POSITION PAPER

The Dilemma of the Direct Answer

Martin Potthast
Leipzig University

Matthias Hagen
Martin-Luther-Universität Halle-Wittenberg

Benno Stein
Bauhaus-Universität Weimar

Abstract

No Web technology has undergone such an impressive evolution as Web search engines did and still do. Starting with the promise of "Bringing order to the Web"[1] by compiling information sources matching a query, retrieval technology has been evolving to a kind of "oracle machinery", being able to recommend a single source, and even to provide direct answers extracted from that source. Notwithstanding the remarkable progress made and the apparent user preferences for direct answers, this paradigm shift comes at a price which is higher than one might expect at first sight, affecting both users and search engine developers in their own way. We call this tradeoff "the dilemma of the direct answer"; it deserves an analysis which has to go beyond system-oriented aspects but scrutinize the way our society deals with both their information needs and means to information access. The paper in hand contributes to this analysis by putting the evolution of retrieval technology and the expectations at it in the context of information retrieval history. Moreover, we discuss the trade offs in information behavior and information system design that users and developers may face in the future.

1 Introduction

The term "direct answer" is used to characterize situations where an information need, expressed as a keyword query or a natural language question, is answered without requiring any further interaction from the user. The (typically short) answer is displayed on the search results page or is spoken out loud, eliminating the need to browse the actual search results or to refine the query.

> The dilemma of the direct answer is a user's choice between convenience and diligence when using an information retrieval system.

Resulting from the manifold of constraints regarding time, intellectual capacity, economy, etc., users willingly accept direct answers, while, on the other hand, suffering potential harm if not having spent the effort for a thoughtful, deliberative research. It is a true dilemma since it cuts both ways: spending more time for research may incur significant inconvenience. The everyday tradeoff between the two extremes strongly depends on task and context: Consider a search for "CPR"[2] while a person nearby suffers a heart-attack, as opposed to while studying health care.

[1] Subtitle of the famous technical report introducing the PageRank algorithm of Page et al. [1999].
[2] CPR $\sim$ cardiopulmonary resuscitation

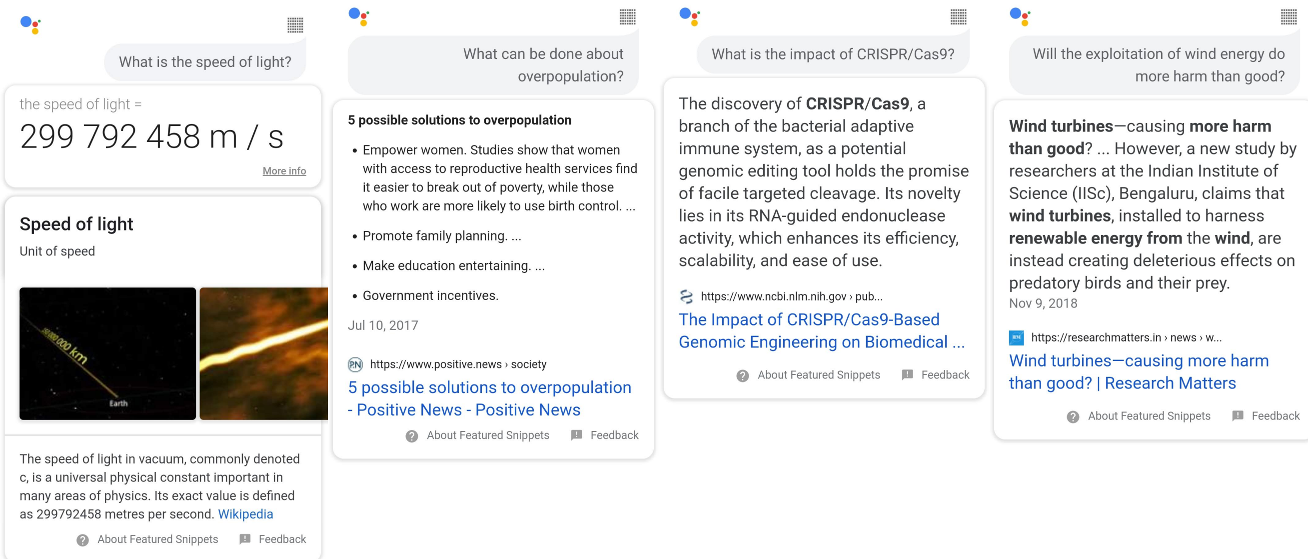


Figure 1: Examples of direct answers in the form of featured snippets on Google Assistant. Observe that only for the left-most—a factoid question—a list of results is returned, whereas, for the remainder, nothing more than the shown result forms the answer. An even shorter summary is spoken out loud. Can we accept these results as "responsible treatments" of the respective questions?

There are queries or questions, whose answers are straightforward, unanimously clear, and well-accepted, such as *"What is the speed of light?"*. These are qualified by nature to be answered directly; information needs of this kind are usually factoid. Crucial, and by far more challenging, are information needs on the other end of the complexity spectrum—questions whose answers are not straightforward, such as *"What can be done about overpopulation?"*, *"What is the impact of CRISPR/Cas9?"*, or *"Will the exploitation of wind energy do more harm than good?"*. These and similar questions are of significant importance for our society. The possible answers are controversially discussed, their backgrounds are analyzed, put in historical and cultural contexts, and relevant as well as irrelevant facts are mentioned in thousands or even millions of documents. Providing direct answers to them reduces the manifold of information items to just one (see Figure 1). Such a reduction must fall short. Complex information spaces require the careful balancing of facts and arguments in order to critically examine and assess the hypothesis space of potential solutions. Or conversely: Giving single direct answers is at odds with this requirement.

In this respect, recent large-scale deployments of technologies that give direct answers—despite the strong progress in these areas from the perspective of computer science—leave the impression of a system-oriented mindset.[3] It appears unlikely that it has been analyzed ahead of deployment what are the potential repercussions of giving direct answers to questions across the entire complexity spectrum on the information behavior of individual users, let alone on societal discourse at large. Unwitting users who receive a direct answer may take it a face value: The answer given, and not others, is the one to be taken seriously.

[3] The user- versus system-oriented discussion in information retrieval research [Voorhees, 2001] contrasts the technically feasible on the one hand and the wish to maximize user satisfaction on the other.

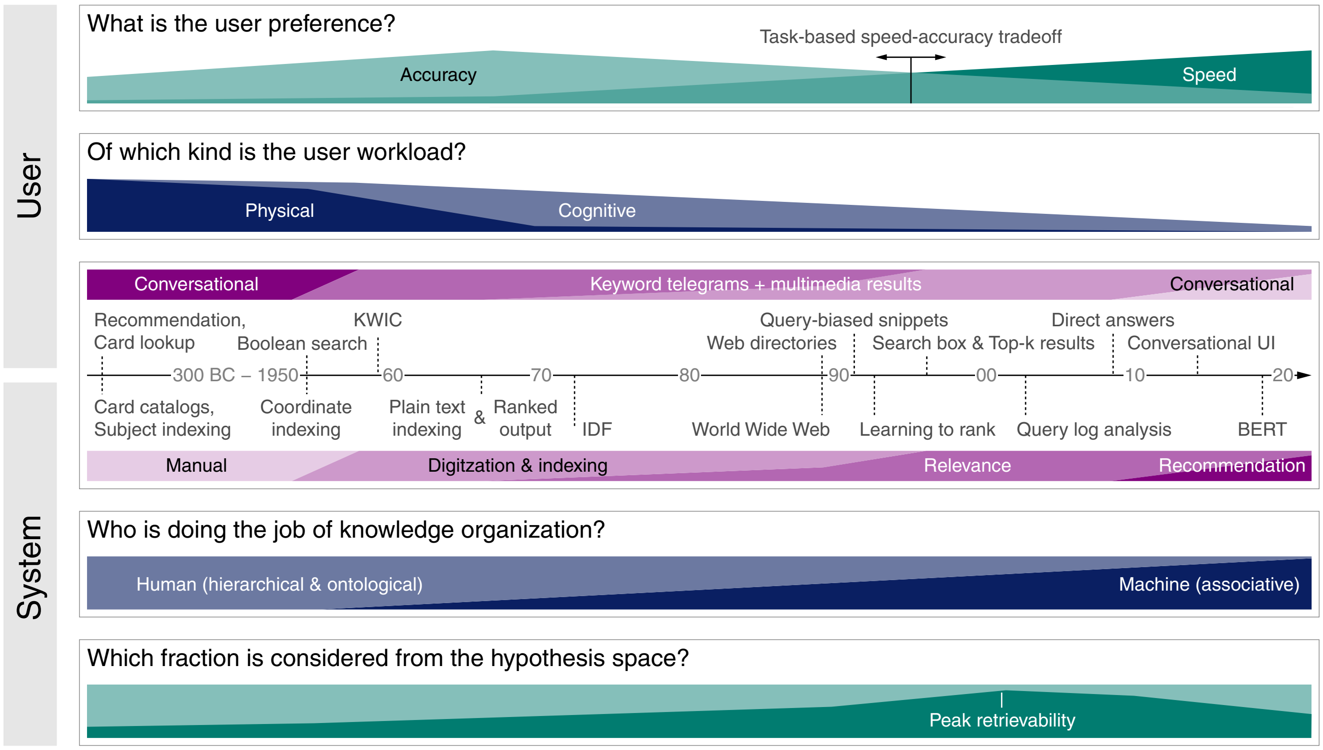


Figure 2: A brief history of retrieval technology (as timeline of innovations in the middle of the figure) along with the consequences of automation, and related to the direct answer dilemma. Directly above and below the timeline, the amount of human involvement in information storage and retrieval is appraised. Top and bottom, we estimate, qualitatively, the trade-off between speed and accuracy of solving a given task on the one hand, and the retrievability of information on the other: We hypothesize that the sweet spot of optimum information exploitation (see "Peak retrievability" in the bottom diagram) has already surpassed for a number of tasks.

In this paper we interpret these developments in context of the long history of information storage and retrieval. By outlining the trajectory of certain advances along with the driving forces behind them we shed light on the the direct answer dilemma, and its instantiations, effects, and consequences both at user and at system side.

2 From Conversational Search to Conversational Search

The development of techniques and technologies for information retrieval has always been driven by information overload, namely, the difficulty of getting a comprehensive overview of something in the face of an abundance of information sources that exceeds ones consumption capacity. In fact, the history of information retrieval [Sanderson and Croft, 2012, Harman, 2019], and particularly that of search engines, can be read as a continuous line of innovations to cope with this problem at an ever increasing scale (see Figure 2, center purple row).

For librarians, the question has always been how to organize their stocks of literature in ways that make information retrieval efficient as well as efficacious. From ancient times at the Library of

Alexandria until about the mid-twentieth century card catalogs (or their ancient equivalents) were the only means of searching for information in libraries, besides asking the librarians themselves. Until today, libraries organize their collections following standardized procedures of subject indexing. Card catalogs were, and their digital counterparts still are, maintained by trained librarians, who thereby gain an invaluable overview of a library's inventory. For centuries, librarians also acted as information intermediaries: Any non-trivial library search started with a conversation to a librarian, a conversational search, as it were.

When digital computing was introduced as a means to speed up the laborious indexing process as well as to enable more end users to search for themselves, it marked a qualitative departure from tradition: Coordinate indexing replaced complex subject descriptors with "uniterms" to be combined in a "query" with Boolean semantics [Taube et al., 1952], allowing for a significant reduction of controlled vocabularies. Misunderstanding of the concept caused some to derive terms from the to-be-indexed texts, instead of selecting them from a controlled vocabulary, which, surprisingly, worked well. Between 1962 and 1966 [Cleverdon et al., 1966, Cleverdon and Keen, 1966], in a series of experiments, it was shown that indexing all words of a text, and ranking documents by an estimate of their relevance to a query, provided for superior retrieval performance—a revolution in library science. Slowly, over the course of decades, as computing technology evolved and digitization efforts took hold, keyword search gained significant importance both within libraries and especially without, in organizations and companies that lacked the indexing traditions of libraries. Interestingly, on the system side, plaintext indexing was adopted into practice, whereas relevance-based ranking was not. For decades, Boolean search and ordering of search results according to meta data prevailed [Sanderson and Croft, 2012]. Similarly, on the user side, with ancestry dating back to the invention of keywords in context (KWIC) [Luhn, 1960], the question of whether or not query-biased snippets help users of retrieval systems had not been settled until about the turn of the millennium [Tombros and Sanderson, 1998].

Progress dramatically accelerated with the onset of the World Wide Web; its rapid growth caused information overload to rise to new heights, and to be felt more widespread throughout society. For a brief period, web directories, a mixture of yellow pages and non-standardized subject classification, dominated the search market, yet search engines were subject to fervent research. Google departed from the typical portal pages of incumbents like Yahoo, that combined directories with search, placing a single search box center stage, and finally ushering in all of the search technologies that had been known to work well for decades, plus many new ones. Although machine learning has been suggested early as a means to improve search result ranking [Fuhr, 1992], the lack of ground truth rendered any such attempts infeasible in practice, until, despite their noisy nature, query logs of search engines were found to provide for a strong relevance signal [Joachims, 2002]. Today, query logs are one of the most valuable assets of search engine companies, and Google has grown to virtually dominate the global search market.

Just like the rapid improvement of system-side retrieval technologies, so became the user interfaces of search engines subject to large-scale user studies, due to the commercial search engine's ever growing user bases, to understand and to improve user experience. A key performance indicator for a search engine is, for example, whether its users browse to a document retrieved for a query, which implicitly indicates that the document has been relevant (i.e., if the user dwells there for a sufficiently long time). Google's researchers noticed that a significant portion of users did not browse its retrieved documents at all, but abandoned their search after viewing the search results page. It turned out that in many such cases, the search engine did not fail to retrieve relevant results, but rather, the query-biased snippets already fulfilled the user's information need [Li et al., 2009]. Calling this observation "good abandonment," Google worked to further this way of satisfying information needs in an attempt to maximize user experience and satisfaction. In particular, Google introduced rich snippets[4] as well as direct answers[5] on its search results pages.

The mobile revolution, and more recently, the emergence of conversational agents, provide for a fertile ground for direct answers to thrive. Li et al. found good abandonment particularly often on mobile devices; mobile search was still a new paradigm at the time, only two years after the iPhone had been introduced. On small screens, and in situations where users have little time and patience to peruse search results, giving direct answers improves user experience. Today, mobile computing and thus mobile search makes up for the majority of the respective markets,[6] and not only the focus of web development has shifted to adopt a mobile first doctrine, but also that of search engine development.[7] An even "narrower," voice-only communication channel is offered by the recent conversational technologies. While research on conversational search is still in its infancy [Anand et al., 2020], Amazon, Google, Microsoft, and others already deployed corresponding technologies and compete on this emerging market worldwide. Due to the user interface constraints, direct answers are best-suited for conversational search. At the same time, language modeling and question answering technology based on deep neural transformer models has dramatically improved, allowing Google to deploy its latest BERT model to aid in answering question queries.[8]

Market estimations promise a strong growth of conversational interfaces and AI. Perhaps, in the future, conversational agents will serve as acting librarians of humanity's digital knowledge, and as capable information intermediaries, ultimately coming full circle on our long way from conversational search to conversational search. In any case, the newly developed technologies will, and already are propagated to the traditional search interfaces: Direct answers are given as a matter of course across all devices today.

3 The User's Tradeoff

Regarding the user's side, Figure 2 (top half) shows a qualitative assessment of the workload that "hypothetical users" have had throughout history to obtain information from the socio-technical information systems that were available at the time. The key takeaway is that the workload to obtain information has continuously and dramatically decreased. Apart from the fact that, until modern times, access to libraries was restricted to only a small fraction of society, at all times one had to physically visit the library, or the library had to physically have books delivered to lenders. Using a library meant planning a visit, talking to librarians, studying on site, and copying insights into one's own notes. With the emergence of public libraries and inter-library loan systems, as well as the cheap availability of books from publishers worldwide, the personal physical workload was reduced. However, the strongest impact on the physical workload necessary to obtain information

[4] https://webmasters.googleblog.com/2010/04/better-recipes-on-web-introducing.html
[5] https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html
[6] https://en.wikipedia.org/wiki/Usage_share_of_operating_systems
[7] https://developers.google.com/search/mobile-sites/mobile-first-indexing
[8] https://www.blog.google/products/search/search-language-understanding-bert/

was due to the digital revolution. When it comes to information retrieval, today, the vast majority of people hardly need to move away from their devices.

In terms of the cognitive workload (coming up with queries, analyzing search results), the trend generally points into the same direction. Though each technological revolution and the resulting changes in user interfaces required users to adjust their behaviors and to retrain, the cognitive overhead has also been continuously reduced: from talking to librarians over keyword queries in (mobile) web search to, currently, conversational search. Actually, computer-generated direct answers may be considered as pretty close to the minimum conceivable work overhead for users, reducing information retrieval to asking a question and reading or listening to a short answer.[9]

Considering as two ends of a spectrum of retrieval workload an "all-manual search, collection, and exploration of sources" on one end, and "direct answers" on the other, we raise the question:

> Given a retrieval task, what share of the retrieval workload should be carried out by the user to maximize the accuracy of the solution that the user selects or generates?

To enable the distribution of workload between user and system, the technological legacy of information retrieval could be harnessed, providing for a rich set of approaches of varying degrees of retrieval quality, each requiring more or less additional work from the user to complete a retrieval task. We believe that the extent to which more or less sophisticated information retrieval technology can and should be employed is highly task-dependent. And although more technological sophistication may optimize certain success metrics, like convenience and low cognitive workload, this is bought at the cost of other success metrics, such as solution accuracy. This is where the user has to make a tradeoff due to the dilemma: When given a choice between speed and accuracy, going into either direction may deteriorate the other, dependent on the task in question.

Maximizing one of the two is not economical, given budget constraints with respect to time, money, and least quality requirements. In practice, users are faced with the so-called speed-accuracy trade off, a concept frequently studied in cognitive psychology [Heitz, 2014], with applications to user interface design. Suppose, we are given a "degradable information retrieval system", the retrieval quality of which can be gradually adjusted between "random ranking" and "direct answering", then the aforementioned question can be recast as follows:

> Given a retrieval task, which tradeoff between retrieval quality and result accuracy achieves optimum costs for a user, or across the user population of a retrieval system?

While the final responsibility for the accuracy of one's solutions that result from information behavior rests with each individual user, it is clear that for certain retrieval tasks, other stakeholders have a vested interest: Consider a company whose business processes heavily rely on accurate research, or a society whose welfare depends on a well-informed population. In this regard, the affordances presented to users by direct answers, dependent on the task at hand, may set detrimental incentives. From the perspective of information foraging theory [Pirolli and Card, 1999],

[9]Maybe the only way to significantly speed things up any further is to envision a brain-computer interface that inserts information as needed into the brain, bypassing the senses as well as conscious thought. This way, retrieval would be indistinguishable from remembering. Arguably, the "smaller vision" of thinking a question to oneself that triggers an out-of-brain retrieval, and then "listening" to an answer "spoken" into one's consciousness is not much faster than speaking the question out loud and then listening to an answer read out loud.

direct answers may overwhelm the information scent of organic search results, which may particularly affect users with low self-efficacy [Kuo et al., 2004]. Under continued exposure to direct answers, users, who are satisfied with the answers given, can be tempted to skip significant parts of the actual reading process. Part of the user population will gradually simply get used to them. Giving accurate answers to simple and basic questions may further be the cause for an inductive extension of this trust to all answers the system gives. In light of this, an information society, which vitally depends on the accurate results from information behavior and technology, should more closely investigate the key performance indicators that are optimized by commercial search engines, and whether there are limits to their optimization, or heretofore disregarded indicators should be mandated.

4 The System-side Tradeoffs

Retrievability Regarding the system's side, Figure 2 (bottom half) shows a qualitative assessment of the shift of responsibility for knowledge organization from manual, human organization to a knowledge organization by machines. Due to the explosive growth of information, knowledge can no longer be organized manually by humans, but the organization is handed over to machines, which store the knowledge and provide interfaces to access it. While manual forms of organizing knowledge were and are often top-down hierarchies of categories, representing an agreed-upon ontology, knowledge organization by machines is mostly associative.

For all practical purposes, the way knowledge is automatically organized for retrieval is not meant to be human-readable, but to enable effective retrieval. The designated access path to documents is through hash keys that are derived and associated with documents during indexing. Nevertheless, most of the existing "traditional" indexing technologies can still be considered symbolic knowledge representations. With the onset of artificial intelligence and the embedding of language into large-scale language models, such as BERT, GPT, and others, knowledge representation may become more and more sub-symbolic. Presently, language models are often not (yet) queried directly for information, but serve to improve query understanding and matching of queries with sentences from documents that have been retrieved using a classical retrieval model. In recent work, Zamani et al. [2018] propose the use of neural networks for index construction.[10]

Our reliance on machine-organized knowledge affects our ability to access it. The interfaces available to users, and the more in-depth access paths available to engineers who have direct access to the indexing and organizing technology itself, define the possibilities and limits of which and how well retrieval tasks can be solved, in short: the retrieval system's affordances. In this regard, Azzopardi and Vinay [2008] introduce the concept of retrievability as a measure of the number of access paths by which an indexed document can be retrieved. We contextualize the concept of retrievability with that of a hypothesis space spanned by the documents and their contents within which retrieval takes place. Again, supposing we could pick and choose retrieval systems

[10]Conceivably, future retrieval technology may encode the contents of documents right within neural networks (e.g., a neural memory). The benefit of an end-to-end integration of indexing documents, language modeling, and question answering can be expected to be a severely improved "understanding". But the sub-symbolic nature of neural networks may prevent the enumeration and exact reproduction of indexed documents. Rather, information retrieval would be akin to the retrieval system "remembering" information by association in the form of queries or questions, and then reproducing faithfully (i.e., textually entailed) to the original.

with different retrieval characteristics from the stock of proposed systems, each configuration of a retrieval system would achieve a different "overall retrievability" within the hypothesis space.

Without a doubt, the large-scale infrastructures that support web searching have dramatically improved retrievability compared to previous search engines. But there is no reason to expect that every new innovation necessarily improves retrievability. We conjecture that "peak retrievability", a point in the history of information retrieval, where overall retrievability was at a peak, has already passed. For example, searching the web on a desktop or laptop computer has by far higher retrievability than searching the web with the same search engine on a mobile device. Using the latter, there is just not enough display space to interact with and compare search results; users are quickly fatigued. In the wake of the mobile revolution, a considerable fraction of users have altogether given up using desktops and laptops in favor of mobile devices, and thus, unwittingly, limited the space of hypotheses available to them in solving retrieval tasks.

But mobile devices are not an integral part of a retrieval system. Consider the trend toward giving direct answers. For the three examples on the right in Figure 1, only the answer shown was given.[11] On other occasions, Google Assistant does show a list of results, so that, by giving just a single answer, a strong message is sent to the user: "This is it. There's nothing else to be said." Obviously, much else needs to be said in answer to these particular questions, but the format of direct answers is entirely insufficient to do so. From this observation, we derive the following, important research question:

> Given a retrieval system, for which retrieval tasks should the user be warned to use a different system and/or device that better suits them to solve their tasks?

The technology to enable this distinction is not out of reach: After all, Google Assistant manages perfectly well to avoid giving direct answers to any question even remotely related to elections. Arguably, this avoidance of direct answers should be extended to all controversial issues, and perhaps to all topics whose hypothesis space is of sufficient complexity.

However, on mobile devices direct answers are, apparently, what provides for a better user experience; and on voice-only devices, direct answers are *the* enabling technology for conversational search. And, after all, if users prefer direct answers over the traditional ten blue links, then why not just give them what they want? This is where system developers must make a tradeoff due to the dilemma: When given a choice between overall user experience and overall retrievability, going into either direction may deteriorate the other. While in the past, commercial search engines have had strong incentives to provide for the best-possible service, the economic environment that reinforced these incentives has dramatically changed. From search engine perspective, if giving direct answers impacts the bottom line, this is what their shareholders will demand.

Relevance Feedback In this regard, there's a second tradeoff due to the dilemma: As we outlined above, one of the key innovations in web search was the use of query logs and interaction logs with search results as ground truth for learning to rank approaches. The user base of a commercial search engine, by interacting with search results to identify the documents that are most relevant to them, takes part in a continuous "crowdsourcing" of these logs. This form of implicit relevance feedback has been the most valuable relevance signal for search engines to date.

[11] These were the very first questions we tried.

If, however, more and more of the user population adopts direct answers, this short-circuits the relevance feedback loop. At best, the user reaction when receiving a direct answer may give weak hints at whether the answer was satisfactory or not (e.g., when a user reformulates the question to obtain a different result, or when affect measurements can be taken to gauge affective or cognitive arousal). Here, the dilemma is a choice between giving users what they want (direct answers), and undermining the very basis on which the retrieval model is built and improved. This, too, may in the long run affect shareholder value, unless there are other strong relevance feedback signals we are unaware of.

Responsibility The third tradeoff due to the dilemma pertains to responsibility. When a retrieval system gives a single direct answer, besides suggesting there is nothing else than that, the answer may be mistaken as the truth. Clearly, commercial search engines act only as intermediaries between their users and information sources, which does not change when, instead of ten blue links, they decide to show only one answer. There is no legal obligation to check whether an answer is true. However, the manner in which the answer is presented endows it with verisimilitude, a semblance of truth, thus affording a situation in which the user might stop double-checking ("good abandonment"). This is likely even the case for highlighted snippets on traditional search results pages, whereas, a pure ranked list of search results which all look the same will not let one result look more truthful than another (besides position bias). Although we cannot prove it, we believe, that the traditional view of "ten blue links" and some shown larger estimated number of results communicates the fact better that a search engine just ranks by relevance, not truth. Direct answers are just a semantically matching answer to a question, but by no means a pragmatically good answer, let alone a true one. Should direct answers come with an explicit disclaimer? For example:

> "This answer is not necessarily true. It just fits well to your question."

Referral Lastly, further afield, direct answers also contribute to developments that appear to undermine the otherwise well-balanced relationship between information intermediaries like commercial search engines, and their information sources. When users get answers to their questions directly on the search results page, there is no reason to visit the page on which the original information has been published. This way, the user and the search engine benefit from the work invested by an information source in creating a piece of information, whereas the information source is deprived of page visitors, who would otherwise be directly exposed to the source's business model. However, direct answers are but one aspect that plays into this phenomenon; also social networks are implied in this respect. Nevertheless, perhaps the most tangible backlash that commercial search engines currently face comes from the European Union, where an ancillary copyright law has been passed that exempts snippets extracted from news articles from fair use. Not only direct answers, but all snippets from news articles may thus be outlawed.

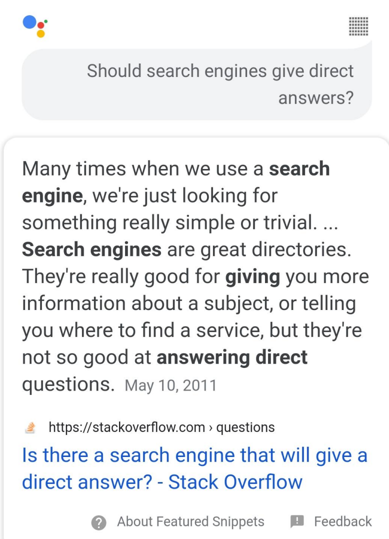
5 Conclusion

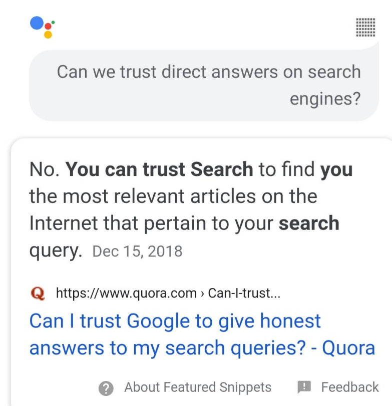
The provision of direct answers marks a paradigm shift and the onset of a new era of information storage and retrieval. In relevance-based information access, users are given a choice among many relevant pieces of information, having to decide for themselves what helps them best. With direct answers, that choice is reduced or even given up entirely, leaving the impression that the information system already knows what's best and what's true, and giving the user no further incentive to double-check. This development is driven to its extreme by the emergence of AI-based conversational agents and conversational search. As no actual conversations are currently supported by conversational search agents, every query is an ad hoc query that is met with one single answer. The technology to generate direct answers is mainly driven by the search industry who compete for their share of a prospectively large new market. The progress that has been made in these domains is impressive: For the first time, information retrieval technology paired with artificial intelligence, when given a question, is capable of identifying text that answers the question from within billions of documents. But for all the exciting new opportunities afforded by these technologies, their repercussions on society due to their large-scale deployment are not well-understood.

In this paper, we outline a dilemma which arises as a result of the paradigm shift: the direct answer dilemma. It takes different forms both at the user side as well as the system side, requiring a trade off between convenience and user experience on the one hand, and accuracy and retrievability on the other. While giving direct answers is a necessity in many user and system contexts, for just as many retrieval tasks this is not warranted, or even detrimental to upholding a healthy competition of ideas within societal discourses on controversial issues. Moving forward, we recommend a careful assessment of the prototypes, their capabilities, and especially their limitations. For many types of questions and retrieval tasks no direct answers should be given due the complexity of their underlying hypothesis space.

More generally, dependent on the retrieval task at hand, the level of technological sophistication that leads to an optimal tradeoff between retrieval speed and accuracy may differ dramatically. For example, during learning, less sophisticated retrieval technology might benefit learning success by requiring users to organize and structure the material themselves instead taking shortcuts, whereas, when checking a fact (e.g., during a discussion), direct answers are better-suited.

It is a good sign, that the retrieval systems themselves are at least aware of their limitations.





References

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford InfoLab, 1999.

Ellen M. Voorhees. The Philosophy of Information Retrieval Evaluation. In Carol Peters, Martin Braschler, Julio Gonzalo, and Michael Kluck, editors, *Evaluation of Cross-Language Information Retrieval Systems, Second Workshop of the Cross-Language Evaluation Forum, CLEF*, volume 2406 of *Lecture Notes in Computer Science*, pages 355–370. Springer, September 2001. doi: 10.1007/3-540-45691-0_34. URL https://doi.org/10.1007/3-540-45691-0_34.

M. Sanderson and W. B. Croft. The history of information retrieval research. *Proceedings of the IEEE*, 100(Special Centennial Issue):1444–1451, May 2012. doi: 10.1109/JPROC.2012.2189916.

Donna Harman. Information retrieval: The early years. *Foundations and Trends in Information Retrieval*, 13(5):425–577, 2019. ISSN 1554-0669. doi: 10.1561/1500000065. URL http://dx.doi.org/10.1561/1500000065.

Mortimer Taube, C. D. Gull, and Irma S. Wachtel. Unit terms in coordinate indexing. *American Documentation*, 3(4):213–218, 1952. doi: 10.1002/asi.5090030404. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090030404.

Cyril W Cleverdon, Jack Mills, and E Michael Keen. Factors determining the performance of indexing systems, volume 1: Design. Technical report, Cranfield: College of Aeronautics, 1966.

Cyril W Cleverdon and Michael Keen. Factors determining the performance of indexing systems, volume 2: Test results. Technical report, Cranfield: College of Aeronautics, 1966.

Hans-Peter Luhn. Key word-in-context index for technical literature (kwic index). *American Documentation*, 11(4):288–295, 1960. doi: 10.1002/asi.5090110403.

Anastasios Tombros and Mark Sanderson. Advantages of query biased summaries in information retrieval. In W. Bruce Croft, Alistair Moffat, C. J. van Rijsbergen, Ross Wilkinson, and Justin Zobel, editors, *SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28 1998, Melbourne, Australia*, pages 2–10. ACM, 1998. doi: 10.1145/290941.290947. URL https://doi.org/10.1145/290941.290947.

Norbert Fuhr. Probabilistic models in information retrieval. *The Computer Journal*, 35(3): 243–255, 06 1992. ISSN 0010-4620. doi: 10.1093/comjnl/35.3.243. URL https://doi.org/10.1093/comjnl/35.3.243.

Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 133–142, New York, NY, USA, 2002. ACM. ISBN 1-58113-567-X. doi: 10.1145/775047.775067. URL http://doi.acm.org/10.1145/775047.775067.

Jane Li, Scott B. Huffman, and Akihito Tokuda. Good Abandonment in Mobile and PC Internet Search. In James Allan, Javed A. Aslam, Mark Sanderson, ChengXiang Zhai, and Justin Zobel, editors, *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, Boston, MA, USA, July 19-23, 2009*, pages 43–50. ACM, 2009. doi: 10.1145/1571941.1571951. URL https://doi.org/10.1145/1571941.1571951.

Avishek Anand, Lawrence Cavedon, Hideo Joho, Mark Sanderson, and Benno Stein. Conversational Search (Dagstuhl Seminar 19461). *Dagstuhl Reports*, 9(12):18–46, April 2020. ISSN 34-83. URL https://drops.dagstuhl.de/opus/volltexte/2020/11983/.

Richard P. Heitz. The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior. *Frontiers in Neuroscience*, 8:150, 2014. ISSN 1662-453X. doi: 10.3389/fnins.2014.00150. URL https://www.frontiersin.org/article/10.3389/fnins.2014.00150.

Peter Pirolli and Stuart Card. Information foraging. *Psychological review*, 106(4):643, 1999.

Feng-Yang Kuo, Tsai-Hsin Chu, Meng-Hsiang Hsu, and Hong-Ssu Hsieh. An investigation of effort-accuracy trade-off and the impact of self-efficacy on web searching behaviors. *Decision Support Systems*, 37(3):331–342, 2004. ISSN 0167-9236. doi: https://doi.org/10.1016/S0167-9236(03)00032-0. URL http://www.sciencedirect.com/science/article/pii/S0167923603000320.

Hamed Zamani, Mostafa Dehghani, W. Bruce Croft, Erik G. Learned-Miller, and Jaap Kamps. From Neural Re-Ranking to Neural Ranking: Learning a Sparse Representation for Inverted Indexing. In Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang, editors, *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pages 497–506. ACM, 2018. doi: 10.1145/3269206.3271800. URL https://doi.org/10.1145/3269206.3271800.

Leif Azzopardi and Vishwa Vinay. Retrievability: An evaluation measure for higher order information access tasks. In James G. Shanahan, Sihem Amer-Yahia, Ioana Manolescu, Yi Zhang, David A. Evans, Aleksander Kolcz, Key-Sun Choi, and Abdur Chowdhury, editors, *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, pages 561–570. ACM, 2008. doi: 10.1145/1458082.1458157. URL https://doi.org/10.1145/1458082.1458157.