

What is the Epistemology of Wayward Web Search?

Robin Hill

University of Wyoming, USA

What is the epistemology of web search, and how does it differ from the epistemology of other search techniques and other methods of knowledge acquisition? The seeker S types a phrase into the Google (or other) search bar in hopes of obtaining a fact (in propositional form) p . The search engine transforms the phrase into a symbol string which is sought in each of a vast pool of pages, returning links found. (If the search engine applies semantic processing, we don't know how, and ignore that possibility.) The search is successful when one or more of those pages contains the search string in some way that answers the question that S had in mind. But what happens when search goes bad?

Wayward results sometimes returned, here called "search blunders", are often glaring, even amusing, sometimes subtle. When S asks, in urgent need of the proper typesetting commands, for "LaTeX formula", she might get the chemical compounds of latex paint. When she searches for "population Park County", intending Colorado, the search might return the value 10,000 embedded prominently in some top page, which is for Park County, Wyoming. If she read as carelessly as she searched, she would formulate a false belief about the Colorado county's population, which is actually 16,000.

These are not falsehoods. The objective of the search engine, as coded by its engineers, is to satisfy the request. These results are veridical – they constitute information, but the wrong information. In a search for "landing on Mars", the pattern-matching might return,

quite correctly, a blog page claiming that men have landed on Mars; the blog does indeed state so, and it matches the search string. On this view, the web search is responsible for no independent proposition, but a fact more like that obtained from the sampling of a sensor.

In practice, however, we do treat web results propositionally, as natural testimony [2]. The seeker S extracts a claim, and risks a search blunder. This sort of blunder is rife in real-world testimony. "How old is your sister-in-law? "Forty-seven". But the questioner was asking about the other sister-in-law. The lack of context in the search string can lead to delivery of accidental truth. Suppose S 's population search returned the value for Park County, Montana, which is 15,500; S attains a reasonably correct belief about the population of Park County, Colorado. In the family case, the other sister-in-law may also be 47 years old. These search blunders look like Gettier problems, where an overlay of luck leads to justified true belief. A mistake of reference (whose?) marks both types, with no intrusive false proposition on which to hang the blame.

These scenarios differ in a gap that appears only in web search. The acquisition of p from a conversation, or a book, takes these steps:

Author or Speaker \rightarrow^{out} Publication or Utterance \rightarrow^{in} Seeker

The acquisition of p from the web takes these steps:

Author or Speaker \rightarrow^{out} Web Page $\rightarrow^{\text{match}}$ Search Engine \rightarrow^{in} Seeker

Blunder-free traversal of the path requires semantics, or grounding, all the way. But there is no such force to carry meaning across the *match* gap in the web case. Formally, we could extend epistemic logic [4] or justification logic [1] to investigate how knowledge holds across some testimonial transfers but not others. Human conversants (and authors) intend, and strive, to overcome the gap with context that is normally ready to hand, but the *match* gap in web search obliterates context.

Could we impose a metric on the amount of context in search that would account for search blunders? Search is an attempt to fill a lacuna, a gap in our knowledge circumscribed by surrounding knowledge. It carries more context than other types of knowledge acquisition, not personal but static body-of-knowledge context: A search for the title of a certain movie cannot exploit idiosyncratic context such as where and when it was seen by *S*, and with whom, but can exploit the names of the actors, the year, the setting or storyline or the name of the director, all associated with the body of knowledge on motion pictures.

The context manifests as these search criteria. A search question may be detailed – “What was the title of that movie about Scotland with Burt Lancaster?” A broader question – “What is Bollywood?” – imposes fewer criteria, exerting less circumscriptive force. Unsolicited (passive) knowledge acquisition, the state of a child or of a relaxed adult sitting in a park, imposes no criteria. Say that retrieval of a given *infor* requires a certain quantity of context, *X*, distributed between the question and the an-

swer. Unsolicited knowledge comes with all of it attached to the answer because there was no question. A detailed search question comes with rich context, so the answer need not supply much, if any. A pattern match comes with almost no context in the question, so all of *X* must be obtained elsewhere.

A web search result (if truthful, propositionally) is semantic information, under Luciano Floridi’s analysis – it consists of data that address a deficit, and it is well-formed, meaningful, and veridical. It does not depend on the source; even an automaton can generate information. The search blunder can be called aleatory, involving elements of chance. That issue is resolved by the Network Theory of Account, which requires the “right sort of information”, that is, relevant information. Our search blunders exhibit a failure of relevance that we can trace to the erotetic nature of information, modeled as questions *Q* and answers *A*, normalized so that the question holds all of the context: [3, p. 193]. Without the transfer of context to the question, we cannot deploy the model that yields correctness, and without that system-and-model verification, we cannot form knowledge.

In theory, human communication can load the context onto the question, which makes knowledge possible under Floridi’s analysis. Web search – ironically, because the web is the network par excellence – strips the context out of the question such that it cannot be rendered meaningful by the Network Theory of Account. The question is whether the knowledge goes with it.

References

- [1] S. Artemov, “The Logic of Justification”. *The Review of Symbolic Logic* 1(4), 2008, pp. 477-513.
- [2] C.A.J. Coady, *Testimony: A Philosophical Study*. Oxford University Press/Clarendon Press, 1992.
- [3] L. Floridi, *The Philosophy of Information*. Oxford University Press, 2011.
- [4] V. Hendricks and J. Symons, “Epistemic logic”. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition, 2015.