

Evaluation

An encore

Evaluation

I found a way to
make the results
better

No you didn't



Evaluation

“google founder”
should return my
website as #1

No it should
return my site



Evaluating Interfaces by Bucket Testing



Evaluating Search Results by Human Labelers and nDCG

1. Humans assign relevance scores to query × document
2. Compare nDCG scores for that query between different search systems (or versions)

nDCG Example

EVALUATION REVIEW

PRESENTATION

BREAK

ADS OVERVIEW

- how ads work
- 3 stakeholders

- 1) users -- want relevance ads
- 2) advertisers -- want cheap good traffic
- 3) publishers -- want to make money and please users

ECONOMICS OF SEARCH ADVERTISING

AUCTION

ebay auction (overture)

- bad experience for advertisers

generalized second price auction

- pay price of what the ad is worth to the next highest bidder
- nash equilibrium, but allows for cheap bad ads

modified generalized second price auction

- multiple by quality score, sort by value
- quality score based on CTR and landing page metrics
- $price_1 \times q_1 = \max bid_2 \times q_2$. solve for $price_1$
- when quality goes up, price comes down

QUESTION AND BREAK

- how many ads should a search engine show?

CLICK FRAUD

advertisers click competitors' ads, publishers click their own ads

- to detect, model deviations from typical behavior distributions
- how do you know how much fraud you catch?

- compare search ads vs publisher ads, track conversions

ARBITRAGE

- making money off ad price differentials

same query, different price

different queries, aggregator landing pages

- make money when landing page CTR \times earnings $>$ cost