Find What I Mean, Not What I Type

Chris Sherman

Executive Editor, SearchEngineLand.com

Semantic Search

search engine land

How we search today

search engine land

What we really want

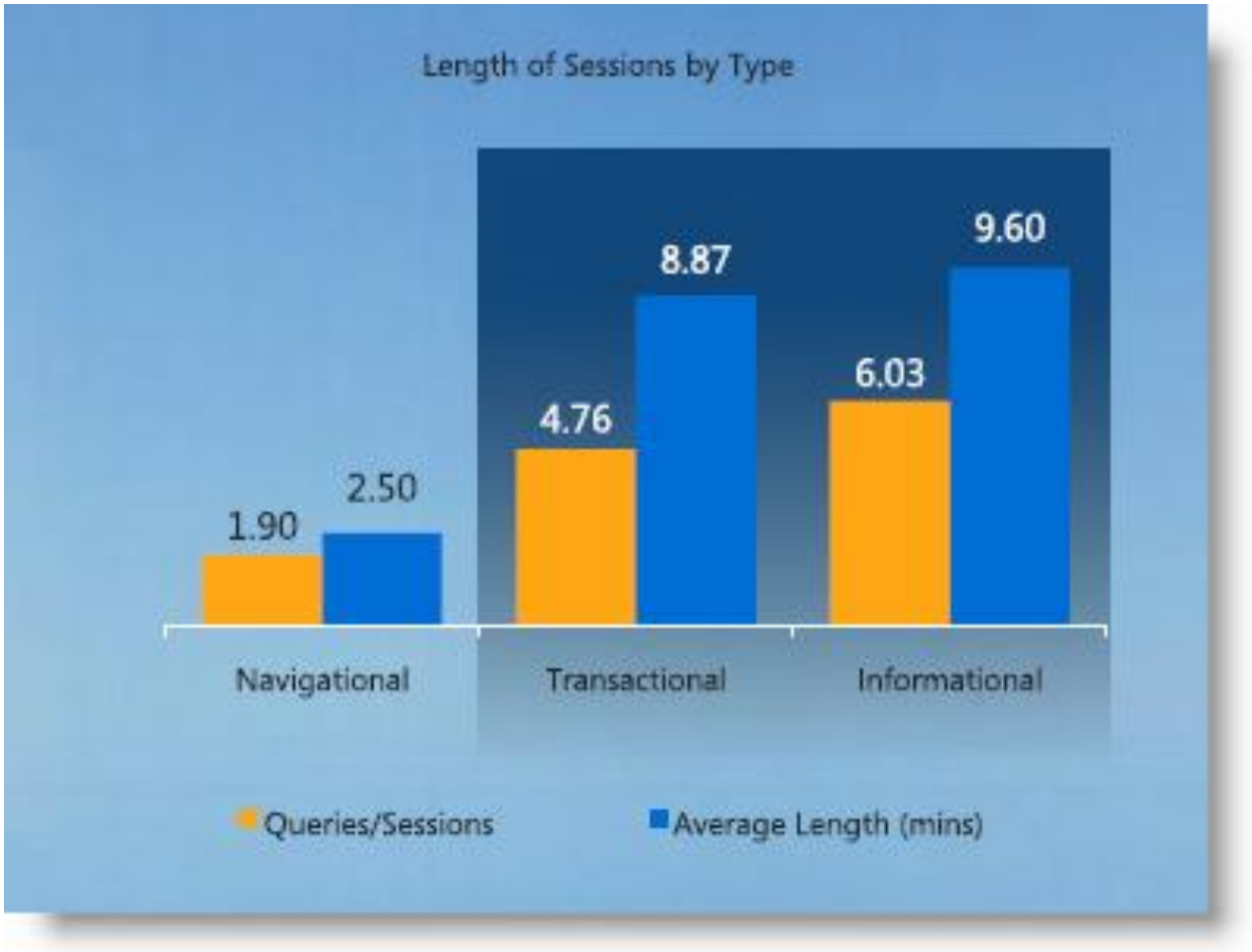
bing™

Google™

YAHOO!®

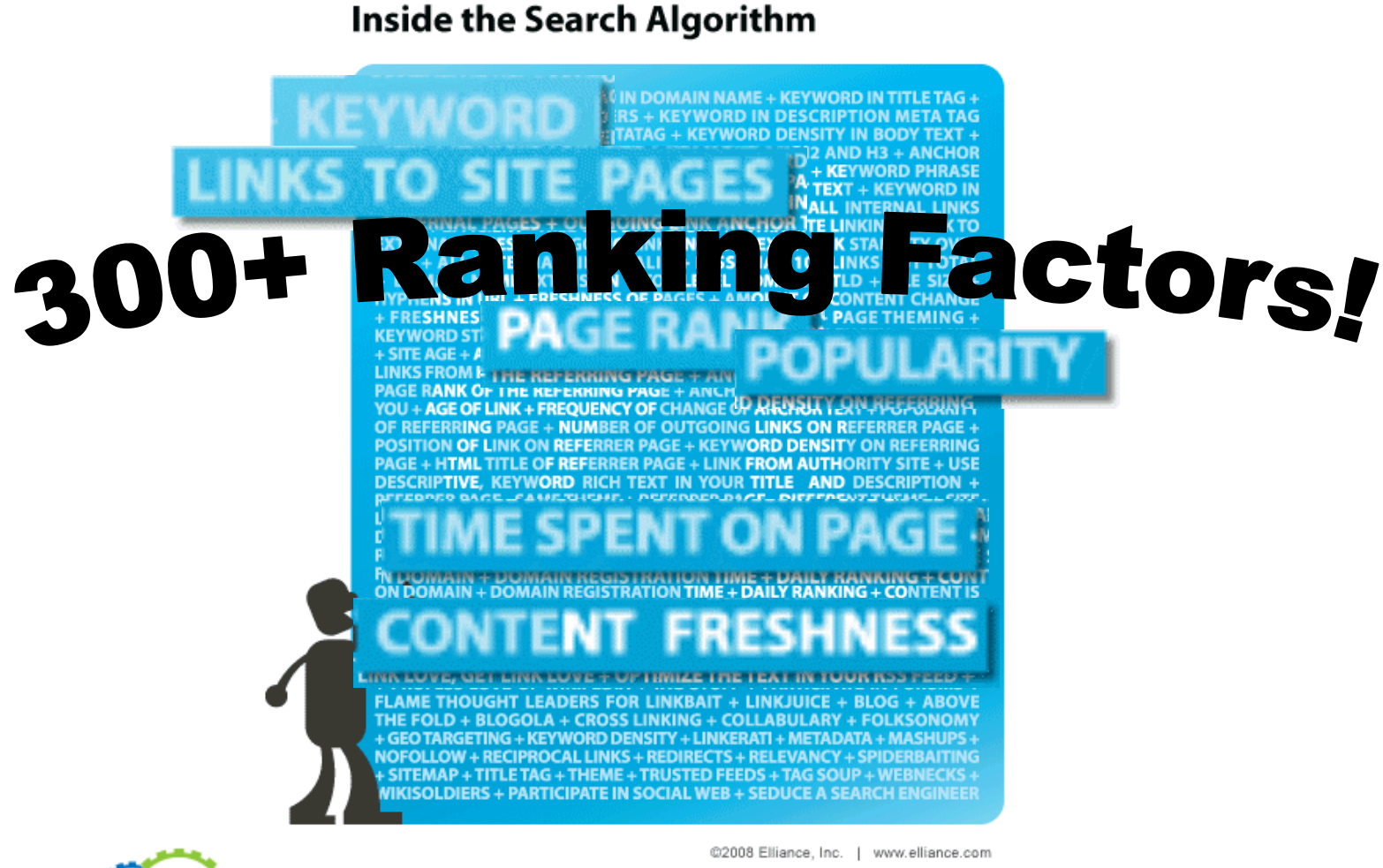
search engine land

What we're searching for



Source: Microsoft Internal Research

How search works today



search engine land

Current search algorithms

- **Rely heavily on text pattern matching, statistics, and observations of searcher behavior**
- **Some semantic analysis, but it's limited**
- **What's needed: True understanding of language and meaning**

search engine land

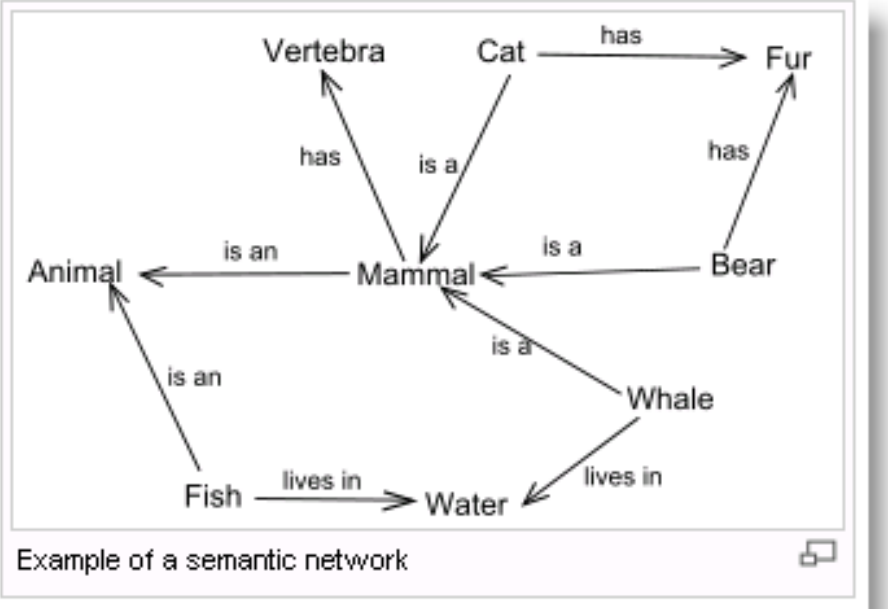
What is “semantic search?”

- **Focus on delivering *answers*, not *search results***
- **Key: Disambiguation – of both queries and content on the web**
- **Heavy emphasis on natural language processing and understanding *meaning* and *intention***

search engine land

The technical definition

- **Semantic search uses a semantic network to map meanings & relationships of words**



Example of a semantic network

search engine land

Why semantic search now?

- **Building an accurate semantic network is time consuming and costly**
- **Semantic search requires *lots* of computing power and storage**
- **Until recently, traditional web search techniques were perceived as "good enough"**

search engine land

Google Data Center, 1999

search engine land

Google Data Center, 2008

Source: New York Times

search engine land

What semantic search is not

- **Semantic search is not what is traditionally known as NLP**
- **Semantic search is not a replacement for navigational queries**
- **To a lesser extent, semantic search is not a replacement for transactional searches**

search engine land

Semantic search ≠ semantic web

- **The semantic web is Tim Berners-Lee's vision of "web 2.0"**
- **The semantic web uses extensive metadata and the Web Ontology Language (OWL), making it possible for the web *itself* to "understand" and satisfy the requests of people and machines to use web content**

search engine land

Caution!

- A number of players offering "semantic" search are just refining search results into categories
- Northern Light and Vivisimo pioneered these efforts years ago
- Not "true" semantic search

search engine land

A Look at Some of the Players

- **Wolfram Alpha**
- **Powerset ***
- **Hakia**
- **TrueKnowledge**
- **Kngine**
- **GoPubMed**
- **DeepDyve**

search engine land

Wolfram Alpha

- **Strengths**
  - **Can combine lots of disparate data types – extracts meaning & relationships on the fly**
- **Weakness**
  - **Relies on "curated" data – "calculates" implied semantic relationships based on a limited data set and its own natural language engine**

search engine land

Powerset

- **Strengths**
  - **Better, more comprehensive view into Wikipedia & Freebase**
  - **"Factz" great for ready reference queries**
- **Weakness**
  - **For now, appears to work best on small, relatively structured corpora**

Hakia

- **Strengths**
  - **For some queries, results are "resumes"**
  - **"Credible sites" recommended by librarians**
- **Weakness**
  - **Tends to default to "galleries" on broad but unambiguous queries**
  - ***Really* bad results for some queries (but it's in beta…)**

search engine land

True Knowledge

- **Strengths**
  - **Probably the first true "answer engine"**
  - **Also truly a semantic web application thanks to its API returning structured results to machine-based queries**
- **Weakness**
  - **Still in beta – will it really work as a public service?**

search engine land

Kngine

- **Strengths**
  - **Very good at aggregating structured data**
  - **Also good at disambiguation**
- **Weaknesses**
  - **Relies heavily on Freebase for facts; can be incorrect or out of date**

search engine land

Transinsight (GoPubMed.org)

- **Strengths**
  - **Excellent disambiguation of queries and categorization of search results**
  - **Identifies foremost authorities in particular subject areas**
- **Weaknesses**
  - **Only works with highly structured corpora**
  - **Busy result page may confuse some users**

search engine land

DeepDyve

- **Strengths**
  - **Really good semantic understanding – queries can be words, sentences, paragraphs or even whole pages**
  - **Excellent "more like this" refinement**
- **Weaknesses**
  - **Confusing UI; primarily a showcase for a paid information retrieval service**

search engine land

Others worth a look

- **Cognition**
- **Duck Duck Go**
- **SenseBot**

search engine land

How will Google, et al, use SS?

- First, semantic search will be folded in to existing technologies
- Key benefit: Providing superior results to "long tail" queries
- Better natural language capability
- Better results for news, real-time & other dynamic content

search engine land

Google's Semantic Search Efforts

- **Suggestions for related search queries**
- **Longer snippets for queries that are longer than three words**
- **Not really "true" semantic search – more a combination of brute force applied to a vast data set**

search engine land

A boon for advertisers

- **Perhaps the biggest benefit offered by semantic search will be for advertisers**
- **Combining semantic search techniques with behavioral, demographic or geographical targeting will offer unprecedented opportunities for pinpoint delivery of ads**

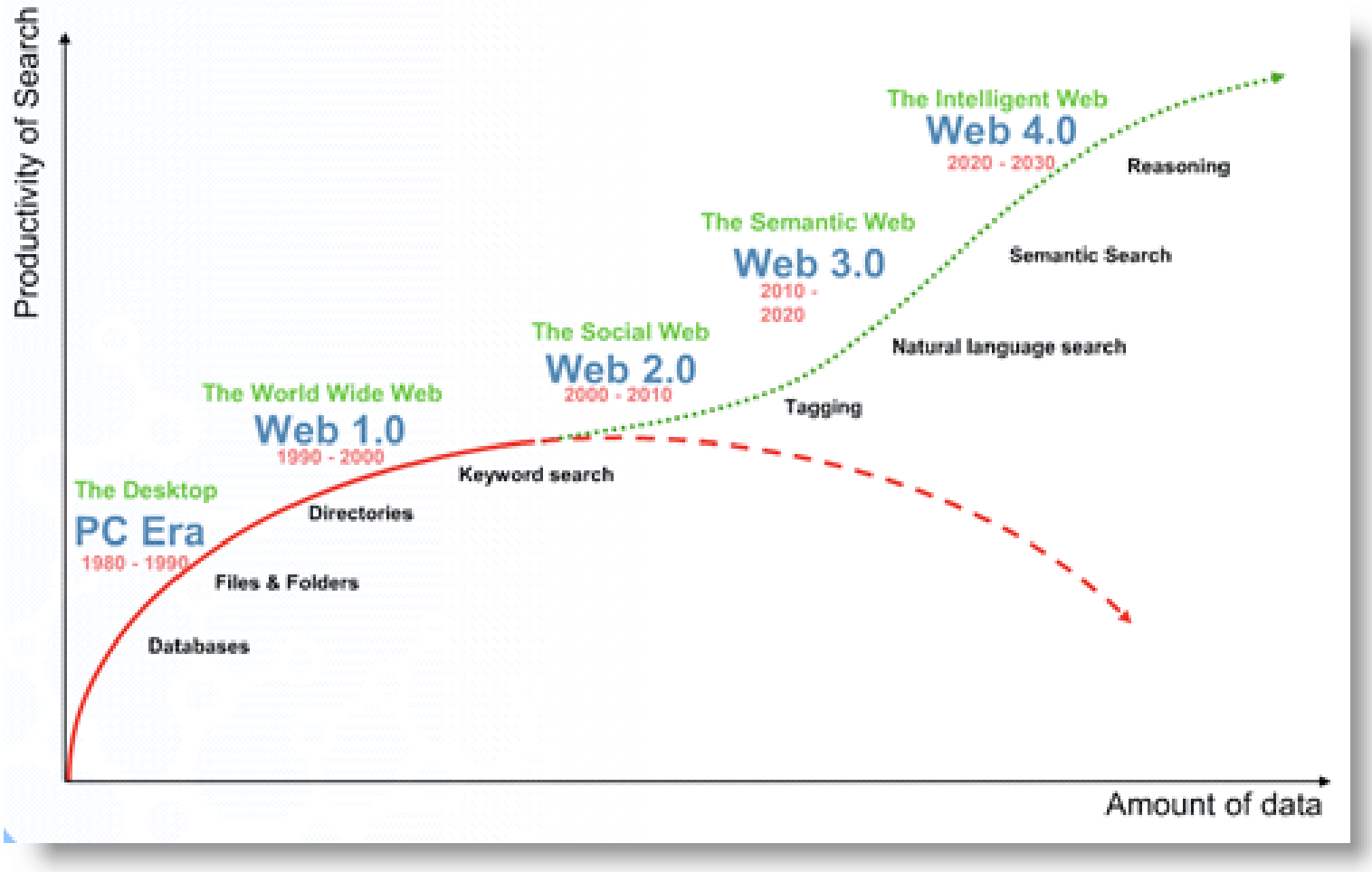
search engine land

Future trends

- **Is semantic search the key stepping stone to realizing the dream of the true semantic web?**
- **It's a good start**
- **With the huge data centers being built by the majors, we may get to the semantic web without all of the (currently required) metadata**

search engine land

The path forward



search engine land

Conclusion

- **Semantic search is an incremental improvement for research-based queries**
- **It will enhance, but not replace traditional web search**
- **It's also a step on the path to realizing the vision of the semantic web**

search engine land

Websites Shown

http://bit.ly/bdGtG5