The future of texts is relationships

April 20, 2020

Daniel M. Russell, Google Research Scientist

The future of texts is also the past of text—all of the textual ancestors, the data accessors, the précis, the commentaries—as well as all of the other texts and interrelationships that go into forming your understanding of the text at hand. They lead from the past into the ongoing present, subtly influencing the interpretation of a text. The future of texts is relationships, both explicit and implicit.

Here's what I mean.

To a degree that none of us expected, global search changes everything about text because it has become possible to discover relationships and interpretations that were effectively invisible before easy and fast search. Search changes text, it changes reading, it changes understanding, and It changes what it means to be literate.

The things we write (or draw, animate, record, or film) don't exist independent of everything that's gone on before. After all, your personal experience re-colors how you interpret what you see/read/hear. In the same way, everything that connects to your present text is in many ways

connected with the texts that have gone on before. You can read Engelbart without Nelson, but knowing the connections between both enhances comprehension in the same way that reading Shakespeare is enhanced by knowing about Milton. Hyperlinking (like citations and references) make these connections explicit, but hyperlinks lets your mind move fast—one click, and you're at the destination of the link. Yes, in the old days you *could* have followed that citation to other work in that non-computational age, but people rarely did, and as we know, changing the cost-structure for analyzing complex concepts fundamentally changes the way we think. [Bush, 1945; Russell, et al, 1993]

Hermeneutics and sociology teach us that meaning of a text (or any media) is time- and place-dependent because the embedding of that text is within the universe of other texts. In an earlier, less civilized age, making connections was the ability of the privileged class that had the time and leisure to read many texts, taking many notes, finding correlations, connections, and intertwinglings. [Shneiderman, 2015] But times change, and now we all live in a world of large textual relationships.

For example, when I was studying for my doctorate in Computer Science in the 1980s, it was an age of disconnection. There were vast collections of texts in libraries, and there were links between documents in the form of citations and references. You could (and I did) create collections of all the relevant documents in your specialty area of interest. Hypertext had been invented, but it wasn't universally available. Following a reference usually meant a trip to the library and the local photocopier.

But then the internet happened and rapid access to links happened. Nearly free online publishing systems happened. Most importantly, nearly instantaneous global search happened—and that changed the way we think about text.   When someone cites Johnson as saying "*Knowledge is of two kinds, we know a subject ourselves, or we know where we can find information upon it.*" you can now do a quick search and find that this is actually the first part of a paragraph that goes on to say *"[and] When we enquire into any subject, the first thing we have to do is to know what books have treated of it. This leads us to look at catalogues, and at the backs of books in libraries*." These three sentences are not as pithy a quote, but the next two sentences change the common meaning. Johnson reads the backs of books so he'll know which volumes he can access in an age when access to texts was rare and difficult. Once he knows who has what books (because he read the book spines), he'll know "where we can find information upon it."   When you follow that cite, you'll also find that the quote is from Boswell's "The Life of Samuel Johnson," so it's really Boswell's recollection of the story. Search lets you not only find the original, but also understand the context and setting. [Boswell, 1873]

In essence, texts are simple to find and link with a straightforward search. Search engines let readers and writers make connections where there are no explicit linkages but simply allusions or echoes of previous work. As the saying goes, "everything is connected," even when the connections aren't made by the author, but can be discovered by the reader.

One of the implications of this is that we now live in an information triage culture. One really can't explicitly connect everything to everything because the links would obscure the meaning and return us to a lost-in-hypertext world view. Instead, you, as a reader, HAVE to figure out where to make the cut. In the world of many explicit links, it's barely possible to follow the links—but in a world where texts are implicitly linked together, the connections can be overwhelming. Readers can become like the man with perfect memory in Luria's "Mind of a Mnemonist" (1987) where the patient with a perfect, unforgetting memory becomes overwhelmed by the imagery that's evoked whenever he reads something. A chance phrase, "like tears in rain," could be linked to Blade Runner (the 1982 movie, from whence it comes), or Rutger Hauer (the actor), Roy Batty (the character), or the meaning of mortality to a replicant (the point of the scene). Do you, like the mnemonist, recall that "like tears in rain" was scored by Vangelis? Given so many options, there's a real skill in knowing which will be valuable to the reader. Skill is required not just on the part of the writer in anticipating the reader's needs, but also on the part of the reader to choose a valuable path among the options.

We often say about search that "everything is available," but we don't really mean it. Not only do readers not pursue every possible path from a well-turned phrase, but there are real limits to what's still active and working at read-time. 404s (missing page errors) are inescapable, even in a hypothetical information universe without walled gardens, where backups are made and robust archival mechanisms are in place. We recognize that ownership and copyright laws change over time, authors and governments change their minds about publication and retract documents or fragments thereof, some documents can be only accessed from particular geolocations, there's a certain amount of simple loss, and a certain amount of encoding attrition that happens with any encoding system.

But unlike images and interactive systems which will end up being fragile over the long course of time, text is easy to read far into the future. The encoding is simple and robust—it's straightforward to encode text in a format that will for the next 10 millennia. [The Long Now Project] The page-level formatting might be wonky a couple of thousand years out, but by contrast the ability to read images will become progressively tougher as time passes. Yes, one might *technically* be able to read and render an image file format from the 1970s, but it's going to get harder and harder (and more costly) to preserve the ability to recover an image or other media encoding. I'm trying to imagine rendering a Quicktime VR file from 1995 (the year of its introduction). This rendering problem won't happen with text.

Long after files of images, VR and AR file formats are illegible or unviewable, text will still be accessible and useful. We will still be able to analyze the texts and discern relationships between the texts and the fragments of knowledge encoded therein. That is, with time, we will be able to extract and encode even more value from the texts as we do deeper and richer analysis of the text. I hope this happens with other media forms, but I'm not optimistic. Even if we solve the perma-link problem, there will be a backwards set of dead links and uninterpretable formats that we still have to be able to find and navigate.

In addition to the constant growth and evolution of online content, texts will grow increasingly valuable through improved text-mining and concept extraction. In the future, look forward to being able to pose queries like "what did people think about Darwinian evolution in the 1990s?" Or "what prior art is relevant to a nanotechnology for creating self-organizing displays"? That is, while a text itself might be canonical and unchanging, its interpretation, relationships, and discoverability—especially with respect to other texts—will continue to change. The downside of all this accessibility is, as Weinberger writes, "We can all see the Sausage of Knowledge being made, one link at a time." [Weinberger, 2011] The ability of global search has changed the way researchers can discover, study, and refer to other texts. This exposes deep truths, as well as the false paths and mistakes made along the way to sausage-framed knowledge making.

From a reader's perspective, the future of text is being able to understand how to ask intelligent questions of our collected texts.   Critical reading skills are more than just about asking incisive questions of what we read and deliberating on them. Instead, as knowledge grows increasingly rich and connected, so too does the need for the reader to have strong search and research skills. [Russell, 2019]

Search skill is an important kind of expertise. We've all looked up information on a just-in-time basis, and had to work our way around the 404 errors and found lost content in online archives. Knowing how to frame a question (which has always been an important skill) know means knowing how to pose a question for search, then conduct the search, and not get lost in the plethora of results that you find. Text is more abundant, richer, and connected than ever. For the skilled reader, search skills are an important constituent of reading and understanding.

The future of text is in enriching the way we find, use, and understand relationships to other texts. Now more than ever, texts are part of a larger explied and implied web of knowledge. No text stands alone, especially now that its close relatives and ancestors can be easily and quickly discovered.

References:

Boswell, James. *The Life of Samuel Johnson*. William P. Nimmo, 1873.

Bush, Vannevar. "As we may think." The Atlantic Monthly 176.1 (1945): 101-108.

Luria, Aleksandr Romanovich, and Lynn Solotaroff (trans). *The mind of a mnemonist: A little book about a vast memory*. Harvard University Press, 1987.

Russell, Daniel M., et al. "The cost structure of sensemaking." Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems. 1993.

Russell, Daniel M. *The Joy of Search: A Google Insider's Guide to Going Beyond the Basics*. MIT Press, 2019.

Shneiderman, Ben. "Odes to Ted Nelson." Intertwingled. Springer, Cham, 2015. 7-11.

Weinberger, David. *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. Basic Books, 2011.