Image Search Live
http://images.live.com

Nick Craswell
Microsoft Research Cambridge
Problem: Text Query to Image Grid

Queries: e.g. a 35 day log
- Most frequent: 200k/794k
- Median: 2 occurrences
- Popular queries: lindsay lohan, scarlett johansson, angelina jolie, sex, jessica simpson, kate beckinsale, paris hilton, britney spears, shakira, sexy, jessica alba, jennifer lopez
  - Around 60 of the top 100 queries are adult or celebrity
- Random queries: bridge, rodolfo font, playboy, douwe egberts, jesus, tanning, beauty, oakenfold, priyanka chopra, ACTORS

Results: Image thumbnail grid

Image Search Live

I'm responsible for the ranking: 200 parameters, tuned via gradient descent (RankNet)
  - Features (nearby text, basic image features, misc)

This talk:
A. The user interface
B. Visualising the click graph
C. The importance of crawl quality
A: The User Interface

• Around 75% of Web search result pageviews are page 1.
• For image search it is 43%
• The 75% threshold in image search is reached around page eight
“Infinite scroll”
- Just: scroll

Resizable thumbnails
- Just: zoom

Metadata hiding

Filmstrip View and “Show Image”
B. Visualising the click graph
C. The Importance of Crawl Quality
Normalised Discounted Cumulative Gain (NDCG)

Gain [ green day ]

- 2 points
- 1 point
- 0 points
- 0 points

Discount

\[
\text{DCG} = \sum_r G_r \cdot D_r
\]

Cumulative over ranks

Normalise according to max possible DCG

NDCG distributions of two systems
What if we had perfect ranking?

Summary

• Interesting UI, motivated by click logs
• Click graph gives an indication of overall user needs
• Crawl quality is important

http://images.live.com