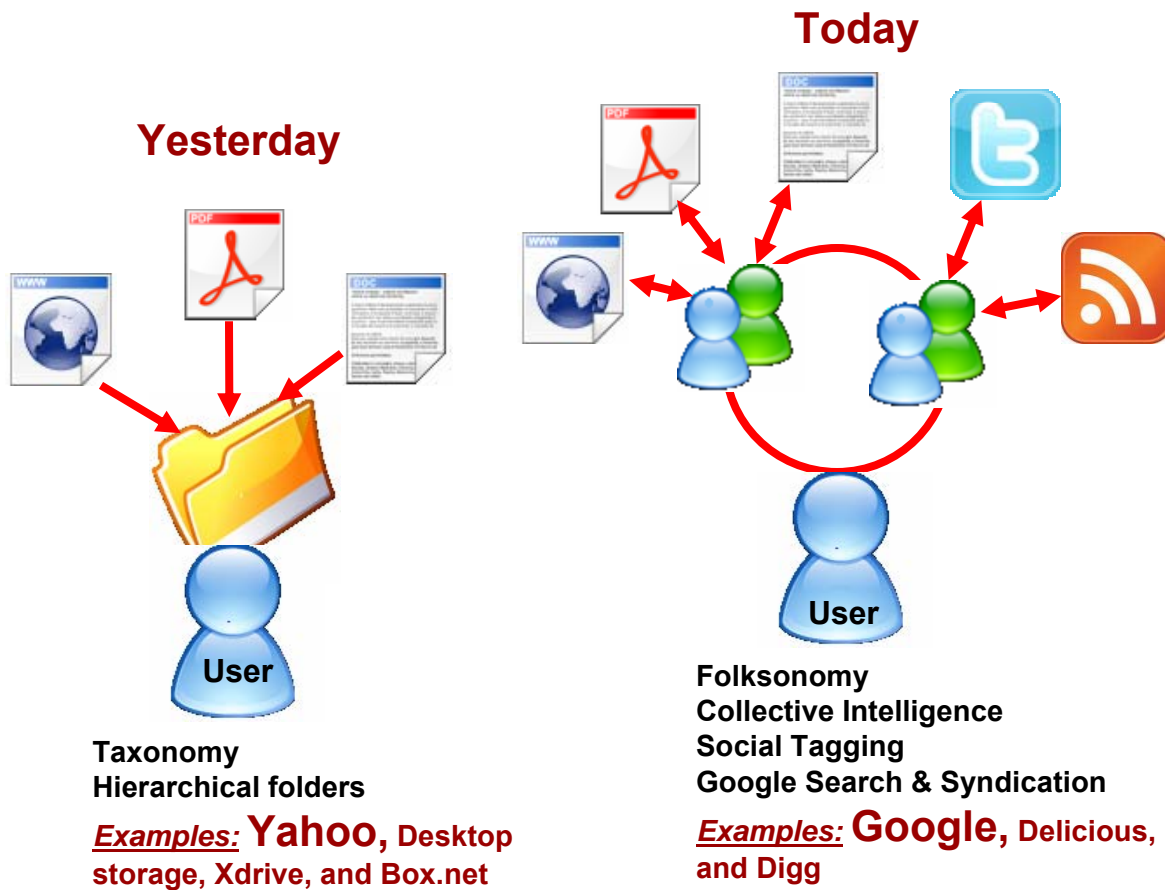
- 5 new web sites added per second
- Blogs double every 6 months
- 247 billion emails are sent per day in 2009

But the tools we use to find or manage information and make decisions have stayed basically the same.



Source: comScore Media Metrix

Information Discovery & Management



Simon's Law of Information

The rapid growth of information causes scarcity of attention

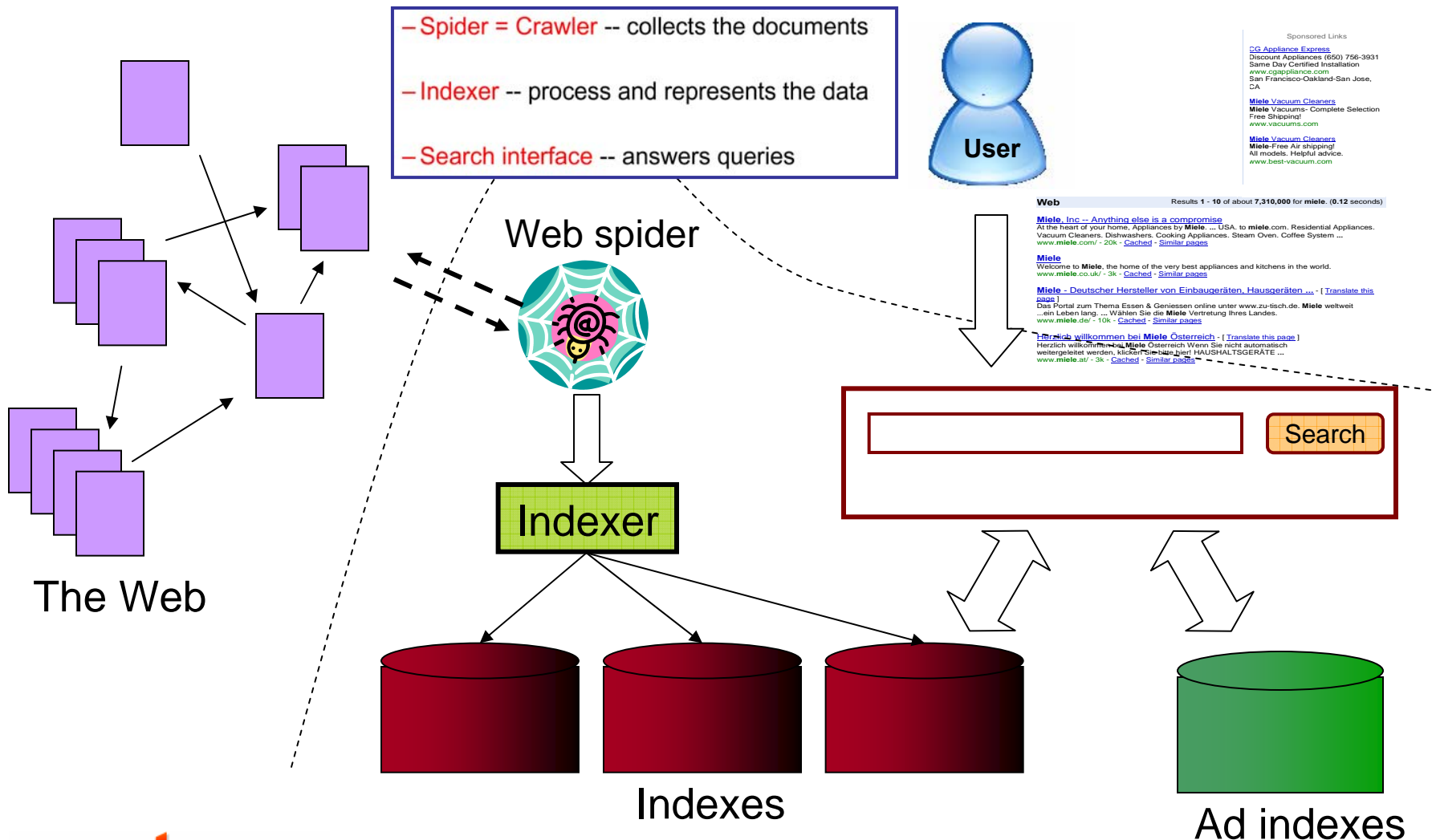
"... a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."

-- Herbert Simon, 1971

Attention and Personalization

- Attention Crisis
 - Information imbalance between publishers and consumers.
 - Information publishers: Information does not get desired attention
=> hurting their economy
 - Information consumers: Consumers do not get the desired information
=> impacting their decision-making
 - There is an urgent need to allocate attention efficiently among the overabundance of information sources
- Personalization
 - Adapts to the individual needs, interests, and preferences for information consumers by matching the right information to their attention.
 - An individually-tailored process for bringing the most relevant information to the attention of information consumers
 - A win-win for both publishers and consumers to achieve information balance

Web Search: Finding Information



Microsoft: 50% Searches Fail to Deliver

- Users are dissatisfied:
 - 50% searches fail to meet consumer's needs
 - 35% of people express dissatisfaction with search today
- Users need extra effort:
 - 72% of people thought current search results are disorganized
 - Half of all search queries are a repeat of a previous search
 - Half the searches are eventually refined
- Users are focused more on tasks and decisions
 - 66% of people reported using search engines to make decisions
 - Long sessions (> 30 minutes) are becoming more common

Challenges of Web Search: Web

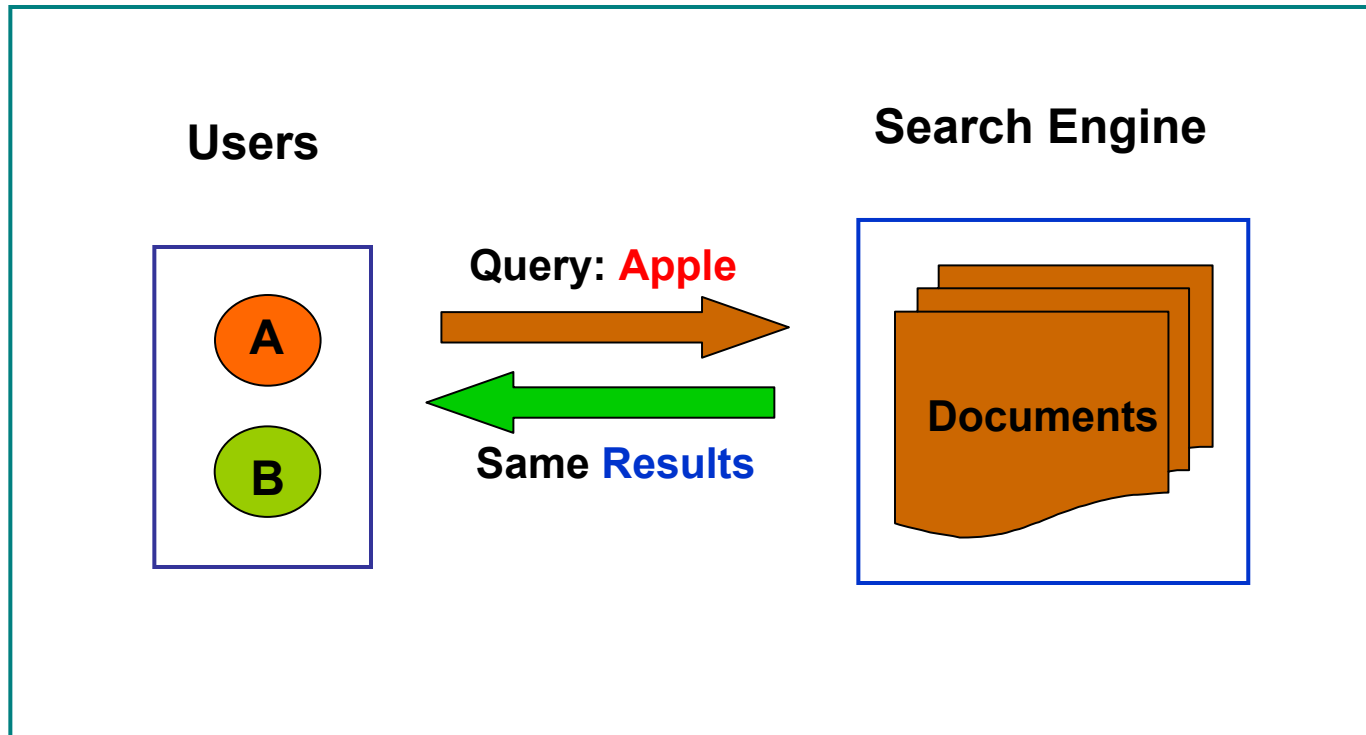
- Dynamic Web
 - The Web is really infinite - Only a fraction of Web can be indexed
 - 5 new web pages are added per second
 - Lack of stability: 25% updates per day and 40% per week
 - High linkage: > 8 links per page
- Diverse Web
 - Diverseness in standard compliance and hosting technologies
 - Heterogeneity: types, quality, languages
 - Duplication: syntactic and semantic – 40% pages are duplicates
 - Others: Spider traps
- SEO, self-linking, cloaking, doorway page, or other spamming measures affects search results and ranking

Challenges of Web Search: User Behaviors

- Diverseness in access and search methodologies
 - Bandwidth and access device
 - Search, search + browse, and filter
- Wide variance in search knowledge and behaviors
 - Queries: short, imprecise, obscure, ambiguous, bad syntax
 - 85% only look over one result page
- Diverseness in search needs:
 - Informational – seek general information on a broad topic (~25%)
 - Navigational – seek a particular website or homepage (~40%)
 - Transactional – want to do something (~35%)

Why Don't the Users Get What They Want?

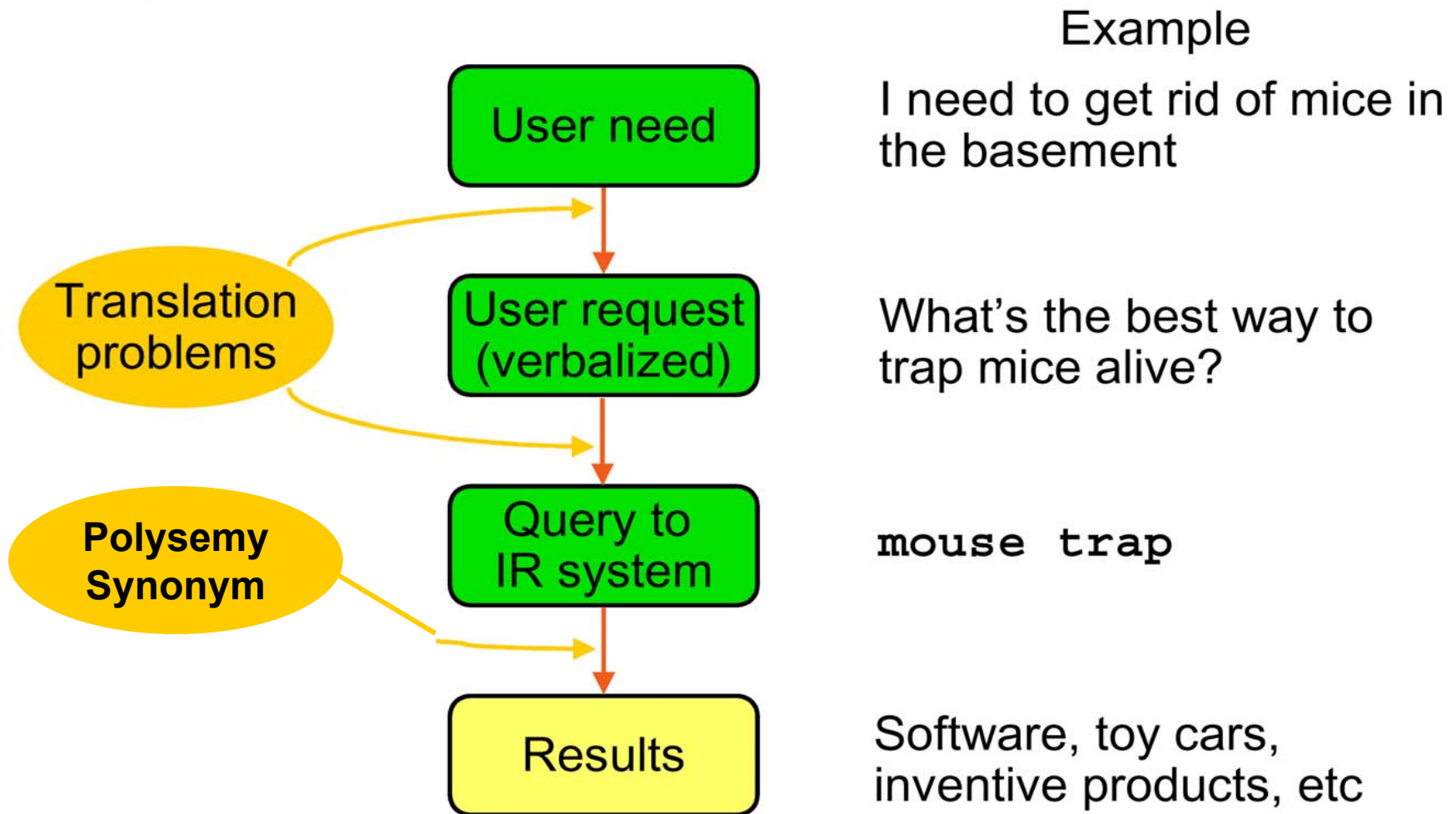
One-Size-Fit-All Solution



- User A wants to know "Apple computer"
- User B wants to know "apple as fruit"

Why Don't the Users Get What They Want?

Semantic Obscurity



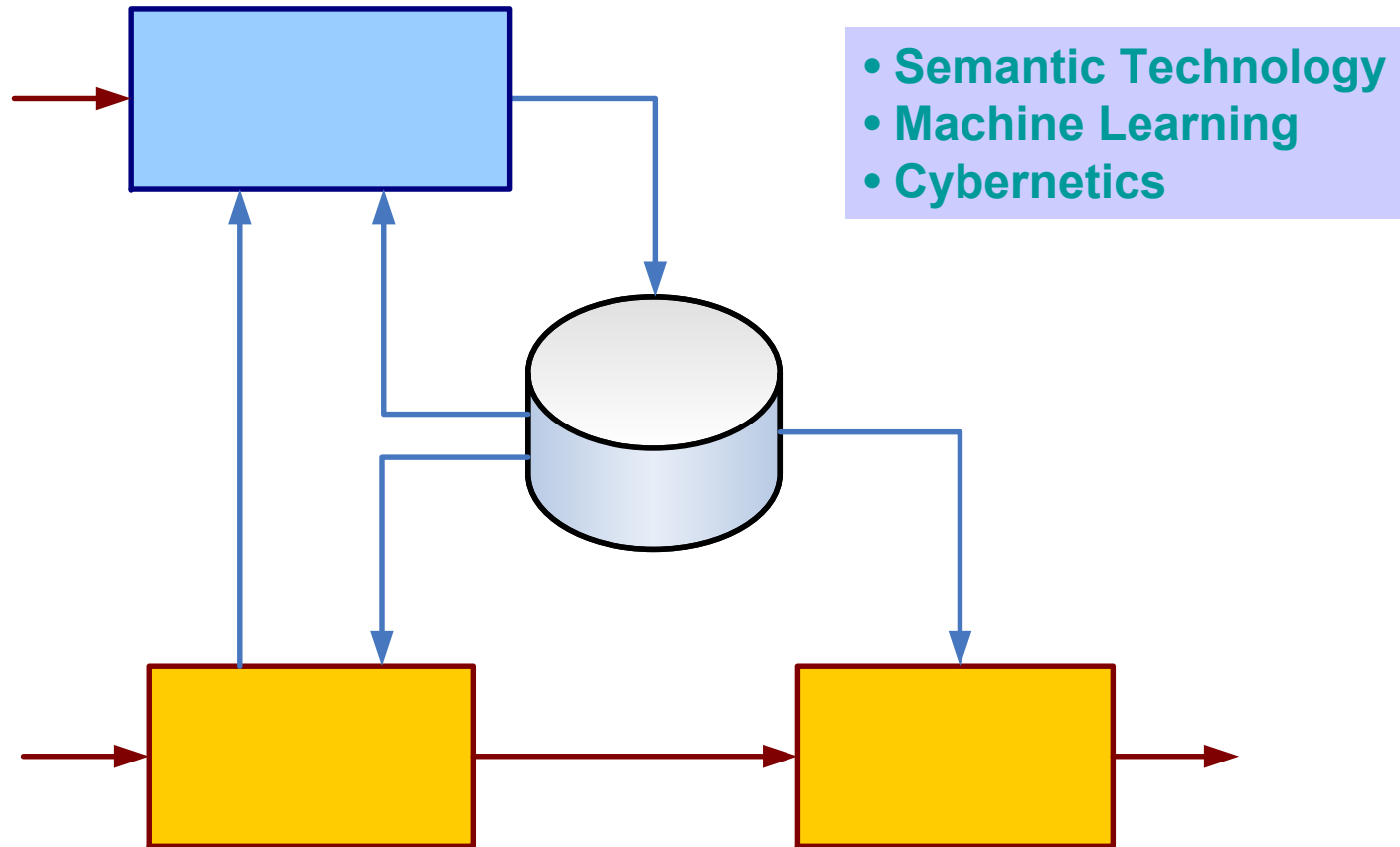
Requirements for Personalized Search: Users

- Search with simple or short queries
- Recommendations
 - Automatic query refinement or generation
 - Search executed on schedule or as needed
- Relevant quality results
 - come from credible Web sites
 - represent the most recent information available
 - remain absolutely relevant to the users
 - are top ranked

Requirements for Personalized Search: Search Engine

- Offline user profile generation
 - Based on attention data
 - It is updated on schedule or as needed
 - It is used to
 - enhance queries and results
 - resolve ambiguity in queries and results
 - determine appropriate search sources and scopes
- Real-time query processing
 - Filter and refine queries
 - Automatically generate query based on personal profile
 - Semantically and syntactically re-define and enhance queries
- Crawling and indexing
 - Smart scheduling to insure freshness of the relevant information
 - *Challenge: crawling 1B pages in a month => 400 pages per second*
- Results and ranking are tailored to meet user's need

A Personalized Search System



General Search v.s. Personalized Search

The image displays two browser windows side-by-side, comparing search results for the query "apple".

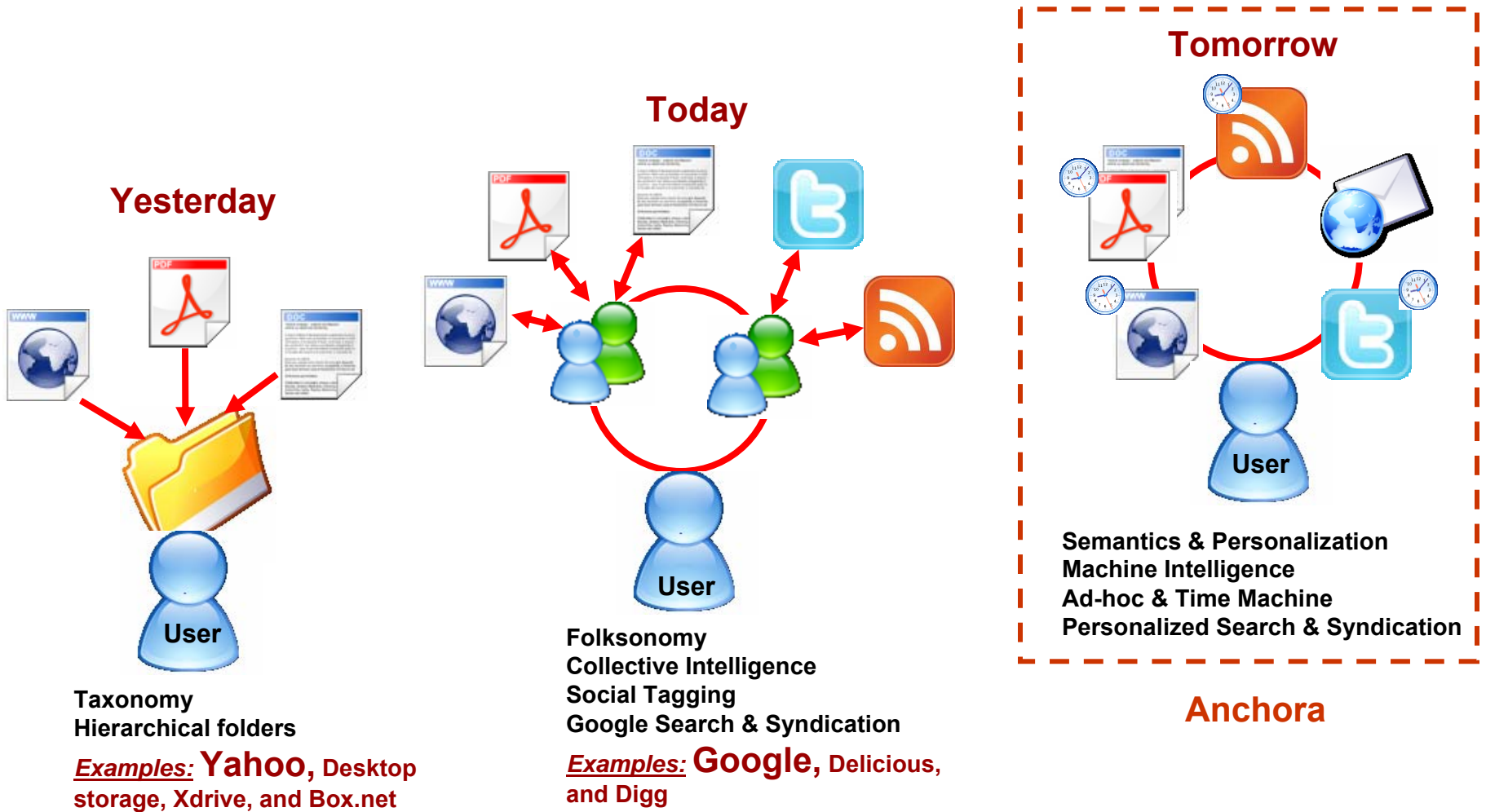
Left Window (Google Search): Shows the Google search interface. The search bar contains "apple". The results page shows a prominent "Apple" result with a description: "Apple designs and creates iPod and iTunes, Mac laptop and desktop computers, the OS X operating system, and the revolutionary iPhone." Below this are links for "Store", "iTunes", "Downloads", "MacBook", "iPhone", "Mac", "Support", and "MacBook Pro". There are also "News results for apple" and "Local business results for apple near San Jose, CA".

Right Window (Anchora Search): Shows the Anchora search interface. The search bar contains "apple". The results page is titled "Googling the Web: 'apple'" and lists several news articles from NYTimes.com, such as "Apple 'Still Pondering' Google Voice App for the iPhone" and "Apple's Management Obsessed With Secrecy". It also includes a "Google Custom Search" box and a list of "Apple" products and services.

Issues with Personalized Search

- **Relevancy & Reliability**
 - Is the attention information “clean” and up-to-date?
 - Machine-learned models prone to catastrophic errors (Google)
 - Are the search results recent and available?
 - Are the indexes for long-tailed information up-to-date?
- **Privacy**
 - Protect consumers information
 - Put the user in control of his/her information
 - The user chooses what services to receive, and in exchange for what attention information

Evolution of Information Discovery & Management



Information Utilization Is Beyond Search

Stay Connected with Information that Matters

- **Find** your relevant quality information persistently
 - Search and Browse
- **Aggregate** your information dynamically
 - Acquire and organize
- **Deliver** your information instantly
 - Retrieve and share



Internet as a Personal Information Utility

A Day with Anchora



anchora

weimlu@gmail.com ♦ (408)-921-0672

Page 20
09/09/09

Summary

- The Need for Personalization
- Personalized Information Search
- Internet as a Personal Information & Communication Utility
 - Find
 - Aggregate
 - Deliver