

Empowering People with Knowledge – the Next Frontier for Web Search

Wei-Ying Ma
Assistant Managing Director
Microsoft Research Asia

Important Trends for Web Search

- Organizing all information →
Addressing user's information need
- Intent
- Knowledge
- Semantic matching & task completion
- Searching content → Searching apps & services
- The cloud platform and developer ecosystem

Library Card Index



- **Search**
 - Paradigm: Query → Indexing → Documents
 - Query: book title, author name, ...
 - Indexing: inverted indices
 - Documents: books
- **Browsing**
 - Documents are organized into categories

The First Generation of Search Engines

- Essentially were invented to replace library card index
 - Based on information retrieval techniques
- **Search**
 - Paradigm: Query → Indexing → Documents → **Ranking**
 - Query: any words appearing in pages
 - Indexing: inverted indices
 - Documents: pages, images
 - Ranking: classical IR techniques + PageRank
- **Browsing**
 - Pages are organized into categories

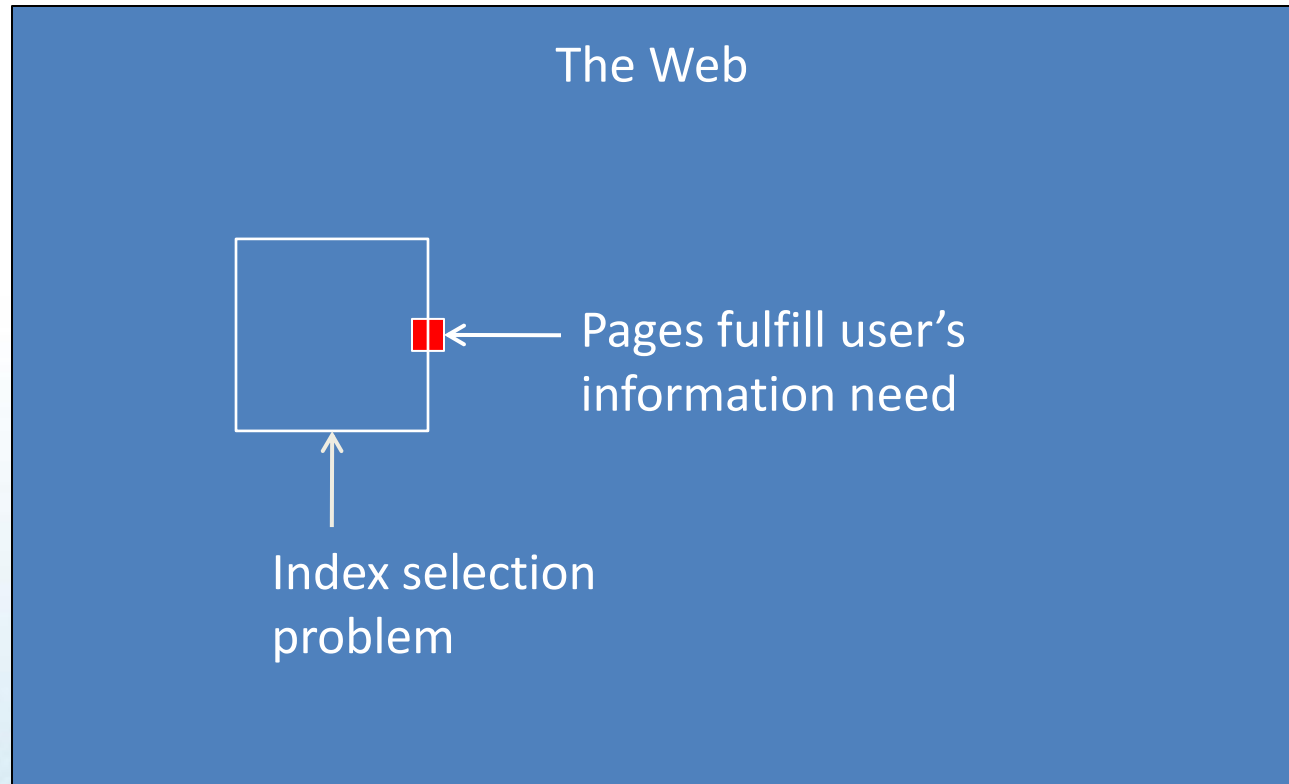


Current Search Engines

- **Search (have not changed much)**
 - Paradigm: Query → Indexing → Documents → Ranking
 - Query: any words appearing in pages
 - Indexing: inverted indices
 - Documents: pages, images, videos, books, answers,...
 - Ranking: More signals (features) are used; machine learning; log mining; human feedbacks, etc
- **Browsing**
 - Authoritative pages are organized into categories
- **Challenges: information explosion and information overload**
 - Index selection, index quality, and freshness
 - Relevance ranking (10 blue links)



Organize all information → Address user's information need



The explosive growth of the Web vs.
The relatively slow growth of human population and their time spent in search

Empower People with Knowledge

Enable people to gain knowledge and creativity from the web by computationally understanding user intent and matching that with published content, apps and services

- **Intent**
 - Computationally understand what the user is trying to accomplish
 - “Knowing” user needs, attitudes, and desires enables us to help the consumer better enrich their lives
- **Knowledge**
 - Computationally distill concepts and entities – such as people, places, products, businesses – and the relationship between them
 - Enable people and businesses to derive insights and knowledge from the web, and take actions
- **Semantic Matching and Task Completion**
 - Routing intent to task (not only content, but also apps & services)

Markets

Languages

...

OEM's

App's

UX

Ranking & Intent

Indexing & Knowledge

Domains

Tasks

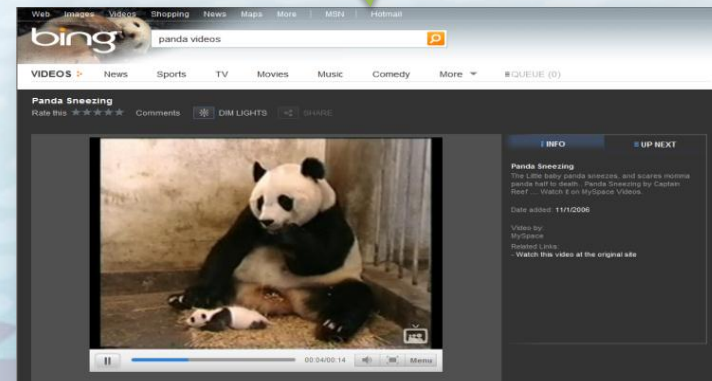
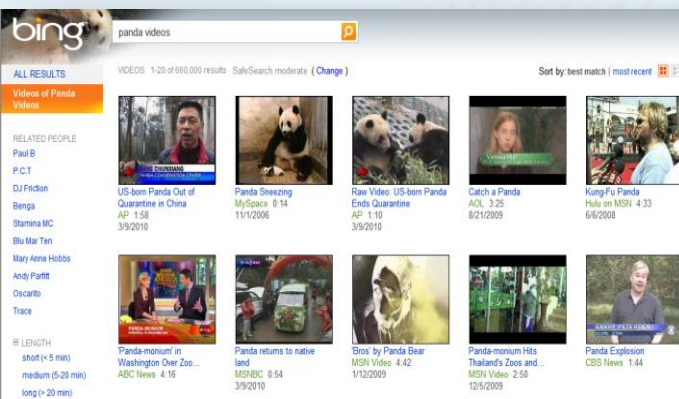
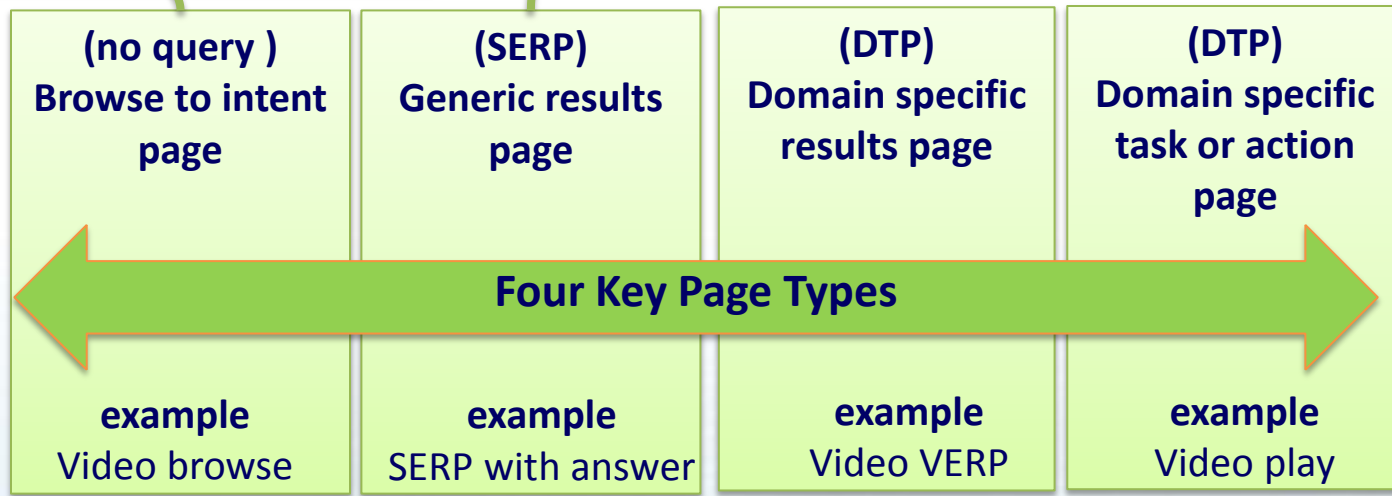
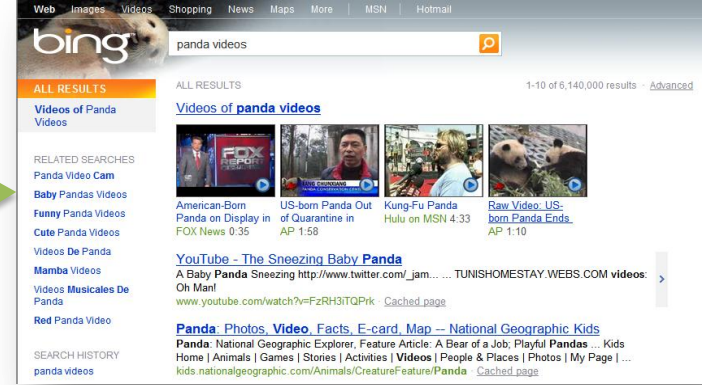
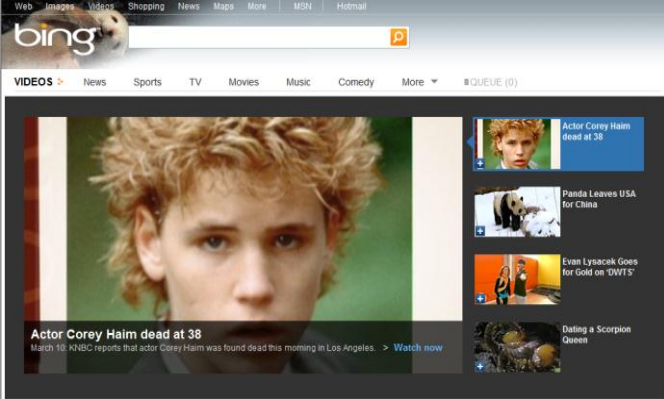
EcoSystem

1st
page

2nd
page

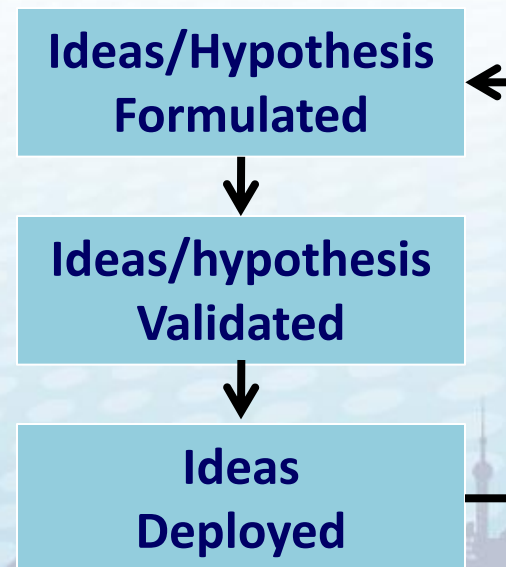
Search Platform

Infrastructure

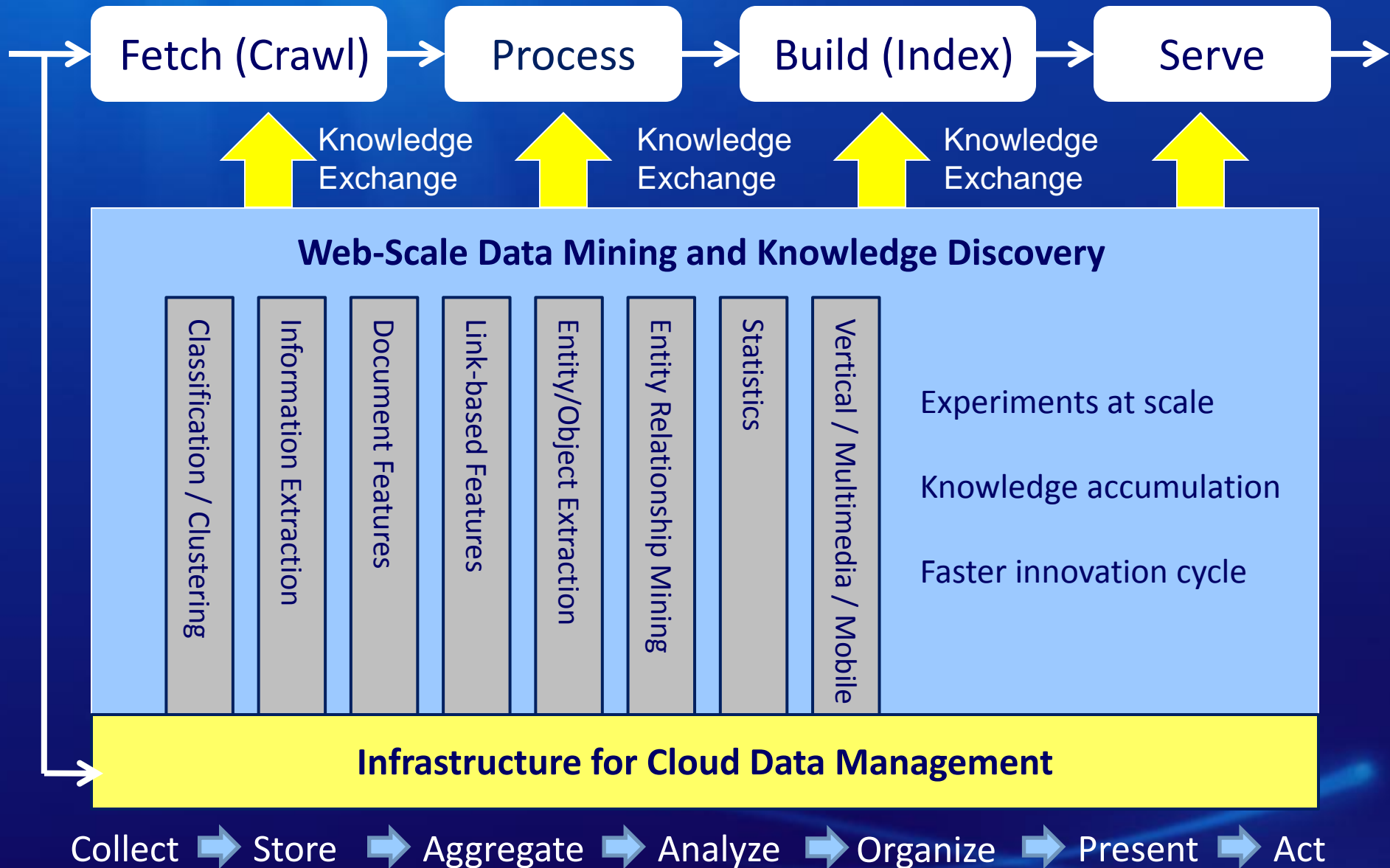


Infrastructure for Web-scale Data Mining and Knowledge Discovery

- **Deep understanding of data**
 - Data -> Information -> Knowledge & Intelligence
 - Queries -> Intent
 - Users -> Audience Intelligence -> Personalized & Targeted
- **Experiments at scale**
 - Offline experiments
 - Online experiments
 - Fast cycle of innovation



Search Infrastructure + Data Mining



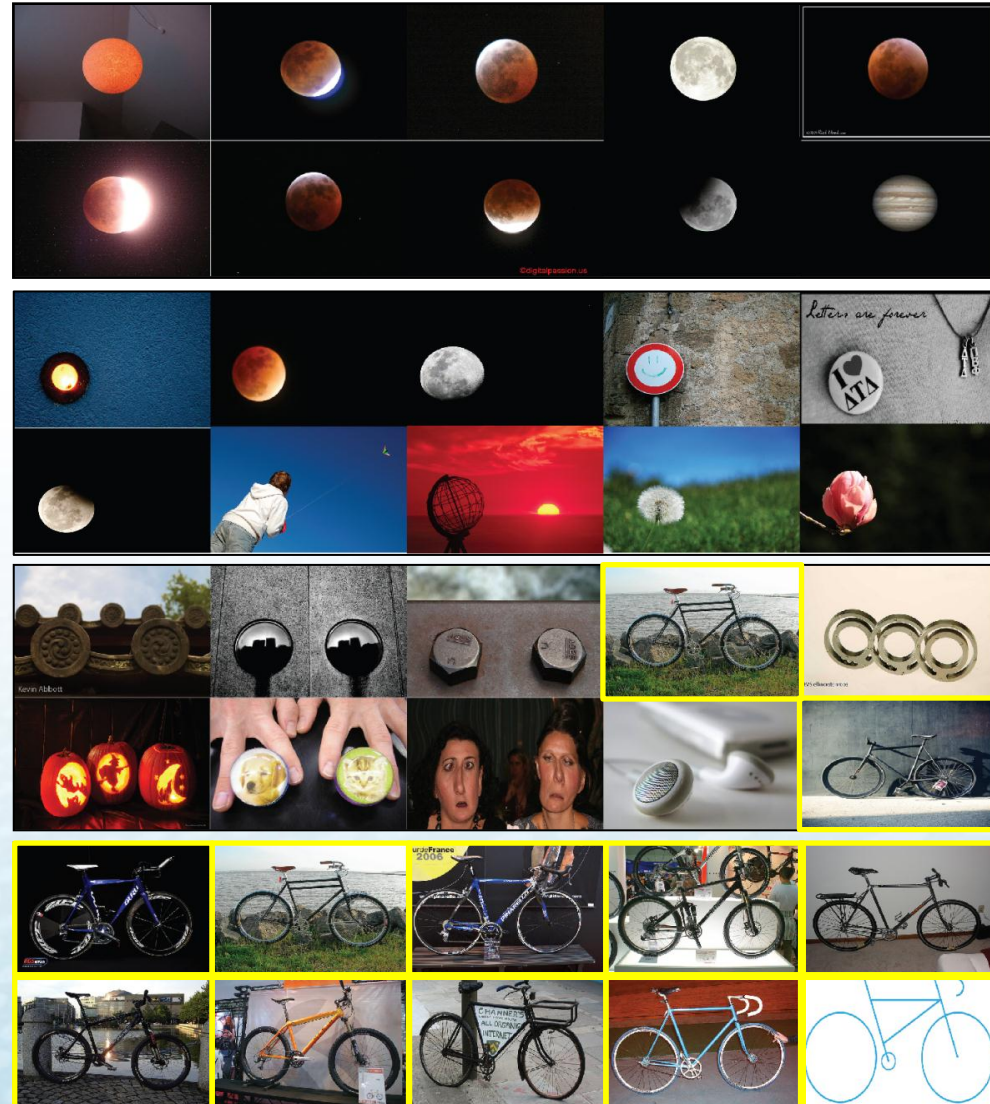
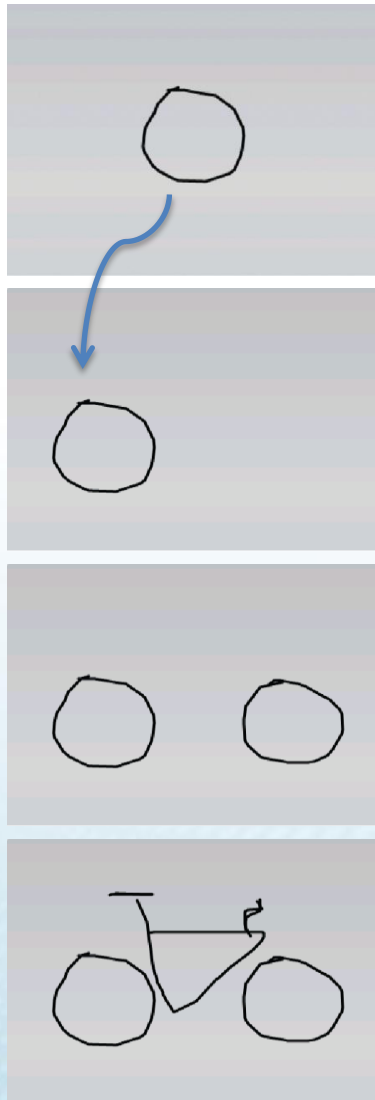
From Web Pages to Web Entities

- Entity search and knowledge mining
 - Web-scale entity extraction, integration, and summarization
 - Entity relationship mining
 - Entity ranking
- Academic search as an example
 - Researchers, papers, organizations, conferences, journals
 - Knowledge and insights
 - Visualization & exploration

DEMO: MICROSOFT ACADEMIC SEARCH

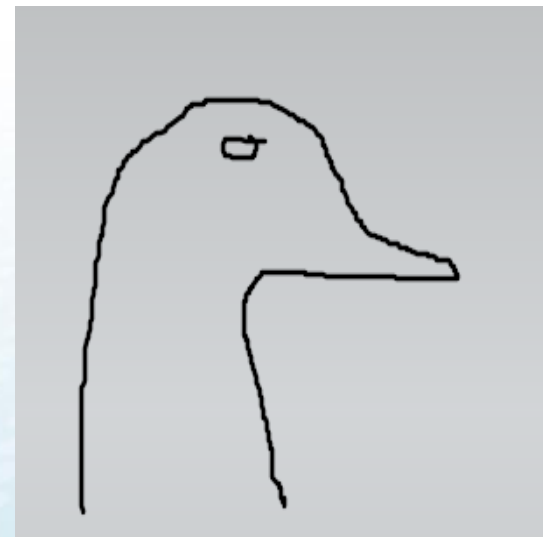
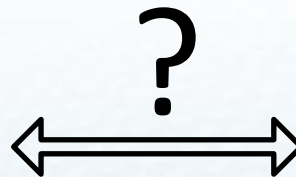
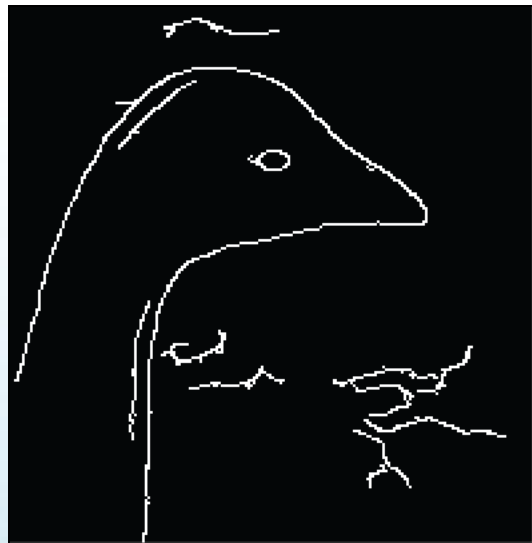


Sketch out Your Search Intent



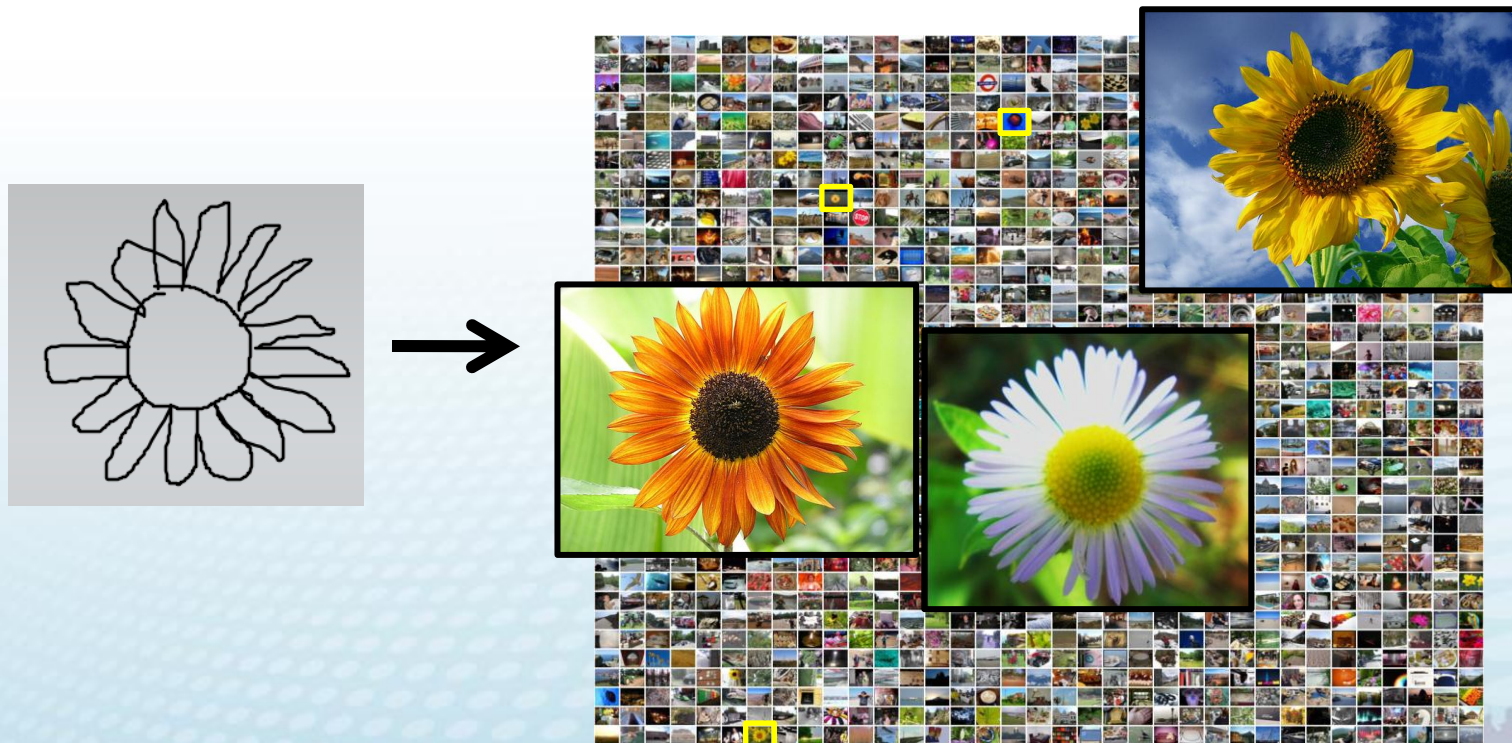
Challenge #1

- The gap between a natural image and a query sketch



Challenge #2

- Scalable solution and efficient semantic indexing to support real-time search in a database of **billions of images**



DEMO: MINDFINDER



Important Trends for Web Search

- Organizing all information →
Addressing user's information need
- Intent
- Knowledge
- Semantic matching & task completion
- Searching content → Searching apps & services
- The cloud platform and developer ecosystem

The Emergence of the Cloud

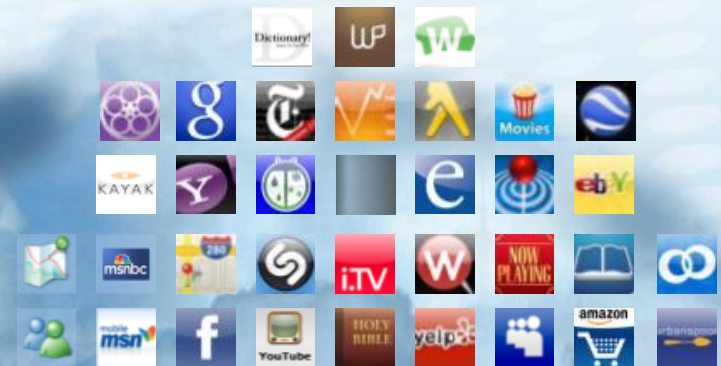
- Software as a Service
- Platform as a Service
- Infrastructure as a Service
- **Information and Knowledge as a Service**



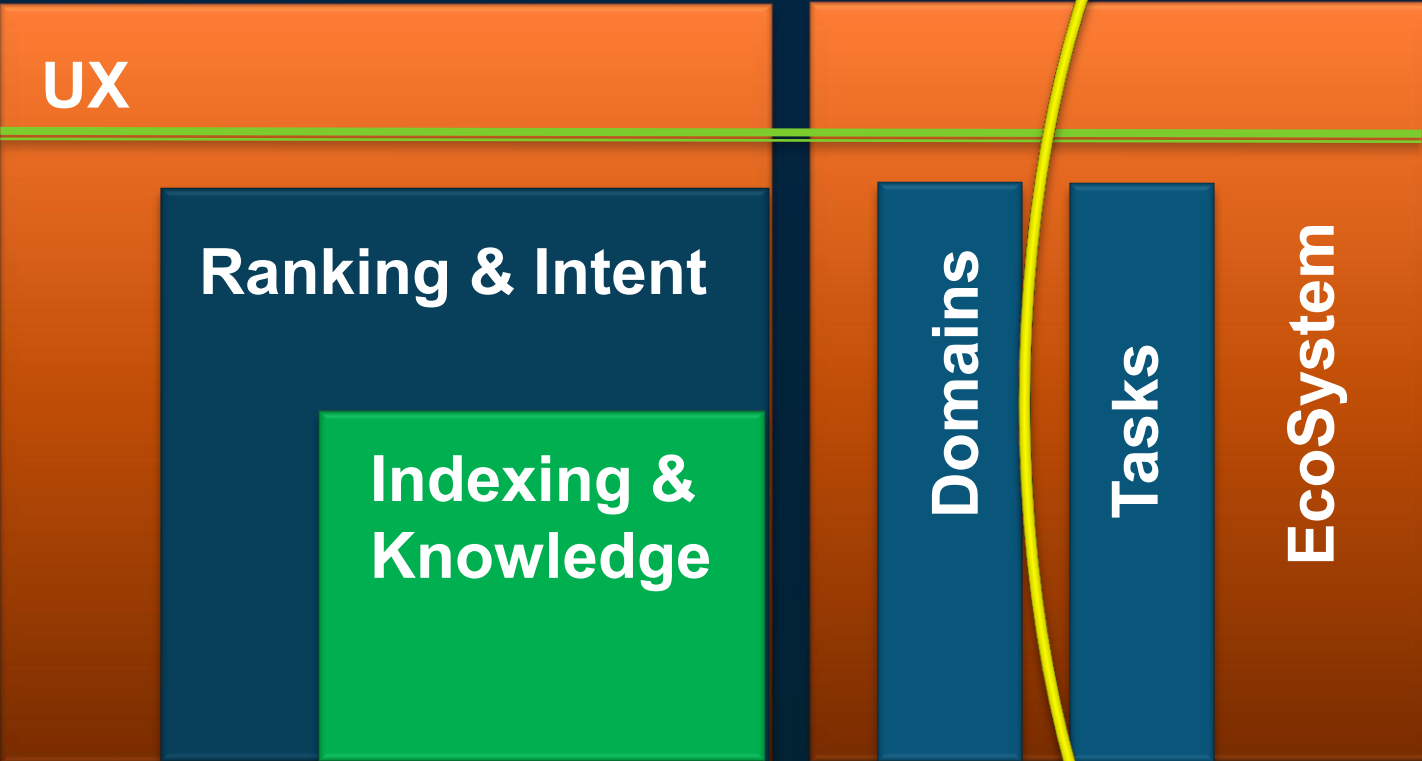
Open Ecosystem for Search

(Platform + Infrastructure + Information/Knowledge)
as a service to developers

- 1-2 developers can build a micro-vertical app and web service to help with users on a specific task
- Millions of apps are easily discoverable and searchable at apps market place
- search engine routes intent to task (apps)



Markets Languages ... OEM's App's



1st page

2nd page

Search Platform

Infrastructure



Evolution of the Web

Topical

Social

Geo-Spatial



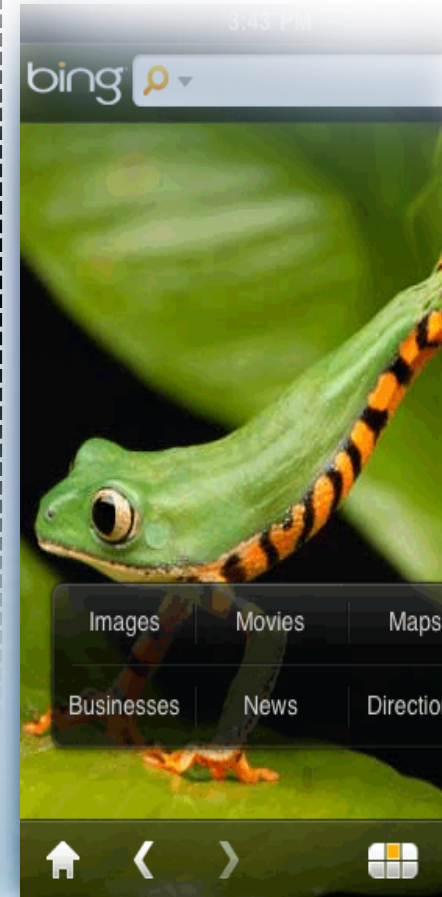
Content & HTML Documents



People & Profiles



Places & Maps



Services & Applications

Intent and Knowledge

- Explicit signals beyond a single query
- Implicit signals from the broader context (social, geo, app, camera...)
- Dialog-based

Intent Understanding



- Move from “a bag of words” to connecting dots (entities)
- “tasklet” and Query Store authoring + entity ecosystem
- Personal content

Knowledge Construction



- Intent-knowledge matching in the social context
- Signed-in cross-screen multimodal experience
- Intent-driven ad + search + browse

Personalization



Understanding

Construction

Personalization

Make Search Actionable

- Search is mostly based on “a bag of words” method
- Statistical and super scalable (breadth)
- Document-centric

Search



- Task/app has deep engagement but not as scalable as search (depth)
- Authored with schemas and entities like NLP
- Action-centric

Tasks



- Marry search and tasks seamlessly (breadth + depth)
- Alleviate the coverage challenges in NLP
- Enhance the flow of tasks with recommendations and social

Search +
Tasks



Re-Organizing the Web for Task Completion

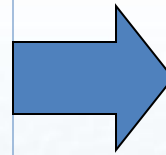


What's Next for Search?



What's Next for Search?

- Organize the world's information
 - Relevance ranking – 10 blue links

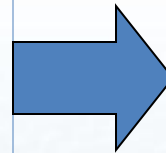


- Directly address user's information need
 - User centric innovation
 - Answers & tasks
 - Understand the query space and organize knowledge according to query space (instead of document space)

What's Next for Search?

- Information

- Search content

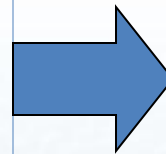


- Knowledge

- Entity (people, places, things) and concepts, and relationship between them
- Search apps & services

What's Next for Search?

- Keyword matching
 - Inverted index

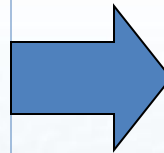


- Understand intent
 - Query understanding
 - Natural language
 - User's context and history
 - Intent modeling
 - Semantic indexing & matching

What's Next for Search?

- Search engine

- Get the relevant information (a website)
- Get out of SERP with a simple click
- Challenge: query – URL matching



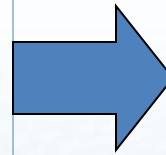
- Decision engine

- Complete the task by fulfilling user intent
- Exploring search results by clicking & browsing
- Whole page relevance
- Search interaction model (dialog)

What's Next for Search?

- **Archived Web**

- Offline mining & knowledge discovery

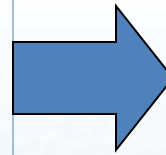


- **Real Time Web**

- Analyzing streaming text data such as tweets

What's Next for Search?

- Software

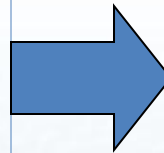


- + Hardware
 - More advanced index serving using hardware acceleration

What's Next for Search?

- **Close**

- Internal engineers
 - Experimental platform
 - Experimental infrastructure
 - Shared data and storage

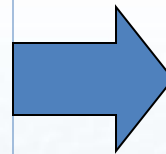


- **Open + Ecosystem**

- 1st and 3rd party developers
 - Platform as a Service
 - Infrastructure as a Service
 - Information and knowledge as a Service (Web data and meta data)

What's Next for Search?

- Impression-based advertising
 - Pay per click

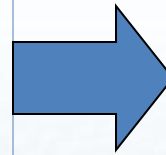


- Transaction-based advertising
 - Deeper understanding of user's intent
 - Routing intent to apps or services for task completion

What's Next for Search?

- **Web Graph**

- Links between web pages
- PageRank



- **Cloud Graph**

- Information cloud
- Social cloud
- Communication cloud
- Entertainment cloud
- Productivity cloud
- Commerce cloud
- Fusion & graph mining

Summary

- Organizing all information →
Addressing user's information need
- Intent
- Knowledge
- Semantic matching & task completion
- Searching content → Searching apps & services
- The cloud platform and developer ecosystem

THANK YOU

