What is the Epistemology of Wayward Web Search?

Robin Hill <hill@uwyo.edu>

What is the Epistemology of Wayward Web Search?

History and Philosophy of Computing 3

Pisa, October 2015

Robin Hill, hill@uwyo.edu

An Epistemology Question

To broaden the subject matter of the philosophy of computer science, let's ask an obvious question regarding the two fields involved:

Can we apply epistemology to web search?

An Epistemology Question

To broaden the subject matter of the philosophy of computer science, let's ask an obvious question regarding the two fields involved:

~~Can we apply epistemology to web search?~~

How do we gain knowledge from successful web search?

An Epistemology Question

To broaden the subject matter of the philosophy of computer science, let's ask an obvious question regarding the two fields involved:

~~What about knowledge from web search?~~

~~How do we gain knowledge from successful web search?~~

What is the epistemology of wayward web search?

What is the Epistemology of Wayward Web Search?

- Web Search: A seeker *S* submits a *search term* to a search engine like Google.

  - "start date Crimean War"

  - "convert mysqldump access database"

  - "history anthropology women's rights feminism Apache Navajo Hopi Nation"

    If a page is returned that allows *S* to formulate a proposition answering the intended question correctly, the search is successful.

- Google processes 40,000 search requests per second.

- Search, in general, addresses a specific deficit in our knowledge.

- The web search mechanism is pattern-matching.

Testimony

We conflate sensor sampling and propositional claims for the current purposes. The results, to a search engine, are indistinguishable. They are veridical in the sense that they report the contents of the source accurately. In either case, we treat web search results as testimony (T. Simpson 2012).

Upon accepting the result, the seeker S holds a qualified commitment to p, easily defeasible, and equivalent in this regard to face-to-face testimony.

An open epistemological problem-- Why does testimony work?-- is not addressed here; we just remark that it does work. That is, testimony can be a source of knowledge, whether it serves as a justification for knowledge inherently or by virtue of some reduction.

Web Search versus Other Testimony

This means that:

- Our doxastic attitude toward web search is the same as toward other testimony.
- Truth in web search carries the same issues as truth in testimony.

Yet many of the quirks encountered seem unique to web search.

So: We seek problems with web search that cannot be attributed to testimony alone.

Note | We equivocate on the form of the results of web search, whether presentation of pages or something that looks like direct question-answering.

Wayward Web Search

To seek those problems: Let's see how web search can go wrong.

Examples of search terms and wayward results from pattern match:

1. *"LaTeX formula"*

   Meaning syntax of the typesetting system, but may return the chemical compound of paint.

2. *"population Park County"*

   Meaning Colorado (16,000), but may return Wyoming (10,000) instead.

3. *"financial status of FOSL"*

   Meaning the Friends of Sierra Leone, but may return info related to person or stock or other organization.

These are *search blunders*.

Mistake of Reference

Search Blunder: When the reasonable result (based on the pattern match) does not correspond to the reasonable expectation (based on the intention).
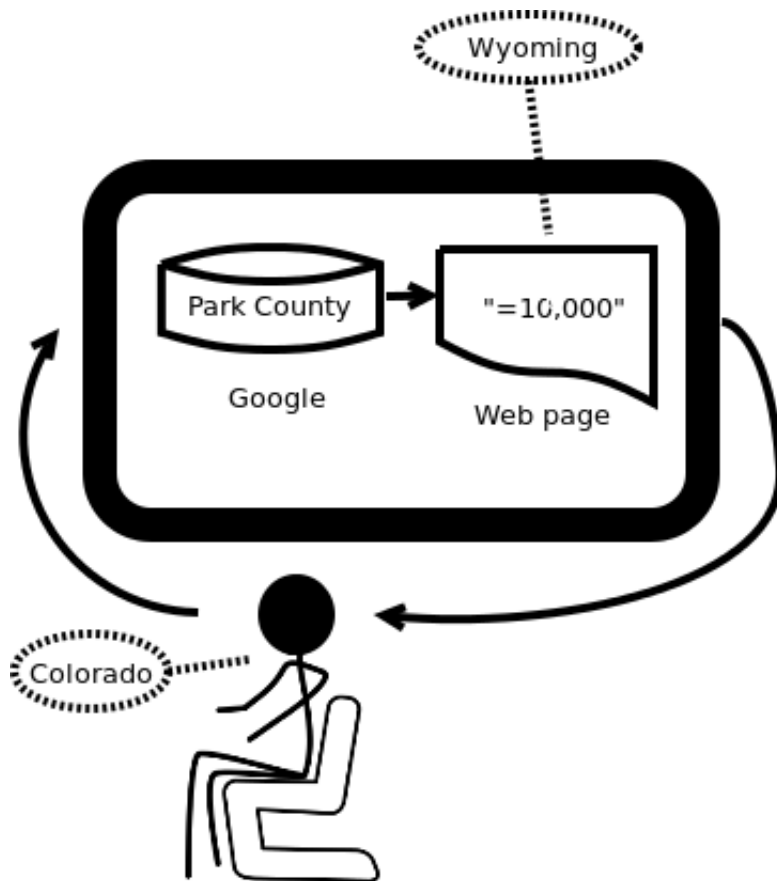
Whose fault?
No error or lie can be charged to either side.

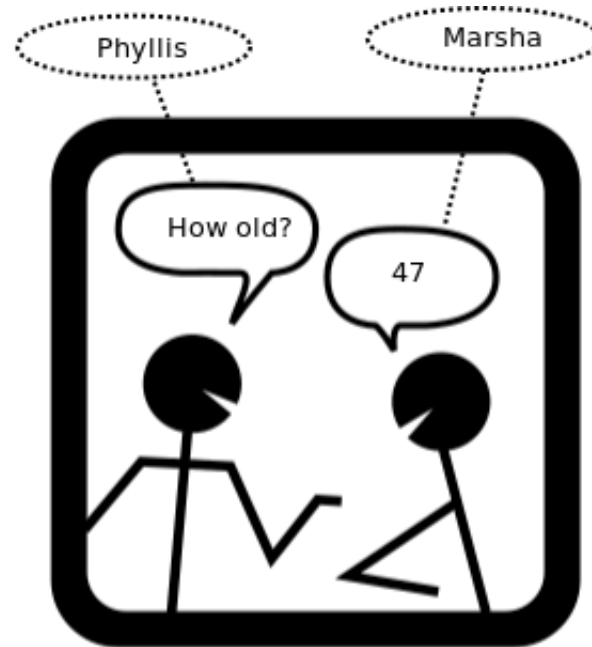Seated stick figure by kattekrab, from openclipart.org

Common in Face-to-face Discourse

But this happens in face-to-face conversation, as well.

"How old is your sister-in-law?"

"47"

But the questioner meant the other sister-in-law.

So the blunder in general is not the difference between web search and conversational inquiry.

Chat icon by Hermann Schwarting, from openclipart.org

In F2F, Context Provides Rectification

Here's the difference: Context provides easy resolution. Context, even if not immediately present in non-web interaction, is immediately accessible, and to the appropriate degree.

To wit: The Park County population mistake is unlikely to take place via a reference book. *S* will notice if the book refers to Wyoming rather than Colorado.

Web search is devoid of intentionality, under both the philosophical connotation of *intention* as content-ful, and vernacular meaning of *unintentional* as "not intended" or "accidental."

Gettier Resemblance

A search blunder, as interpreted by the seeker *S*, may provide a *true* proposition, resembling the good luck aspect of a Gettier case.

- Search string "population of Park County" may return 15,500, the population of Park County, Montana, giving a virtually true proposition as a result.

- The sister-in-law to whom the questioner refers may also be 47.

- The financial status of the Friends of Sierra Leone may be the same as the person or stock.

> This manifestation of the Gettier problem due to a mistaken reference is a side issue related to testimony, not to web search *per se*.

The Different Scenarios

The difference between inquiry in a face-to-face conversation (or library) and web search can be illustrated thus.

The acquisition of *p* from a conversation, or a book, takes these steps:

$$\text{Author or Speaker} \xrightarrow{out} \text{Publication or Utterance} \xrightarrow{in} \text{Seeker}$$

The acquisition of *p* from the web takes these steps:

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

The web search cannot detect a mistaken reference in a pattern match.

The Gap

The pattern match step, from the web page to the search engine, severs the flow of relevant information by stripping context. This is not deliberate, but inevitable. Web search does not allow for mistaken references; they are all correct.

But face-to-face disambiguation rules out possible worlds, worlds of interpretation.

Can we fix this by fiat? Could this be codified, or even rectified, by multi-agent versions of epistemic logic or justification logic? With suitable axioms:

1. We can declare that a unintentional agent can provide knowledge.

2. Or we can declare the opposite.

3. We can declare that adequate context be solicited along with the search term.
   Or can we?

Scope of Context

Can we develop a metric for adequate context? The "amount" of context differs widely.

Questions

1. What was the title of that movie about Scotland with Burt Lancaster?

2. What is Bollywood?

3. What's it all about?[...diffuse wondering...]

Context

1. Detailed context in question, little required for answer ("Local Hero").

2. Little context in question, much required in answer.

3. No context at all, open to anything.

Quantification of Context

To block search blunders, we would have to require some variable amount of context to be preserved across the pattern match for the purpose of rectifying search blunders. But how much?

Low Quantity

> The pattern "FOSL" found in a web page listing charitable organizations in Sierra Leone cannot be disambiguated from the pattern "FOSL" found on a stock market report.

High Quantity

> Extensive (pattern-matching) context will pick up both sisters-in-law, and will associate the Park Counties of Colorado and Montana, and may find common members of the two FOSL organizations.

Analysis under Philosophy of Information

Web search result is information:

- datum filling a deficit

- well-formed

- true (in the somewhat murky sense of testimony)

- meaningful

That is, the result conveys some fact, useful in some circumstances.

Philosophy of Information: Question Normalization

Floridi's information is erotetic, modeled by binary question-and-answer,

with the question normalized to hold all of the context:

$$Q_{0/1}^{CLP} + A_{0/1}$$

Where C is context, L is level of abstraction, and P is purpose. Q is a request to erase a data deficit through saturation, that is, communication to maximum capacity, which allows a unque Boolean answer.

We see that the scope of the context C is problematic, but it may be the purpose P or the level of abstraction L that is more pertinent. This locates the issue, and suggests that distinguishing these three parameters may be futile.

Philosophy of Information: Network Account

In Floridi's account (still very roughly):

Relevance of information is the probability that an agent would ask for it upon being informed that it might be available.

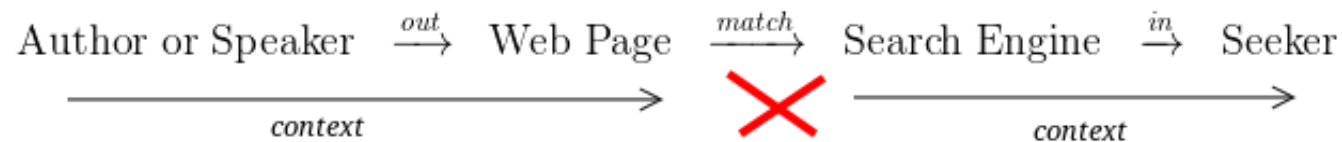The aleatory relevant information is resolved into knowledge by the Network Theory of Account.
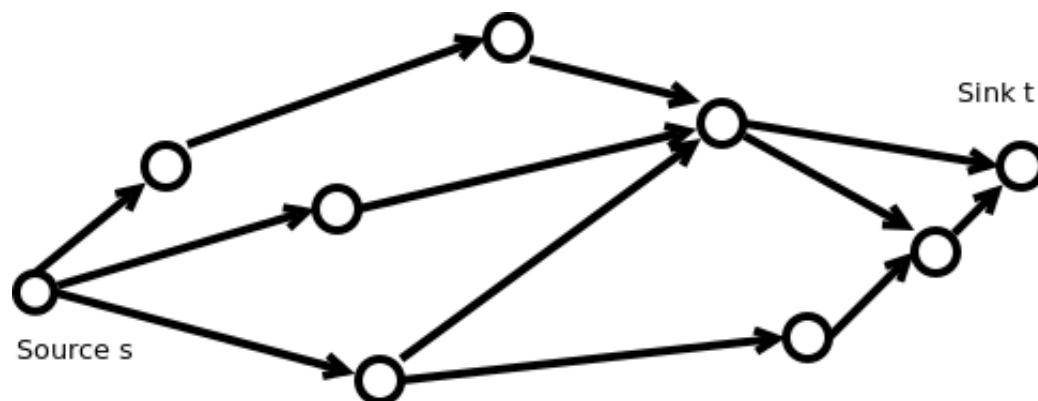
In this flow network, each arc *(u,v)* has a capacity *c(u,v)* that measure the answers available. From the source *s* to the sink *t*, each node is an information source and each arc a "how-come" question, giving Boolean answer.

The sink *t* poses the questions and the source *s* provides correct answers; and *t* is knowledge if there is a network providing such an account.

Web Search Blocks Question Normalization

Since we cannot transfer the appropriate context to the question in a way that carries across the pattern-matching gap, we cannot successfully normalize the question in the way required for relevance and hence, for aleatory resolution into knowledge.

Author or Speaker $\xrightarrow{out}$ Web Page $\xrightarrow{match}$ Search Engine $\xrightarrow{in}$ Seeker

$\xrightarrow{}$ ✗ $\xrightarrow{}$

context   *context*

The match strips out the context.

The question is: Does knowledge go with it?

Conclusion

We have exposed a difficulty in the acquisition of knowledge through web search-- the loss of appropriate context and meaningful reference as the info passes through the unintentional pattern match step.

Web search is akin to knowledge acquisition by testimony, except that when this difficulty occurs (routinely) in face-to-face scenarios, such issues are easily rectified. Pattern-matching does not recognize ambiguity, or even the notion of ambiguity.

We have expressed this quandary in terms of the Philosophy of Information, noting that the inability to normalize the inquiry embodied in the search term blocks the development of knowledge.

Our hope is that this brings us closer to understanding the foundational epistemological question, "When does S know that p?"

And some open questions...

Open Questions

1. What is the epistemology of search in general, and library or reference search in particular?

2. What is the epistemology of testimony from non-intentional (as opposed to intentional) agents?

3. Can some responsibility be placed somewhere?

4. Is the pattern-match gap related to the Gettier gap meant to be closed by some *+G* factor?

Discussion

Thank you for your attention. Questions?

Robin Hill

The author welcomes comments at any time.
hill@uwyo.edu