Ranking for Research

Access to this document is restricted to specific people in Research who are working actively to improve search, not people with a general interest in the topic.
Purpose

- People in Research are now making major contributions to Google search, delivering benefit to hundreds of millions of people around the world.

- Deepening this collaboration would be awesome for everyone! This talk provides some context around work on web ranking to further that goal.

- Work on we ranking has involved hundreds of people over 20 years, so this presentation is far from comprehensive. Please add comments on confusing points or areas where more information would be useful.
Outline

- Be careful in discussing search
- Get a feel for search
- Great search, not great metrics
- The structure of ranking
- Approaches to relevance
Be careful in discussing search.

Having a big, positive impact on the world is a joy of working on search.
But this impact makes our words and actions carry heavy consequences.
So we must tread carefully...
Be careful in discussing search.

- Attempts to manipulate search results are continuous, sophisticated, and well-funded. Information about how search works should remain need-to-know.
- We are sued often. All emails involving people in Search are likely to be retained in connection with multiple, ongoing lawsuits.
- Internal communications have been repeatedly leaked to the press and cast in a bad light. Around the world, we endure multiple negative press cycles weekly.

Email threads involving search product counsel in my inbox this month.
My email and docs are currently retained in connection with four different lawsuits.
"Tucker Carlson raised the question Thursday night, reporting exclusively on internal Google emails in which employees discussed potentially manipulating the search engine's results to undermine President Trump's travel ban..."
Be careful in discussing search.

- The EU fined Google 2.2 billion Euros in connection with an aspect of search. People you work with were involved.
- Search issues can inflame world leaders who have power over Google, demand Congressional hearings, etc.
- We never know when the next regulatory action or subpoena will hit.

Tweet from the US house majority leader.
Facebook turned over some internal communications as part of a lawsuit.

Those communications were then seized from the opposing party by the serjeant-at-arms of the House of Commons.

Never saw that coming!
Particularly Sensitive Topics

- Keep talk about how search works on a need-to-know basis. Everything we leak will be used against us by SEOs, patent trolls, competitors, etc.

- Discuss social fairness in search results with great caution. This is an important topic that we must all take very seriously. But how will your email read to reporters, regulators, and litigants?

- Avoid statements that might be perceived as anticompetitive, e.g. why aren’t we ranking this Alphabet-related product higher? How can we knock down such-and-such site? We do not do such things.
Sensitive Topics

- Do not discuss political bias in search in writing. This is another important topic that we take seriously, but is just too volatile for email.

  Postscript: Two weeks after this talk was initially given, Google’s CEO was called to testify on this topic before Congress.

- Do not discuss the use of clicks in search, except on a need-to-know basis with people who understand not to talk about this topic externally. Google has a public position. It is debatable. But please don’t craft your own.
Search Quality Has Many Aspects

- Relevance
- Page quality
- Popularity
- Freshness
- Localization
- Language
- Centrality
- Topical diversity
- Personalization

- Web ecosystem
- Mobile friendly
- Social fairness
- Optionalization
- Porn demotion
- Spam
- Authority
- Privacy
- User control of spell correction

Capturing everything in a metric is tough!
Some Known Shortcomings of Live Traffic Eval

- The association between observed user behavior and search result quality is tenuous. We need lots of traffic to draw conclusions, and individual examples are difficult to interpret.
Basic Control Flow

Search Query

↓

Query Rewriter
interprets query

↓

Muppet / Ascorer
retrieve and score documents

↓

Superroot
adjust ranking

↓

Web Search Results

Web Documents

↓

Crawl & Indexing
prepare for fast retrieval and scoring