



AgriKG: An Agricultural Knowledge Graph and Its Applications

Yuanzhe Chen¹, Jun Kuang¹, Dawei Cheng^{3,4}, Jianbin Zheng¹,
Ming Gao^{1,2(✉)}, and Aoying Zhou¹

¹ School of Data Science and Engineering, East China Normal University,
Shanghai, China

{yzchen,jkuang}@stu.ecnu.edu.cn, {jbzheng,mgao,ayzhou}@dase.ecnu.edu.cn

² Key Laboratory of Advanced Theory and Application in Statistics
and Data Science - MOE, Shanghai, China

³ Shanghai Jiao Tong University, Shanghai, China
dawei.cheng@sjtu.edu.cn

⁴ Keydriver Inc, Shanghai, China

Abstract. Recently, with the development of information and intelligent technology, agricultural production and management have been significantly boosted. But it still faces considerable challenges on how to effectively integrate large amounts of fragmented information for downstream applications. To this end, in this paper, we propose an agricultural knowledge graph, namely AgriKG, to automatically integrate the massive agricultural data from internet. By applying the NLP and deep learning techniques, AgriKG can automatically recognize agricultural entities from unstructured text, and link them to form a knowledge graph. Moreover, we illustrate typical scenarios of our AgriKG and validate it by real-world applications, such as **agricultural entity retrieval**, and **agricultural question answering**, etc.

1 Introduction

Agriculture is the industry that accompanied the evolution of humanity, and fulfilled faithfully its core mission of the food supply. With decreasing workforce in the rural areas, advancing in the artificial intelligence, and developing the IoT technologies, it is desired to improve the efficiency and productivity of the agricultural industry. An agricultural knowledge graph repository will work as the foundation to achieve these goals.

Knowledge graph, which can be general-purpose and domain-specific, is a backbone of many applications, such as search engine, online question answering, and knowledge inference, etc. As a result, there are various knowledge graphs, including Wikidata¹, DBpedia², etc., for accessing to structured knowledge. Although there are some general knowledge graphs which contain some entities

¹ <https://www.wikidata.org>.

² <https://wiki.dbpedia.org>.

and relations about the agriculture, there is not a domain-specific knowledge graph for agricultural applications.

With the development of Web and IoT techniques, a wealth of fragmented data is crawled from Internet, generated by sensors or collected by agricultural drones. It is helpful and valuable to extract the agricultural knowledge from the fragmented data. Based on the agricultural knowledge, farmers will be able to take more informed and rapid decisions, make decisions to maximize return on crops, and be provided the advice and recommendations on the specific farm problems. Therefore, in this paper, we demonstrate an Agricultural Knowledge Graph in Chinese, namely AgriKG, which can be applied to support some agricultural applications, and further improve the efficiency and productivity of the agricultural industry.

The goal of this Demo system can be summarized as follows:

- Automated knowledge growth: AgriKG is able to identify the agricultural entities and relations from raw text, and incrementally adds the incoming knowledge triples into the knowledge base.
- Agricultural entity retrieval: AgriKG provides the entity retrieval in different fashions. Users are allowed to retrieve the agricultural entities via submitting a keyword search or image retrieval.
- Agricultural question answering: to enable AI-driven agriculture, AgriKG is able to address the questions via applying the subgraph matching.

2 System Overview and Key Techniques

As illustrated in Fig. 1, AgriKG consists of five key components: (i) crawlers collect the raw text and semi-structured data from Web; (ii) NLP module is a key component which provides a set of tools for the raw text understanding; (iii) entity recognition identifies the agricultural entities from the raw text; (iv) relation extraction aims at finding the attributes of entities and extracting relations from the raw text; (v) the applications of AgriKG include agricultural entity retrieval and question answering, etc.

Crawler. To construct an agricultural knowledge graph, AgriKG crawls the taxonomy from Wikidata, collects the attributes and images about entities from Hudong Baike, and acquires the massive agricultural raw text from some agricultural Web sites, such as China Agriculture, Xinnong, China National Seed Association, etc.

NLP Module. Since massive agricultural information appears in the raw text, the NLP module is applied to extract information, understand the raw text. It consists of a set of tools, such as text representation [1], word segmentation, and POS tagging [2], etc.

Entity Recognition. All entities in AgriKG is grouped into 16 predefined categories, including *animal*, *plant*, *chemical*, *climate*, *agricultural products*, *disease*, *nutrients*, *agricultural implements*, *agricultural terminology*, etc. Given a piece of

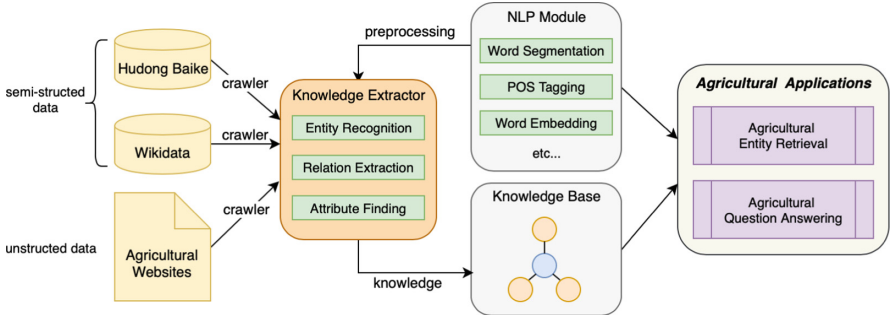


Fig. 1. The framework of AgriKG

text, we enumerate all spans, which are considered as the candidates of entities, after word segmentation and POS tagging. If a span is an entry in Hudong Baike, it is considered as an entity, and further classified into one of the 16 categories. In addition, to collect the ground-truth data for training, an auxiliary tool is developed to help the entity annotation.

Relation Extraction and Attribute Finding. One part of relations, such as *instance of*, *has part*, *subclass of*, *parent taxon*, *material used*, *natural product of taxon*, etc., in AgriKG extracts from Wikidata. The other part of relations, including *suitable planting*, *growing climate*, etc., extracts via using the remote-supervised approach [3] to train a neural relation extractor [4]. All entities and relations are stored in Neo4j, and the remaining data is stored in MongoDB.

Agricultural Applications. To achieve the precision farming, we develop two smart agricultural applications: agricultural entity retrieval and question answering for agricultural knowledge.

To support smart farming applications, such as weed monitoring and pest controlling, users can retrieve agricultural entities via submitting the traditional keyword search or image retrieval. For a keyword search, AgriKG returns the exactly matched entity. For an image retrieval request, AgriKG recognizes the most similar entities via using ImageMatch API and Elasticsearch for image similarity search³.

AgriKG also provides question answering, which consists of three key components: entity linkage, user intention understanding and answer ranking. A question request will trigger AgriKG to recognize the entities mentioned in the question [5]. Furthermore, the user intention is modelled as a multi-constraint question graph. It will be constructed based on the detected entities after the question annotation [6]. By doing so, question answering is transferred into a subgraph matching problem. Finally, after the ranking scores of candidates are calculated by a Siamese convolution neural network (CNN) [6], the answer will be subgraphs of the knowledge graph with the largest ranking scores.

³ <https://image-match.readthedocs.io>.

3 Demonstration Scenario

Our constructed AgriKG consists of more than 150,000 entities and 340,000 relations. To demonstrate the system, our GUI not only visualizes the architecture, but also lets the users interact with it.

Knowledge Extraction. In AgriKG, the raw text is crawled from the Web, and the extracted knowledge will be stored into the knowledge base. To illustrate the process of knowledge extraction in AgriKG, when a piece of text in Chinese is given, Fig. 2(a) demonstrates the recognized entities and extracted relations from the input text.

Entity Retrieval. In AgriKG, we can retrieve entities from the knowledge graph in two manners. In the traditional manner, it is a keyword search, which returns the exactly matched entity to us. In the other manner, it is an image retrieval. We can require an image retrieval in AgriKG when we have some photos of plants or pests. When we upload a picture of *agave*, Fig. 2(b) illustrates the result of the image retrieval. AgriKG will tell us exactly what the species is. With this functionality, we can identify unknown species whenever and wherever, and access to corresponding agricultural knowledge, such as planting strategy, pest controlling, etc.

Question Answering. Users are allowed to ask some simple questions (only involving single relation) or multi-constraint questions (involving multiple relations). AgriKG transfers a question into a multi-constraint query graph, and returns the most similar subgraphs via subgraph matching. Figure 2(c) demonstrates the answer of question “*what plants are suitable for growing in Chongming County*”. Therefore, AgriKG enables us to obtain the answers of agriculture-related questions in real time.



Fig. 2. System demonstration

4 Conclusion

To overcome the challenges on how to effectively integrate large amount of information for agricultural applications, in this paper, we propose a knowledge-based system, namely AgriKG, to automatically integrate the massive agricultural information into a knowledge graph, and to provide some services, such as agricultural entity retrieval, agricultural question answering, and so on.

Acknowledgments. This work has been supported by the National Key Research and Development Program of China under grant 2016YFB1000905, and the National Natural Science Foundation of China under Grant No. U1811264, 61877018, 61672234, and 61502236. It has been also supported by the Shanghai Agriculture Applied Technology Development Program, China (Grant No. T20170303).

References

1. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint [arXiv:1607.04606](https://arxiv.org/abs/1607.04606) (2016)
2. Li, Z., Sun, M.: Punctuation as implicit annotations for Chinese word segmentation. *Comput. Linguist.* **35**(4), 505–512 (2009)
3. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: *ACL*, pp. 1003–1011 (2009)
4. Lin, Y., Shen, S., Liu, Z., Luan, H., Sun, M.: Neural relation extraction with selective attention over instances. In: *ACL*, vol. 1, pp. 2124–2133 (2016)
5. Yang, Y., Chang, M.: S-MART: novel tree-based structured learning algorithms applied to tweet entity linking. In: *ACL 2015*, pp. 504–513 (2015)
6. Bao, J., Duan, N., Yan, Z., Zhou, M., Zhao, T.: Constraint-based question answering with knowledge graph. In: *COLING 2016*, pp. 2503–2514 (2016)