

Question answering systems in biology and medicine—the time is now

Jonathan D. Wren

Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104-5005, USA

Associate Editor: Alex Bateman

Contact: jdwren@gmail.com

Received on March 16, 2011; revised on May 25, 2011; accepted on May 26, 2011

On February 16th, 2011, it could be argued that the world quietly changed as a milestone in human history was reached. A question answering (QA) system, able to deconstruct a natural language question into information retrieval and analysis tasks, implemented across 2880 CPUs and embodied as an IBM-engineered system named Watson, handily defeated the top two human champions on a game show called Jeopardy. Similar to the achievement of IBM's Deep Blue in beating chess champion Gary Kasparov in 1997, it was not a surprise to technology enthusiasts that such a feat was possible, but it publicized the progress that has been made and the capabilities on information retrieval that are within reach. Thus far, the most salient examples of QA system implementations, as well as publications, are outside the biomedical domain, yet I would argue the sheer number of studied entities, heterogeneous nature of the data, exponential growth of information and emphasis on generation of new knowledge makes biomedicine the field in most need of good QA systems. Anecdotally, I was giving a talk to a couple hundred biomedical scientists and clinicians a week after the Watson challenge and only a couple dozen were aware of it when I asked. There seems to be a gap of awareness between the eventual beneficiaries of advanced QA systems and the developers in terms of recognizing its potential. This, unfortunately, might lead to marginalization of biomedical QA research in terms of publication and funding venues. We should take the Watson milestone as an opportunity to consider the ways biomedical research—including bioinformatics—can benefit from QA systems and some of the possible non-technical hindrances to progress.

On the surface, Watson's performance would seem more a victory for trivia retrieval than scientific research. The most pressing scientific questions are not those that are limited by factual retrieval. However, the scientific endeavor is predicated upon both observation and inquiry. As we gather observations, we are naturally curious if they are consistent with other observations, if what we observed has been reported before, and if what we believe may be the implications of our observations have been explored by others. This requires us to search the peer-reviewed literature which is not only large and growing exponentially in terms of publications (MEDLINE is currently growing ~5%/year), but the total searchable domain is

becoming rapidly larger with an increasing use of supplementary information and availability of full text. Thus, the quest for answers can not only be time consuming, but fraught with difficulties since the average research lexicon is filled with synonyms, acronyms and variations in naming conventions. A priority is placed on thoroughness (i.e. sensitivity/recall) because not being aware of relevant prior art can render a completed project moot. We are accustomed to trying to translate our questions into Boolean-based keyword searches that we hope will yield the best specificity to sensitivity trade-off in PubMed since it is not practical for us to sort through hundreds or even thousands of results when we know most of them are probably not directly relevant. The most difficult part, however, is not locating documents with single facts, but locating and connecting strings of facts that may be in different documents. The potential to engage in factual inference is probably the most powerful and appealing advantage of QA approaches over traditional information retrieval (IR) techniques. While keyword-based IR is a process researchers are accustomed to despite the inefficiencies, what the Jeopardy experiment showed us is that the technological know-how is here to automate, at least partially, this part of the process. And that, relative to human experts, it can perform well—it is specific enough to get answers correct, sensitive enough to find data hidden in mountains of text and smart enough to know when it does not know the answer.

Even though the QA field itself goes back, arguably, to the 1960s, development of QA systems for biomedical applications is a more recent phenomenon that has seen some but not much activity within PubMed (Cao *et al.*, 2011; Olvera-Lobo and Gutierrez-Artacho, 2010; Overby *et al.*, 2009; Zweigenbaum, 2003), even within bioinformatics journals. Yet, there are many vibrant research programs across the world working on the issue. Much of their progress is being reported in conferences such as TREC, which had a genomic QA track from 2006 to 2007 (Hersh and Voorhees, 2009) and BioCreative (Leitner *et al.*, 2010), which focuses on information retrieval, a foundational technology for QA. Yet, these are not often accessed by (and perhaps not even targeted to) biomedical researchers in general, the largest end user audience for the product. In defense of this apparent gap between producers and consumers, it could be argued that biomedical QA systems are not as ready for prime-time as Watson was. The issue is not one of whether or not outreach has taken place, but how effective it has been or will be. As a commercial endeavor, IBM cannot afford to spend effort advancing QA capability for purely academic reasons, and Jeopardy was their showcase for their product. Sometimes, it is argued the value of an advance is independent of its application, but I would argue

that the two are intertwined. Biomedical QA would benefit from its own Jeopardy equivalent, to showcase its potential and advertise to its ultimate end user audience. This visibility factor will likely play a role in the pace of development by garnering more interest from funding agencies and possibly more mainstream biomedical journals in different application areas (e.g. clinical, neurological, genomic, etc.).

Scientists, in general, are prolific question askers and where answers exist, they can eventually be found, given enough time and effort. However, I would argue scientific productivity, in general, will increase as good QA systems reduce the time it takes to search for answers. The time required to structure queries and sift through results could be spent instead on exploring options and possibilities for experimentation. Current QA systems focus mostly on searches for information within text, but the ultimate goal in scientific QA will be for them to use multiple data sources to answer questions (e.g. ‘how many uncharacterized proteins with GPI anchors are co-expressed with at least one coagulation factor?’), since our answers to questions are structured in many ways. That, however, is further away and blurs the line between QA systems and what might be considered intelligent, thinking systems, which is a worthy, yet far more distant goal (Wren, 2004). For the time being, we need to appreciate that QA systems have matured to the point they can best human experts and are essentially the next generation

of search engines that will have a direct impact on the speed of biomedical research, and that technical factors are not the only hindrances to advancing QA research in biomedicine—visibility to and appreciation by target user audiences may well determine just how soon the future arrives.

Conflict of Interest: none declared.

REFERENCES

- Cao,Y. *et al.* (2011) AskHERMES: an online question answering system for complex clinical questions. *J. Biomed. Inform.*, **44**, 277–288.
- Hersh,W. and Voorhees,E. (2009) TREC genomics special issue overview. *Informat. Retr.*, **12**, 1–15.
- Leitner,F. *et al.* (2010) An overview of BioCreative II.5. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **7**, 385–399.
- Olvera-Lobo,M.D. and Gutierrez-Artacho,J. (2010) Question-answering systems as efficient sources of terminological information: an evaluation. *Health Info. Lib. J.*, **27**, 268–276.
- Overby,C.L. *et al.* (2009) The potential for automated question answering in the context of genomic medicine: an assessment of existing resources and properties of answers. *BMC Bioinformatics*, **10** (Suppl. 9), S8.
- Wren,J.D. (2004) The emerging in-silico scientist: how text-based bioinformatics is bridging biology and artificial intelligence. *IEEE Eng. Biol. Med.*, **23**, 87–93.
- Zweigenbaum,P. (2003) Question answering in biomedicine. In *Proceedings of the EACL*. pp. 1–4.