

Microsoft® Research

# Faculty Summit

10  
YEAR ANNIVERSARY



# Bing: User Intent and Decision Engine

Harry Shum, PhD

Corporate Vice President  
Microsoft Corporation

# Overview

- Why Decision Engine
- Bing Demos
- Search Interaction model
- Data-driven Research Problems
- Q & A

# Opportunities for Search Innovation

1

Imprecise  
Results

2

Research  
Sessions

3

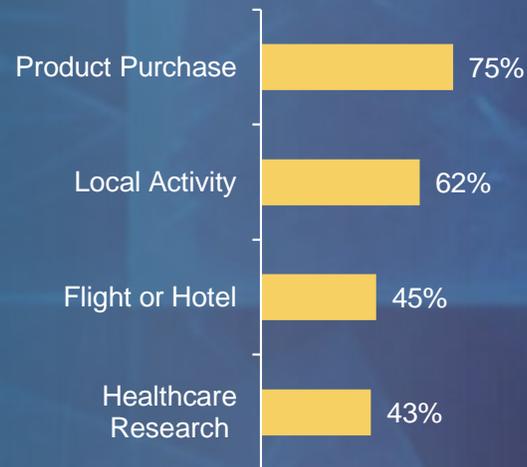
Inform  
Decisions

# Turning to Search to Inform Decisions

## Users Relying More on Search

Q. In the past six months have you used a search engine to help inform your decisions for the following tasks?

**66%**  
of people are using search more frequently to make decisions



## Decision Sessions are Lengthy

Length of Sessions by Type



Users need help with tasks and making decisions

Complex task and decision sessions could be easier



# Demos

Frederick Savoye

Senior Director  
Bing Product Marketing

# Search User Interaction Model

# Search User Interaction Model

- Search engine
  - Objective: getting the user relevant information (a website)
  - Model: getting out of search results page with a simple click
  - Challenge: query – URL matching
- Decision engine
  - Objective: completing the task by fulfilling user intent
  - Model: exploring search results by clicking and browsing
  - Challenge: whole page relevance

# Bing interaction model

- Explore pane (or left rail)
  - TOC: verify user intent
  - Related search: expand user intent
  - Search history: remind user intent
- Task completion
  - D-cards: showing structural information for the site
  - Hover: showing more information for the site
  - Simple tasks: e.g. Fedex tracking number
  - ...

# Data-driven Research Problems

# Some Challenging Problems for Search

- A lot of data...
  - Billions of query logs, documents, pictures, clicks, etc.
  - Processing them is costly and takes time
- Statistical learning + distributed computing
  - Can we train 1 Billion samples (query – URL pairs)
  - Within a few hours? No over-fitting?
- Two examples of Bing-MSR
  - “Bing-it-on” N-gram
  - Statistical model for log mining

# Statistical Language Model for Search

- LM Applications in Search
  - Query processing: alterations, expansions, suggestions
  - Document processing: classification, clustering
  - Matching and ranking
  - Powerset.com
- Leading technology: N-gram
  - $P(\text{next word} \mid N-1 \text{ preceding words})$
  - Better “understanding” = model predicts better

# Challenges to build N-gram for Search

- High quality model needs lots of data at web-scale:
  - Billions of documents, trillions of words, PetaBytes of storage
- Smoothing:
  - How to deal with a very long tail
- Freshness:
  - Web contents are added and revised rapidly
- Bing-MSR innovation: Constantly Adapting LM (CALM)
  - Highly parallel algorithm designed for cloud computing
  - Refine LM as soon as documents are crawled

# "Bing-It-On" Ngram Services

- We are sharing our resources
  - For details, go to <http://www.facebook.com/microsoftresearch> and follow "Bing-It-On Ngram"
- Compare Bing-It-On with Google's Release

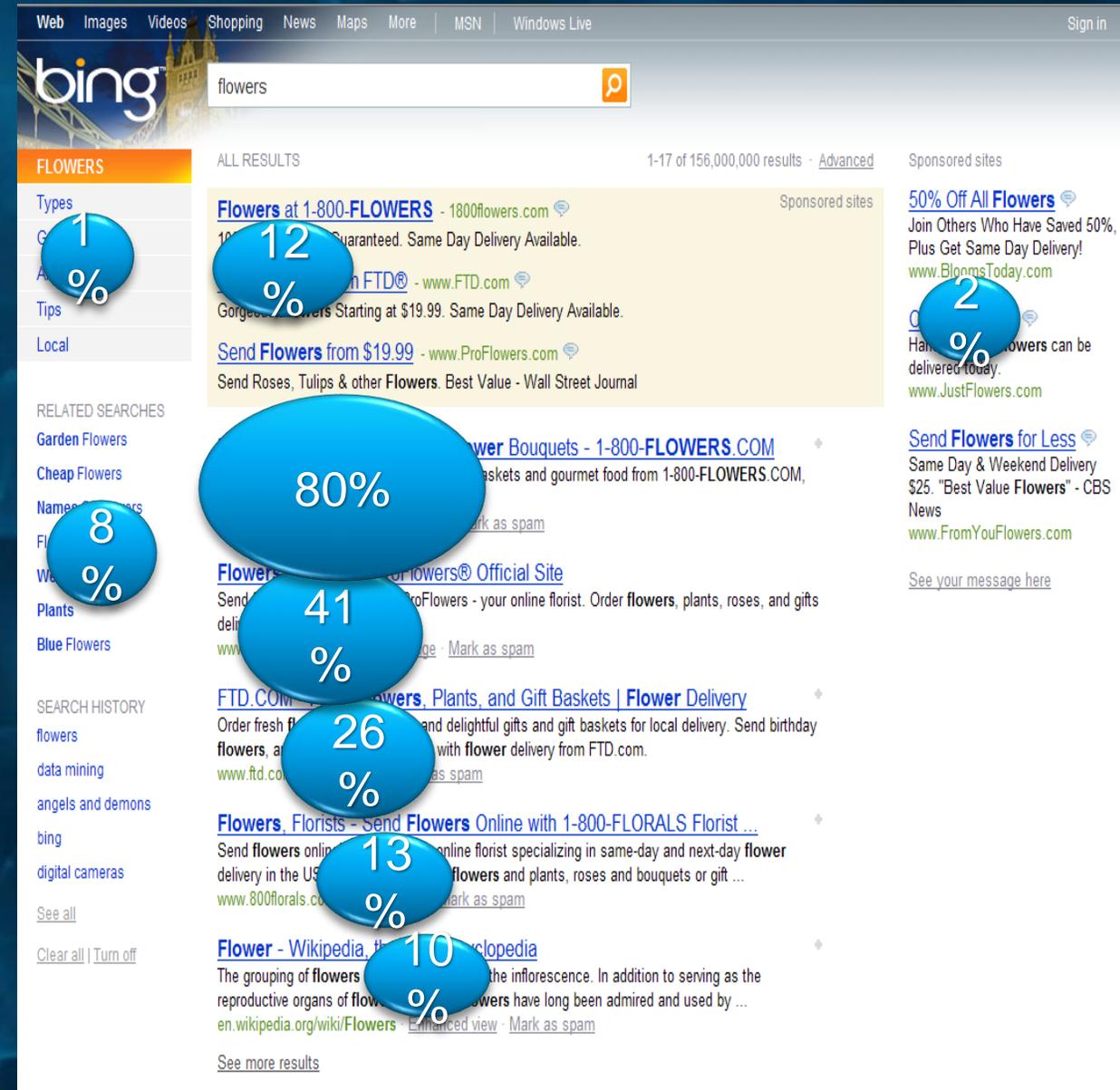
	Google	Bing-It-On Ngram
<b>Content Types</b>	Document Body only	Body, Title, Anchor texts
<b>Model Types</b>	Raw Count only	Count and smoothed models
<b>Highest order N</b>	5	5
<b>Training Size (Body)</b>	~ 1.0 trillion words	> 1.3 trillion
<b># of 1-gram (Body)</b>	13 million	1 billion
<b># of 5-gram (Body)</b>	1 billion	237 billion
<b>Availability</b>	DVDs from LDC	On demand web services hosted by MS
<b>Update</b>	September 2006	Monthly

# Log Mining for Search

- Data driven decision is first class citizen in search
  - Toolbar and search logs: ~10 TB/Day each
- Bing uses log mining to
  - Understand what users want
  - Assess how we are doing
  - Quickly improve Bing
    - Query Processing
    - Ranking
    - User experience
- Examples:
  - Relevance inference using Bayesian Browsing Model (BBM)
  - User behavior understanding with Hidden Markov Model (HMM)

# Mining search log with HMM

- Search log records only clicked results
  - Skipped results are not explicitly recorded
- Hidden data mining
  - Model viewed results as Markov chain
  - Skipped results = hidden data
- How well does it work?
  - Very close to eye tracking data



# What's In the Name?

approachable

can be used as a verb

global



easy to spell

memorable

the Sound of Found

fresh



**Thank you!**