AgAsk: A Conversational Search Agent for Answering Agricultural Questions

Hang Li
The University of Queensland
Brisbane, Australia
hang.li@uq.edu.au

Bevan Koopman
CSIRO
Brisbane, Australia
bevan.koopman@csiro.au

Ahmed Mourad
The University of Queensland
Brisbane, Australia
a.mourad@uq.edu.au

Guido Zuccon
The University of Queensland
Brisbane, Australia
g.zuccon@uq.edu.au

ABSTRACT

While large amounts of potentially useful agricultural resources (journal articles, manuals, reports) are available, their value cannot be realised if they cannot be easily searched and presented to the agriculture users in a digestible form. AgAsk is a conversational search system for the agricultural domain, providing tailored answers to growers questions. AgAsk is underpinned by an efficient and effective neural passage ranking model fine-tuned on real world growers' questions. An adaptable, messaging-style user interface is deployed via the Telegram messaging platform, allowing users to ask natural language questions via text or voice, and receive short natural language answers as replies.

AgAsk is empirically evaluated on an agricultural passage retrieval test collection. The system provides a single entry point to access the information needed for better growing decisions. Much of the system is domain agnostic and would benefit other domains. AgAsk can be interacted via Telegram; further information about AgAsk, including codebases, instructions and demonstration videos can be accessed at https://ielab.io/publications/agask-agent.

CCS CONCEPTS

• **Information systems** → **Information retrieval**.

KEYWORDS

Conversational search, Neural Information Retrieval, Agriculture

ACM Reference Format:
Hang Li, Bevan Koopman, Ahmed Mourad, and Guido Zuccon. 2023. AgAsk: A Conversational Search Agent for Answering Agricultural Questions. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining (WSDM '23), February 27-March 3, 2023, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3539597.3573034

WSDM '23, February 27-March 3, 2023, Singapore, Singapore

ACM ISBN 978-1-4503-9407-9/23/02…$15.00
https://doi.org/10.1145/3539597.3573034

1 WHAT PROBLEM DOES AGASK ADDRESS?

Twenty first century agriculture is increasingly mechanised, data-driven and scientific-evidence based [1, 14, 15]. This had generated a wealth of valuable resources and data that could be used by agricultural users. However, much of these resources are currently locked away in natural language documents: research project reports, communications and scientific publications [9]. These documents are not easily discoverable, interpretable, and synthesised for agricultural users. No easy-to-use service is in place that provides a single entry-point to search these resources. Without such a service, agricultural users are not able to get evidence-based answers to their questions. AgAsk aims to provide this service.

2 GROWERS: THE TARGET USERS OF AGASK

For the purpose of this paper we refer to all our agricultural users as 'growers'. Growers encompass both farmers working the land, agronomists (agricultural scientists) and agriculture consultants who work directly with farmers. The common thread is that all these people have agricultural questions for which they want evidence-based answers, derived from the body of agricultural knowledge out there.

These growers questions are complex and multi-faceted; thus effective systems are required to serve the appropriate information. Scientific-like questions such as "What varieties of bread wheat are most resistant to crown rot?" are hard to answer automatically. Two problems make these questions hard to answer:

- **Complex answer matching**: Growers may express their queries in ways that do not directly match relevant information. The complex information need also comes with many query variations that an automated system must be able to handle.
- **Answer generation**: Growers need easily digestible answers to their questions; presenting a 25-page scientific document will not provide much help, both from a workload perspective and for growers to recognised how it might relate to their need.

3 HOW DOES AGASK HELP?

In this demonstration we present AgAsk: an end-to-end system that offers growers a single entry-point to search agricultural advice. AgAsk is a conversational search agent that integrates the latest developments in neural information retrieval for the effective identification of agricultural advice and the generation of answers.

AgAsk is currently deployed as a Telegram bot; Telegram is a popular messaging platform available for any device. In this paper, we describe the technical architecture of AgAsk, as well detailing and empirically evaluating the neural retrieval model underpinning the retrieval component of AgAsk. While AgAsk is set within the agriculture domain, the methods themselves are generally applicable to any domain where tailored, evidence-based answers must be derived from large collections of scientific documents.

4 SYSTEM OVERVIEW OF AGASK

Figure 1 provides the overall architecture of AgAsk. A grower uses Telegram to ask his question to the 'AgAsk' bot. Overall conversation management is handled by Macaw [16], an open-source framework for building conversational search systems. Macaw passes the query to our custom retrieval pipeline, comprising of a first stage BM25 retriever and the neural TILDEv2 re-ranker [17]. Retrieved passages are then sent to the BART answer generation model which converts the passages into a single coherent answer. The answer is then fed back to Macaw, which is responsible for serving it back to the grower via Telegram.

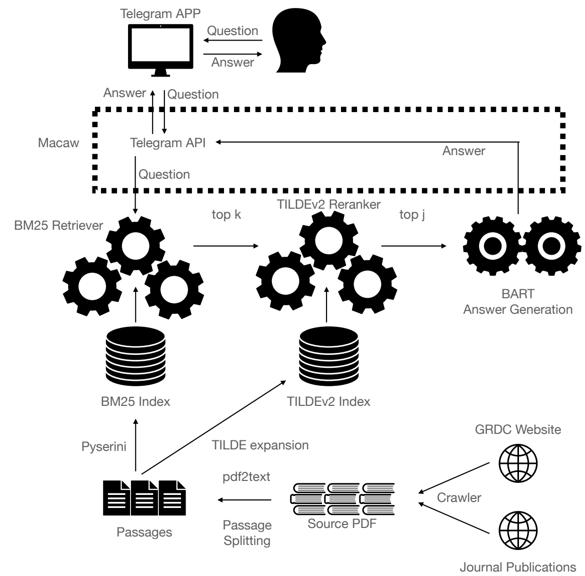

Figure 1: Overall architecture of AgAsk.

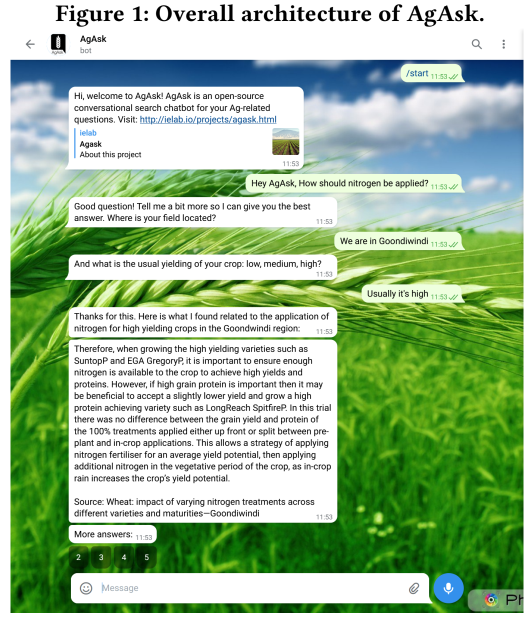

Figure 2: A screenshot of the AgAsk Telegram bot.

4.1 Client and User Interface

AgAsk is accessible to agricultural users via Telegram; an example screenshot is shown in Figure 2. Telegram was chosen because it provides a simple API and Telegram clients are available for every major platform and device. The grower can pose a natural language question and AgAsk will respond with a generated answer.

A demonstration video of AgAsk is available at https://ielab.io/publications/agask-agent. The retrieval of passages is done by the AgAsk bot. The clarifying questions are currently manually inserted to demonstrate what a fully interactive and response system might look like. We are in the early stages of deploying in production such a mixed-initiative conversational system.

We also log all user interactions including clicks, likes and emojis; this provides a source of relevance feedback information that may be used in future feedback mechanisms or online learning to rank.

4.2 Document Collection & Processing

Currently AgAsk uses two source of information: 82,843 scientific articles taken from 33 agricultural journals[1]; and 4,003 agricultural reports taken from an industry based agricultural corporation[2]. Combined, these provide a wide variety of topics related to all aspects of the crop and soil management.

Source documents were all in PDF format. They were converted from PDF to JSON using Apache Tika. From here, the text documents were further split into passages of three sentences (the Spacy sentencizer was used to derive sentence boundaries). From the 86,846 documents, $\approx$ 9.5M passages were produced.

4.3 Indexing

The extracted passages were indexed in two separate indices: one for the BM25 retriever; another for the TILDEv2 re-ranker. For the BM25 index, we used Pyserini [8] to index one passage at a time in a standard inverted file index. For TILDEv2, we first expanded the entire passage collection with TILDE [18] using 200 expansion terms. The expanded collection was then indexed for use with the fine-tuned TILDEv2 model.

4.4 Retrieval Method: BM25+TILDEv2

AgAsk uses BM25 for first stage retrieval and TILDEv2 for re-ranking. TILDEv2 is a sparse neural re-ranker that relies on document expansion at indexing time to avoid the need for neural inference at query time [17]. The expansion of the entire collection is accomplished by TILDE [18], which is cheaper in both time and costs compared to alternative expansion methods used by neural sparse models, like doc2query [11]. Retrieval is achieved by first ranking passages using BM25 and then re-ranking the top $k = 1,000$ passages with TILDEv2. The use of TILDEv2 in place of alternative neural rankers was to achieve an effective and importantly computationally efficient pipeline, as TILDEv2 does not require GPU

[1]For which relevant authorisation from publishers were obtained.
[2]The Grain Development Research Corporation of Australia (https://grdc.com.au) which was one of the partners in the development of AgAsk.

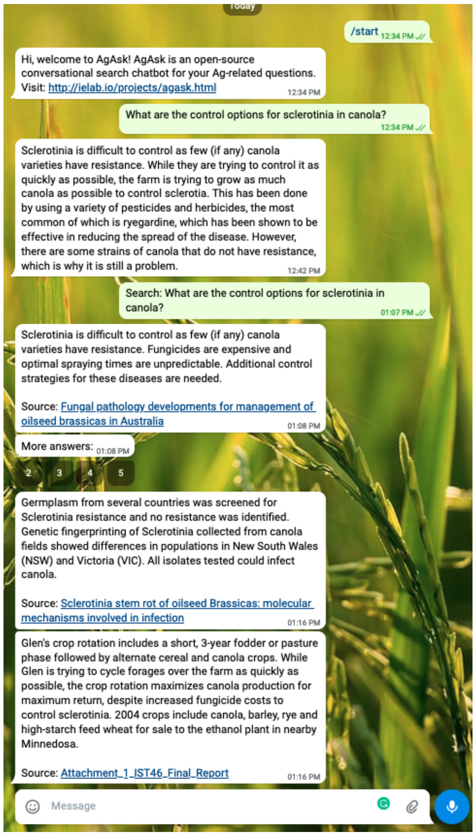

Figure 3: Example of an answer from the current answer generation method in production that has not been fine-tuned on in-domain data (top), along with the top-3 passages identified by the retrieval pipeline (bottom).

at inference time (nor for the passage expansion step performed by TILDE) . The resulting retrieval pipeline has a query latency that is compatible with those expected by users of a live system, and has a low production infrastructure cost as no expensive GPU infrastructure is required to support the retrieval component.

The TILDEv2 checkpoint deployed in production was fine-tuned on MSMARCO and then further fine-tuned on the training set of our agricultural collection with a small set of topics (further details in Section 5), which was out of domain on MSMARCO.

It is possible to use the output of the retrieval pipeline to directly display results to users. Alternatively it can be used as input to the answer generation module.

4.5 Answer Generation

The answer generation component takes the list of top j relevant passages provided by the retrieval pipeline (BM25+TILDEv2) and uses these to generation a single coherent answer to the question. In the module, we employed the BART [6] answer generation model. We are in the process of fine-tuning this model to our collection; for this we are acquiring gold answers to the evaluation topics from agricultural experts. An example answer generated by the BART model not fine-tuned (zero-shot) on our collection is provided in Figure 3: this shows that some of the answers are still of acceptable quality, even though in-domain fine-tuning has not occurred yet.

4.6 Conversation Management with Macaw

AgAsk employs the Macaw conversational information seeking framework [16], as it provides a convenient way of building an entire pipeline from scratch. The Macaw framework consists of several modules, including intent identification, co-reference resolution, query generation, retrieval model, and result generation. Currently, we have disabled the intent identification, co-reference resolution, query generation, file IO, and standard command line IO modules. We have instead instantiated our own retrieval and result generation modules, as detailed above, while we are in the process of deploying in production relevant modules for intent identification, relevance feedback, and question clarification.

5 RETRIEVAL MODEL EVALUATION

We perform an empirical evaluation to understand the effectiveness of the retrieval model used by AgAsk. We use the previously detailed agricultural passage collection. As for topics we form a set of 210 natural language questions provided by two agricultural experts and representing a range of grower related questions. The 210 topics were divided into 160 train and 50 test. The same experts performed a relevance assessment of these topics and passages from the AgAsk collection. The resulting test collection is the subject of a forthcoming publication currently in preparation.

The following retrieval models were evaluated (all these also contributed to the pool for relevance assessment):

- **BM25+TILDEv2**, the underlying model of AgAsk (Section 4.4), fine-tuned on the 160 training topics, with BM25 stage 1 implemented using the Pyserini toolkit [8].
- **BM25+monoBERT**, a comparative baseline using the same BM25 stage 1, that uses a BERT-based cross-encoder pre-trained on MS-MARCO and fine-tuned on the 160 training topics.
- **BM25**, a basic baseline and the same stage 1 retriever for the above two models.

Evaluation results on the 50 test topics are reported in Figure 4. There was a large difference in effectiveness between the first-stage term-based BM25 model and the neural re-rankers: monoBERT and TILDEv2 provided a far more effective top-5 re-ranking than the initial BM25 ranking (differences are statistically significant, two-tailed paired t-test, $p < 0.01$). However, it's worth noting that for measures like success@100, BM25 was highly effective. This meant that BM25 *retrieved* the relevant passages, but was not effective at *ranking* them (low effectiveness for measures that consider top ranked results; e.g., nDCG@5). This tells us that using BM25 for initial retrieval was useful, if followed by a high-precision re-ranker. We recall that the answer generation component, powered by BART, only considers a handful of top-ranked passages (for efficiency reasons), and thus a high effectiveness in early rank measures will likely result in effective answer generation. Furthermore, if the answer generation component was disabled, simply returning passages to the user, then in a conversational setting these would likely only be the first few retrieved ones, e.g. three or five, rather than a long list.

MonoBERT was the most effective model. If you consider a live question-answering system that might provide three possible answers to a user's question (e.g., in a conversational or mobile setting) then success@3 would be the measure to consider. In this setting monoBERT provided a success@3 of 0.96: 48/50 topics had a relevant passage in the top 3 results. We posit this would make for a highly effective production system if the results generalise.

While monoBERT was highly effective, it was computationally expensive. Query latency would make it prohibitive to real users in an online passage retrieval setting; for example, re-ranking the top 1,000 passages from BM25 took on average 23.1 seconds in our experiments, based on a high performance cluster based on Nvidia Tesla V100 GPUs. This time would be further compounded with the time required for answer generation (also requiring GPU for inference) and by the additional modules in AgAsk, e.g., intent identification. TILDEv2, while less effective, was far more efficient, and could be deployed on commodity hardware. These trade-offs between monoBERT and TILDEv2 were shown in previous work too, which have also considered energy (and thus running) costs required by the methods [13, 17].

In this evaluation we have focused on only the passage retrieval and not the answer generation. Recall that one setting of AgAsk was to just display passages; the other is to feed these passages to the BART answer generation model. To evaluate answer generation we need to form a new test collection that contains gold standard answers to each topic (rather than relevant passages). Forming this test collection and the evaluation of BART is left to future work.

6 HOW DOES AGASK COMPARE TO EXISTING SYSTEMS?

While several prior works have targeted open-domain conversational question answering [3, 5, 7, 12], domain-specific question answering in agriculture have attracted less attention.

Mostaco et al. [10] proposed AgronomoBot, which is a chatbot developed for the search and display of data acquired from a Wireless Sensor Network deployed on a vineyard. The bot is also based on the Telegram API and is able to access information collected by field sensors, providing it back to the users through interactions. Unlike AgAsk, AgronomoBot is highly tailored to information from the sensors. It cannot answer broader questions regarding growing decisions; users still need to interpret the results provided by the bot for decision making.

FarmChat [4] provides farming related information through natural speech and text interactions. FarmChat is specifically designed to help potato farmers in India and relies on an expert-curated knowledge base of answers. While it helps to answer grower questions similar to AgAsk, it is highly tailored to one crop and one region; unlike AgAsk which is both crop and region agnostic.

Similar to FarmChat, AgroBot allows farmers to pose questions in natural language and get answers from an underlying agricultural knowledge base [2]. Thus, unlike AgAsk, the knowledge base is not backed by a comprehensive collection of scientific evidence while requiring the manual curation of a domain-specific knowledge-base.

7 HOW WILL AGASK HAVE IMPACT?

The impact of AgAs is three fold: 1) it enables growers, consultants, and domain experts to get accurate answers to their questions via interactions with the agent; 2) The answers provided by the AgAsk agent are gleaned from scientific reports, manuals, and research journal articles; thus providing an scientific and evidence-based source for possible answers; and 3) The AgAsk agent integrates the

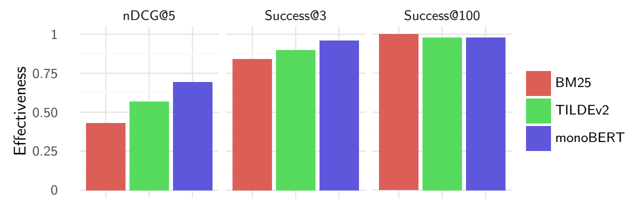

Figure 4: Results of the three considered retrieval methods on the 50 test topics. Differences between BM25 and the other methods are significantly different for nDCG@5 (p<0.01); for Success@3 they are significant only for monoBERT (p=0.012).

latest developments in neural information retrieval for the effective retrieval of agricultural advice and the generation of answers. These three aspects of AgAsk provide a useful tool and to foster future research in the agricultural conversational search agent direction.

We also plan to continue development with directions and assistance from domain experts. Currently, we use the BART answer generation in a zero-shot setting; further fine-tuning should improve answer generation effectiveness. We plan to exploit aspects of the conversation setting by taking into account previous utterances to improve answer generation. Finally, having answers personalised — e.g., by location, crop type, etc. — would allow for far more effective answers.

REFERENCES

[1] Manlio Bacco, Paolo Barsocchi, Erina Ferro, Alberto Gotta, and Massimiliano Ruggeri. 2019. The Digitisation of Agriculture: a Survey of Research Activities on Smart Farming. *Array* 3 (2019), 100009.
[2] Neelam Chandolikar, Chirag Dale, Tejas Koli, Mayank Singh, and Tarussh Narkhede. 2022. Agriculture Assistant Chatbot Using Artificial Neural Network. In *2022 ICACTA*. IEEE, 1–5.
[3] Danqi Chen and Wen-tau Yih. 2020. Open-domain question answering. In *ACL*.
[4] Mohit Jain, Pratyush Kumar, Ishita Bhansali, Q Vera Liao, Khai Truong, and Shwetak Patel. 2018. FarmChat: a conversational agent to answer farmer queries. *IMWUT* 2, 4, 1–22.
[5] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *arXiv* (2020).
[6] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv* (2019).
[7] Yongqi Li, Wenjie Li, and Liqiang Nie. 2022. Dynamic Graph Reasoning for Conversational Open-Domain Question Answering. *TOIS* 40, 4 (2022), 1–24.
[8] Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira. 2021. Pyserini: A Python toolkit for reproducible information retrieval research with sparse and dense representations. In *SIGIR*.
[9] Rob Lokers, Rob Knapen, Sander Janssen, Yke van Randen, and Jacques Jansen. 2016. Analysis of Big Data technologies for use in agro-environmental science. *Environmental modelling & software* 84 (2016), 494–504.
[10] Gustavo Marques Mostaco, Icaro Ramires Costa De Souza, Leonardo Barreto Campos, and Carlos Eduardo Cugnasca. 2018. AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks. In *ICPA*, Vol. 24. 1–13.
[11] Rodrigo Nogueira, Wei Yang, Jimmy Lin, and Kyunghyun Cho. 2019. Document expansion by query prediction. *arXiv* abs/1904.08375 (2019).
[12] Chen Qu, Liu Yang, Cen Chen, Minghui Qiu, W Bruce Croft, and Mohit Iyyer. 2020. Open-retrieval conversational question answering. In *SIGIR*. 539–548.
[13] Harrisen Scells, Shengyao Zhuang, and Guido Zuccon. 2022. Reduce, Reuse, Recycle: Green Information Retrieval Research. In *SIGIR*. 2825–2837.
[14] Matthew J Smith. 2018. Getting value from artificial intelligence in agriculture. *Animal Production Science* (2018).
[15] Jim Virgona, Geoff Daniel, et al. 2011. Evidence-based Agriculture-Can We Get There? *Agricultural Science* 23, 1 (2011), 19.
[16] Hamed Zamani and Nick Craswell. 2020. Macaw: An extensible conversational information seeking platform. In *SIGIR*. 2193–2196.
[17] Shengyao Zhuang and Guido Zuccon. 2021. Fast passage re-ranking with contextualized exact term matching and efficient passage expansion. *arXiv* (2021).
[18] Shengyao Zhuang and Guido Zuccon. 2021. TILDE: Term independent likelihood moDEl for passage re-ranking. In *SIGIR*. 1483–1492.