UrbanFoodKG: human-in-the-loop and LLM-assisted knowledge graph engineering for integrated food-system exploration

by

Aidan J. Beesley

B.S., Kansas State University, 2024

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Computer Science
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Major Professor
Dr. Hande Kucuk McGinty

Copyright

Abstract

The increasing population of the United States and the decline in available farmland has resulted in a rise in urban agriculture. Urban food systems encompass all aspects of growing, managing, and distributing food grown in urban environments. The wide array of urban food systems data, often collected in heterogeneous surveys, reports, and studies, presents a challenge for those gathering the data to organize and understand it. A lack of understanding prevents effective decision-making, resulting in less-than-desirable urban agricultural outcomes.

This report presents the UrbanFoodKG ontology, created as a part of the GRIP Resilient Urban Food Systems project, meant to assist in collating the information gathered and answering what domain-specific questions the experts have. The UrbanFoodKG knowledge graph takes domain-expert knowledge and the survey data they gathered, using the Knowledge Acquisition and Representation Methodology (KNARM) to produce an application based on the UrbanFoodKG ontology that allows users to reason over the urban agricultural data. The application currently supports 37 domain-expert competency questions.

UrbanFoodKG provides an initial infrastructure for standardizing and connecting urban agriculture data while enabling future applications in policy analysis, food-system resilience, and natural-language question answering. The computational approaches employed during this process are geared towards understanding the Human-in-the-loop LLM-assisted Knowledge Graph Engineering better used in concert with the KNARM methodology. Although the current evaluation is limited by the available survey data, the framework can be extended as additional datasets, domain knowledge, and policy information become available.

Table of Contents

List of Figures

Acknowledgments

I would like to thank Dr. Weese, Dr. Amtoft, and Dr. McGinty for taking the time to serve on my graduate committee. Time is a valuable thing, and I'm honored to have received theirs. This is especially true for my advising professor, Dr. McGinty. As many meetings as we could fit in, and this paper made it across the finish line.

I would also like to thank the Urban Food Systems project, lead by Dr. Pliakoni. This paper wouldn't exist without your project, and I hope I've contributed to it admirably, even if I wasn't there for the beginning of it. Aryan Dalal deserves special thanks for acting as a mentor to me on this project.

I'd like to thank Maitland E. Smith Scholarship House, for providing me the environment I've needed to flourish these past four years I've spent at K-State. It's been my home away from home, and I'm sad to leave it now. I'd like to thank my close friends, particularly Zach Moser, Zach Henson, Jade, Thomas, and Nate. Thank you all for helping to keep me sane and not buried too deeply in my work. Lastly, I'd like to thank my family for supporting me through all of this, both financially and even more importantly, supporting me emotionally. The love from my family is a cherished thing, and I wouldn't trade it for the world. Especially my parents, Frank and Sheila Beesley. You both believed in me even when I didn't believe in myself. Thank you.

Chapter 1

Introduction

Urban agriculture is the practice of cultivating food in urban environments, as opposed to the rural environments traditionally associated with growing food. This process can take multiple forms, including but not limited to the use of greenhouses, rooftop gardens, and community gardens. Methods of urban agriculture have existed as far back as the earliest cities of humanity, with modern interest increasing in the 20th century, first with a focus on wartime rationing during the world wars and poverty relief during the Great Depression, then focusing on increasing sustainability into the latter half of the century[2 3]. Despite the mental image that small community gardens may conjure, modern urban agriculture methods can produce comparable amounts of food and can even surpass traditional forms of agriculture per land area for certain crops[4 5]. Modern urban agriculture is growing in popularity as well, having grown over 30% in the last 30 years according to the Senate Agriculture Committee[6].

Urban agriculture can serve a vital role in reducing the impact of food deserts and lowering food prices[7 8], but urban agriculture can have numerous benefits to the surrounding communities beyond simple food production. Urban agriculture also contributes to greater economic development[9 7]. The presence of greater green spaces can cool the microclimate of the surrounding area by 0.5-4°C[10 9], increasing the quality of life of those nearby. Those living nearby will also have a greater sense of community[9 11] and overall healthier lifestyles[9 12] compared to their peers. The end result of an urban farm in an area is revitalization.

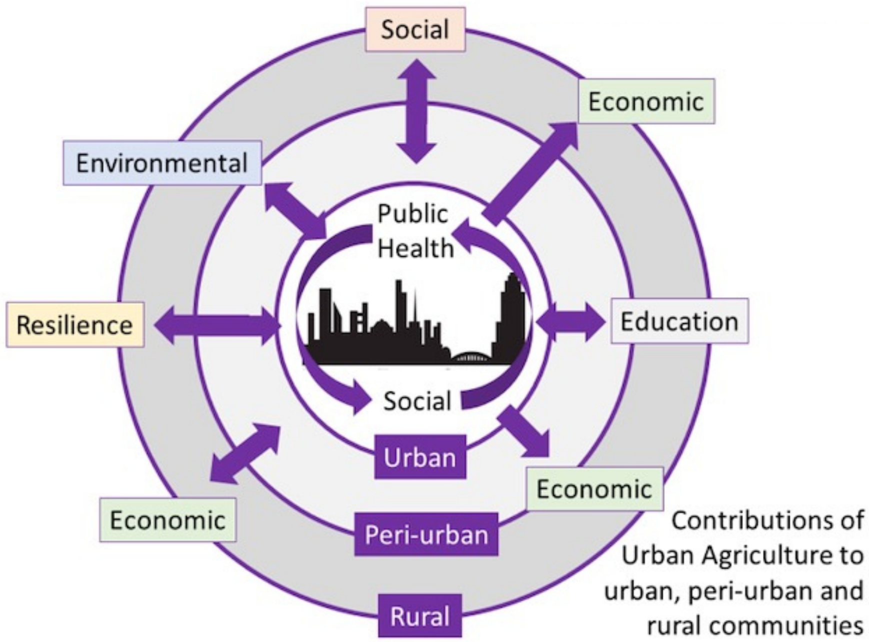

Figure 1.1: *GRIP Urban Food Systems project diagram, detailing the interdisciplinary nature of the project*

Kansas State University's Game-changing Research Initiative Program (GRIP) is a $3 million project from K-State to promote transdisciplinary teams to research problems facing the world and that may have a positive impact on the economy of Kansas[13]. The group that my project has been involved with is "Development of Resilient Urban Food Systems That Ensure Food Security in the Face of Climate Change," which brought together a wide array of specialists to work on it. The particular transdisciplinary aspect of the wider Urban Food Systems project that my report seeks to assist in is producing reasonable action from the data that has been gathered from surveys conducted by other members of this group. The way that the information will be processed and reasoned over is with an ontology-based knowledge graph.

Knowledge graphs are formal representations of terms and the relations among them using formal logic[14], such as "compost is_a typeOf fertilizer." Terms are held in nodes of the graph, while the relations between them are stored in edges between the nodes, creating a spiderweb-like graph from which insights can be gained. The terms used in a particular knowledge graph are gathered in ontologies, which take all terms relevant for a certain topic and create a bank of terms such that they are easily machine-understandable and are ready for further analysis and reasoning tasks that can be applied using applications downstream. Ontologies can be perceived as the schemas around which knowledge graphs are built[1]. My project aims to create a system that enables users to reason over information gathered from urban agriculture surveys to make better decisions, using a knowledge graph and ontology schema. Potential users of such software could range from individual operators seeking ways to improve their farms to policymakers using it to guide their decision-making on urban agriculture policy.

The rest of my report will consist of a review of relevant academic literature, the data and methodology behind my ontology and knowledge graph, an overview of the results of the knowledge graph towards answering the experts' competency questions, and the future work this research could support.

Chapter 2

Related Work

In this section, I will review the related work, divided into two sections: one covering agricultural ontologies and one covering automated ontology creation. The agricultural ontology section was initially going to cover any existing urban agricultural ontologies, but the lack of urban agricultural ontologies proved to be a major gap in the literature and thus the category was expanded. Despite that, agricultural ontologies is only a small section of the literature that exists. The automated ontology generation section is a much more well explored topic in the literature, both pre- and post- popularization of LLMs for the purpose of ontology generation, thus allowing for a larger literature review with greater depth of coverage of the different methods used. Given that my report does feature LLM generation, the papers covered are biased towards that topic for the easier points of comparison between my work and other work done in the field.

2.1 Agriculture Ontologies

In the paper "An effective automated ontology construction based on the agriculture domain," authors Deepa and Vigneshwari propose a method of creating an ontology with a Jaccard relative extractor for natural language processing focused on individual subdomains of agriculture (i.e. soil) which are then combined into a larger unified ontology with a Naive

Bayes classifier algorithm[15]. The methodology begins with a preprocessing phase to tokenize, tag, and organize the data before it is fed to the Jaccard relative extractor to see how closely related certain words and phrases are. The extractor was looking at four categories from which to draw relations, a general ontological use and use in each of the subdomains of soil, weather, and pests. From that output, the ontology can be created. The resulting ontology was evaluated on precision, recall, F1 measure, and ROC area, and compared against the decision tree and k-nearest neighbors methods of generating an ontology. Deepa and Vigneshwari's proposed method showed superior performance on all metrics. Compared to the work done for my project, the ontology is focused on more traditional methods of agriculture rather than urban agriculture. Additionally, the sources used for the methodology were all based on Indian agriculture statistics, which may lead to challenges in trying to apply the specific ontology created to an American context given the multitude of differences between the countries, though the methodology itself would still be applicable. The ontology generated is also broad in focus, starting directly from subject-matter documents, rather than competency questions from domain experts as this ontology did. Lastly, the ontology is generated in an entirely different way than the LLM method used in my project, in part due to the paper's publication year of 2022, before the large increase of use in LLMs for ontology building.

In "Innovative agriculture ontology construction using NLP methodologies and graph neural network" by Saravanan and Bhagavathiappan, the authors propose three methods of natural language processing to construct an agricultural ontology[16]. The initial input data consisting of domain-relevant unstructured textual documents is put through an anaphora resolution to ensure that no data is lost when put through the three ontology creation pipelines. The first methodology uses a NLP based algorithm on input documents to gather a series of candidate terms, which are then filtered, tagged, and reduced to a final list which an ontology is made from. The relationships between terms in that ontology would then be created by the use of Hearst patterns, which recognize word structures in text. The second method uses a Bidirectional Encoder Representations from Transformers (BERT) machine learning model to extract terms from the input documents, using the Hearst patterns to determine the relationships between terms. The third methodology uses the BERT model to extract terms, and an unsupervised Graph Neural Network (GNN) to determine the relationships between terms. The three ontologies created by these methodologies were then evaluated for accuracy by domain experts, of which the BERT and GNN methodology performed the best. The work done in my project differs from that done by Saravanan and Bhagavathiappan in purpose, technology, and inputs. My work is designed to create an urban agricultural ontology, specifically designed to answer the competency questions proposed by experts as a part of the Urban Food Systems project, whereas the authors' main goal is testing their methodology. LLM generation with a human in the loop was used for the creation of the ontology, while the authors use a mix of NLP algorithms, machine learning, and neural networks. Finally, the paper uses unstructured textual documents for input rather than quantitative data and competency questions as my work did.

In "Automatic relationship extraction from agricultural text for ontology construction" by Kaushik and Chatterjee, the authors propose a two-step system of designing agricultural ontologies[17]. The first is using a combination of regular expressions and NLP to extract vocabulary, then identifying semantic relationships in the second step with a reasoning algorithm to fully construct the ontology. The first step, term extraction, is done with the use of the RENT algorithm, requiring regular expressions and a list of twenty patterns to find the list of candidate words, i.e. "production of candidate-word", which will be used in the final ontology after being manually trimmed down. Once the list of candidate words is produced, a modified version of the Open Information Extraction algorithm is used to extract the relations between terms though NLP, where it is assisted in the paper's main contribution, the RelExOnt algorithm. RelExOnt uses domain expert knowledge to gather a possible set of relations between terms and the rules for identifying the relationship. The algorithm will then move on to processing potential relationships with those rules to see if the relationships are valid. With the terms and relationships all defined, an ontology is made. The end result of this process has an average precision of 86.89%, as evaluated by domain experts. Compared to the work done for my project, Kaushik and Chatterjee's approach is an algorithmic one rather than an LLM generated one, which is where the largest difference lies. The other major difference continues to be subject matter, as the ontology made by the authors is one about crops, rather than urban agriculture, though the innovations made by Kaushik and Chatterjee aren't restricted to such a topic.

As detailed above, the lack of an ontology specifically for urban agricultural purposes is a large gap in the currently existing literature. Urban agriculture faces many challenges that may be less prominent or completely absent in traditional agriculture, such as industrial contamination or zoning issues. Additionally, many of the ontologies created in this section were more focused on examining a new methodology and happening to use agriculture as an example rather than being focused on agriculture outcomes. A focus on the agricultural applications is another factor setting this work apart. Furthermore, from a methodological point of view, domain experts in the loop and an LLM-oriented approach were focused on compared to the existing entries in the field. While the ontology created for my project is specific to the purpose of the UFS project, and thus not suitable as an all-encompassing urban agriculture ontology, it provides a springboard for future work in this space done either by myself or other scientists in this field, a subject which will be discussed further in the Conclusion and Future Work section.

2.2 Automated Ontology Generation

In the paper "MASEO: A Multi-Agent System for Explainable Ontology Generation," authors Li et al. propose a system of combining multiple AI agents and existing ontology generation assistance technology to produce explainable ontologies[18]. The system contains an initial ontology generation agent, a syntax validation agent using RDFlib, a reasoning agent using HermiT, and a conflict resolving agent using the OOPS! scanner, where each agent has the capability to fix issues in the ontology without requiring a restart of the entire pipeline. Three ontologies were produced from MASEO, with varying levels of success in fulfilling competency questions ranging from 100% to 54.4% as evaluated by experts. The ontologies generated by MASEO tend towards a more narrow fitting of the competency questions compared to broader definitions in their respective fields that would lead to potential reuse. The authors acknowledge that the system has flaws, such as the inability for an ontology to be sent back to an agent once it has already passed it and the limiting of annotation to simple comments, but concludes with detailing what they would fix in future work. While MASEO does contribute to the field of ontology automation, my approach involves a greater involvement of a human in the loop. The MAESEO system has no human in the loop whatsoever, while my approach has several rounds of human and LLM back and forth in creation of the ontology, covering both the ontology writ large and specific sections. MASEO begins with strictly the competency questions as a starting point for ontology generation, while my approach begins with both competency questions and a human-made otology before any LLM generation.

The paper “Towards Automated Ontology Generation from Unstructured Text: A Multi-Agent LLM Approach” by Talukder, Mridul, and Seneviratne, takes a similar approach to Li et al. with MASEO in combining multiple AI agents with different roles to produce a single ontology[19]. This paper has four LLM agents in their pipeline; one for acting as a domain expert to synthesize a requirements document, one for acting as a manager to transform the domain expert requirements into a concrete steps for creating the ontology, a coder to implement the ontology from the manager's plan, and a quality assurance agent to ensure that there are no issues in the generated ontology with the assistance of tools like the HermiT reasoner and RDFLib. The paper uses a fifth LLM outside of their established pipeline for the generation of competency questions, though in other uses of this system this job could be done by a human. For the purposes of evaluation, Talukder, Mridul, and Seneviratne compared their multi-agent approach with a single-agent generated ontology. The comparison uses a pair of LLMs; the first to generate, validate, and execute a query based on a competency question, and the second to judge the query on how well it answers the competency question compared to a formal rubric. The first LLM can undergo multiple cycles to perfect its queries, and the amount of cycles it takes is another metric used to evaluate the ontology's competence in addition to the judge LLM's decision. To alleviate the potential issue of LLM accuracy in the evaluation step, the authors also use a vector-based RAG evaluation of the ontologies. The evaluation done uses two life insurance contracts that were fed into the proposed and baseline pipelines, where the multi-agent pipeline was shown to be more extensible, less redundant, and more structurally sound ontology. While Talukder, Mridul, and Seneviratne's approach is admirable, it has the same lack of a human in the loop that Li et al. have with MASEO compared to the work done here. Additionally, this pipeline has only been tested on unstructured textual data, whereas the data in my project is mostly quantitative. While this is not an inherent feature of their pipeline, the authors also use AI generated competency questions rather than domain expert generated questions, which provides another contrast between their work and mine.

In “NeOn-GPT: A Large Language Model-Powered Pipeline for Ontology Learning,” Fathallah et al. propose the integration of ChatGPT 3.5 into the existing NeOn methodology for ontology generation, testing their proposed workflow on a wine ontology with a gold standard ontology for comparison[20]. The workflow also integrates existing tools that aid in ontology generation, the HermiT reasoner and OOPS! pitfall scanner to serve as checks on the output ontologies. The Fathallah et al. adapt the NeOn methodology into the following steps for their work; the specification of the ontology requirements, the conceptualization and initial modeling of the ontology, and a full implementation of the ontology with triplets and accompanying metadata. After comparing the NeOn-GPT workflow with an ontology created from natural language description of the domain and an ontology creation guided by the Ontology Development Guide Prompts from Stanford, Fathallah et al. conclude that, while the NeOn-GPT workflow may fall short of the gold-standard ontology in depth and richness, it was vastly superior to the other AI generated ontologies it was compared to. In comparison to the work done here, NeOn-GPT starts from a completely different methodology, NeOn, whereas KNARM was used during the creation of this paper. Additionally, the NeOn-GPT workflow, as it is presented in the paper, has little in the way of information that is fed into the initial generation of the ontology. In my project, competency questions, quantitative data, and a hand-made, domain expert guided ontology were used as input for the generated ontology.

In the paper, “OntoRAG: Enhancing Question-Answering through Automated Ontology Derivation from Unstructured Knowledge Bases,” Tiwari, Lone, and Pal put forward a method of automated ontology generation combined with RAG with a focus on eliminating AI hallucinations[21]. OntoRAG consists of a six-step pipeline for the creation of an ontology; web scraping, PDF parsing, chunking, information extraction, knowledge graph construction, and ontology creation. Once an ontology is created, OntoRAG deploys a retrieval methodology to answer a user's query with the following steps; extracting key elements from the query, performing a similarity search of the ontology based on those elements, retrieving the most similar chunks of the ontology, and the final output is synthesized from those chunks. Twiari, Lone, and Pal evaluate OntoRAG by comparing it to two other forms of RAG, vector RAG and GraphRAG. The three different forms of RAG were compared by their work over a dataset of electrical relay documentation and using an LLM as a judge, showing OntoRAG to be 85% more effective than vector RAG and 75% more effective than GraphRAG. Compared to the work done on my project, OntoRAG starts from a complete blank slate, needing to scrape the web for sources. In my project, data had already been gathered by other teams of domain experts in the Urban Food Systems project and they were kept involved in the loop in the entire ontology creation process. Additionally, OntoRAG approached the creation of ontologies and knowledge graphs in the reverse order of my approach. OntoRAG derived an ontology from a knowledge graph, whereas my work used an ontology as a guiding model for a knowledge graph.

In the paper "Navigating Ontology Development with Large Language Models," Saeedizade and Blomqvist compare six different prompting techniques across eleven different LLMs to ascertain which is the best for automated ontology generation[22]. The prompting techniques tested were Zero-shot prompting, Sub-task Decomposed Prompting with Waterfall and Competency Question by Competency Question (CQbyCQ) variations, Chain of Thoughts, Self Consistency with Chain of Thoughts, and Graph of Thoughts. The best performing prompting technique was CQbyCQ, where the LLM addresses one competency question at a time with an individual ontology, with the final output being the combination of the individual competency question ontologies into one whole ontology for the domain. Out of the eleven different LLMs tested, the best one at the time of publishing was ChatGPT-4. To evaluate the prompting techniques and LLMs tested, a two stage testing method was used. First the authors examined if the LLM could create a model for the given tests regardless of the prompting method used, examining the result with either a reasoner inference verification depending on which test set was used. Only two models, GPT-3.5 and GPT-4, passed this test and were then assessed by specific prompting technique for accuracy by manual testing using a set of eight criteria where GPT-4 outperformed GPT-3.5. GPT-4 was then tested with different prompting methods, of which CQbyCQ performed the best. Both CQbyCQ and the work done on my project start from competency questions and domain information, though the information in Saeedizade and Blomqvist's work is textual data of college courses, rather than the quantitative data used here. Compared to the work done for my project, the CQbyCQ method does not have a human in the loop of ontology generation.

In the paper "Large language models for scholarly ontology generation: An extensive analysis in the engineering field," Aggarwal et al. evaluate four prompting strategies across seventeen different LLMs in an effort to examine the capabilities of automated ontology generation in engineering research[23]. The prompting strategies were a mix of zero-shot and chain-of-thought prompting combined with prompting the LLM to infer either a one-way or two-way relationship, for four total strategies. The method of evaluation for the LLMs and promptings is narrow, Aggarwal et al. task the LLMs with comparing two research topics and deciding a single relationship between them; broader, narrow, same-as, or other. The output of the LLMs were evaluated against a set of 1000 relationships from the IEEE Thesaurus, where they were assessed using precision, F1, and recall scores. Claude 3 Sonnet with zero-shot reasoning performed the best out of the models and prompting methods tested. Compared to the work done in this paper, the work done by Aggarwal et al. has a much narrower scope of outputs, only requiring the generation of four relationships compared to not only the relationships but some terms as well in the ontology made for this paper. The generation itself also didn't have a human in the loop, nor any competency questions involved in the generation in another contrast to the work done in this paper.

The reading above details some ways that the field of ontology development has previously attempted to automate ontology building along with how the advent of LLMs have contributed to that process. Given how laborious, time-consuming, and expert-dependent the field was before the advent of LMMs, adopting LLMs for building Knowlede Graphs, as the research in this section supports, improves the process of Knowledge Graph Engineering. Generative AI automation enables faster, more consistent ontology production with accuracy comparable to that of other forms of automation. But, given both the nature of large language models as general tools rather than specific ones and the fact that generative AI still has the risk of hallucinations, there is still value in a human in the loop. That is why this ontology was created using KNARM[1], a methodology that recognizes the usefulness of automation and the need for experts, both in creating ontologies and in the relevant domains, when making an ontology. The details of the methodology and the implementation of my LLM-aided approach are further detailed in the methodology section of this report.

Chapter 3

Methodology

This section will begin by reviewing the details surrounding data collection and the collection of competency questions that allowed me to evaluate the implementation of my methods. Then, I introduce and explain in detail how the semantic web approaches have been utilized in combination with the generative AI and statistical approaches in accordance with the KNARM methodology to create the ontology in use for my project. Finally, I discuss how these approaches were implemented as part of an application for user friendliness and the adoption of the underlying methods and data that were utilized as part of this research.

3.1 Data and Competency Questions

The GRIP Urban Food Systems (UFS) project is a project sponsored by Kansas State University meant to examine the current landscape of urban agriculture and build tools to assist in the long-term sustainability of such projects. UFS project data was used extensively in the creation of this report. Members of that project gathered data from urban agricultural sites across the state of Kansas and the Kansas City metro area. In total, 23 sites had survey data gathered about them.

All of this data was collected for the year 2024, and covers four broad topics; operations, community health, environment and sustainability, and owner/operator demographics. As a part of preprocessing of this dataset, all of this data was anonymized. This dataset is available upon request, and is integrated into the graph database detailed in the Application section below. The competency questions gathered from this research group were created from collaboration between the domain experts and ontology engineers working on the wider UFS project. The majority of the competency questions the domain experts provided were rooted in the environment and sustainability dataset, with many also requiring data from the operations survey. An example of a competency question is "Is the adoption of remediation measures at residential-use sites dependent on operator education and institutional support (Q57)?" The full list of competency questions can be found in Appendix A. After the data were collected, preprocessed, and made available for my project by the other members of the UFS project, some further processing was needed. Some aspects of the data that had been provided, such as redacted personal data and columns with no entries, were irrelevant to the ontology, and were deleted.

As I improved my familiarity with the data during knowledge graph and application development, a research finding emerged: a discrepancy between the competency questions the experts wished answered and what was possible to answer with the amount and types of data collected. Some of the questions ask for data over time, for example, "Is high upfront material costs [in using clay pebbles as a growing medium] justified by long-term reuse?" Survey data has only been collected in 2024, though this limitation could be mitigated if the UFS project were continued into the future. Some of the questions have an issue of measurement and metrics; many ask about the effects of different factors on yield, but there isn't a baseline in this dataset for what should be considered a normal yield for many of these sites. There isn't a large set of urban agricultural data that can serve as a baseline for these metrics either, and using regular agricultural data risks incorrect conclusions being drawn due to the innate differences between traditional and urban agriculture.

3.2 Ontology

In this research, an ontology is defined as a shared vocabulary of terms, such as classes, attributes, and relations, that is organized hierarchically and both human-understandable and machine-reasonable[14]. The purpose of an ontology is to organize data semantically, which is to say they are stored in the form of classes and relations, rather than logically with a traditional database schema. This allows for easier integration and interoperability with different databases and systems.

One of the ways in which an ontology works is through what are known as inferred hierarchies. This is where a concept will be described through a series of axioms, and the other relations of those axioms can be used to derive knowledge about the things they describe. For example, an ontology could have 'community gardens' as a concept. If, for the sake of an example, all community gardens had the axiom 'uses rainwater,' and I know that 'uses rainwater' is linked to 'rainwater industrial pollution risk,' then the ontology can confidently say that all community gardens are at risk of their rainwater being polluted with industrial contaminants. When applied across every axiom in an ontology, this allows for new insights to be revealed.

For building ontologies, there are many methodologies to choose from; NeOn, Methontology, or CommonKADS for example. With the above definition in mind, the methodology chosen for this paper is KNARM (KNowledge Acquisition and Representation Methodology), as it was previously established by a member of the team working on the UFS project[1]. As a methodology, KNARM has a focus on automation that aligns well with the modern landscape of ontologies that the popularization of LLMs has brought. KNARM still has a place for humans in the loop, though, both ontology experts and domain experts, to provide crucial assistance in the automated generation process. KNARM is also an agile methodology; there may be times in development where one needs to go back and reassess their ontology, such as if new information has come to light that would be useful to include in an ontology. KNARM recognizes this fact and includes it in the methodology.

For my project, KNARM was initially followed by the prototype ontology built for the

UFS project. Since some of the steps of the KNARM methodology had already been completed by the prototype ontology, and my project inherited the work done by the prototype, those steps did not need to be retread on this ontology. Additionally, due to the differences between the prototype ontology and the gradual changes in the final ontology that would be generated for my project, some of the steps in the KNARM methodology had to be revisited, as will be detailed in the appropriate sections below. Even with those factors, the KNARM methodology still served as the guideline by which my project was created, and as such, the remainder of this section will use it to explain how the ontology was produced. The steps followed will be listed and explained in their order in the KNARM methodology, even if a step has been skipped. If a step was skipped or if there were serious deviations taken in a step, then they will also be detailed in that section.

3.2.1 Sub-Language Analysis (1)

Sub-language Analysis is the process of identifying what information already exists and evaluating if it is relevant for the ontology development. This can be in the form of preexisting ontologies, domain-specific vocabularies, technical documentation, and published scholarly literature. For my project, given the prototype ontology that had already been developed, this step was not done any further than taking that ontology and using it as the basis for the ontology my project produced.

3.2.2 In-House Unstructured Interview (2)

The In-House Unstructured Interview is meant to help those creating the ontology better understand the data, domain context, and the objectives of the project[1]. This step was accomplished by the ontology engineers working on the previous ontology, the findings and insights of which were then inherited by this ontology. Thus, conducting a new In-House Unstructured Interview would have been irrelevant. Discussions were held with fellow ontology engineers and domain experts about my project and the work they had done on it as an introduction to the project, which could be considered analogous, even if it isn't strictly the

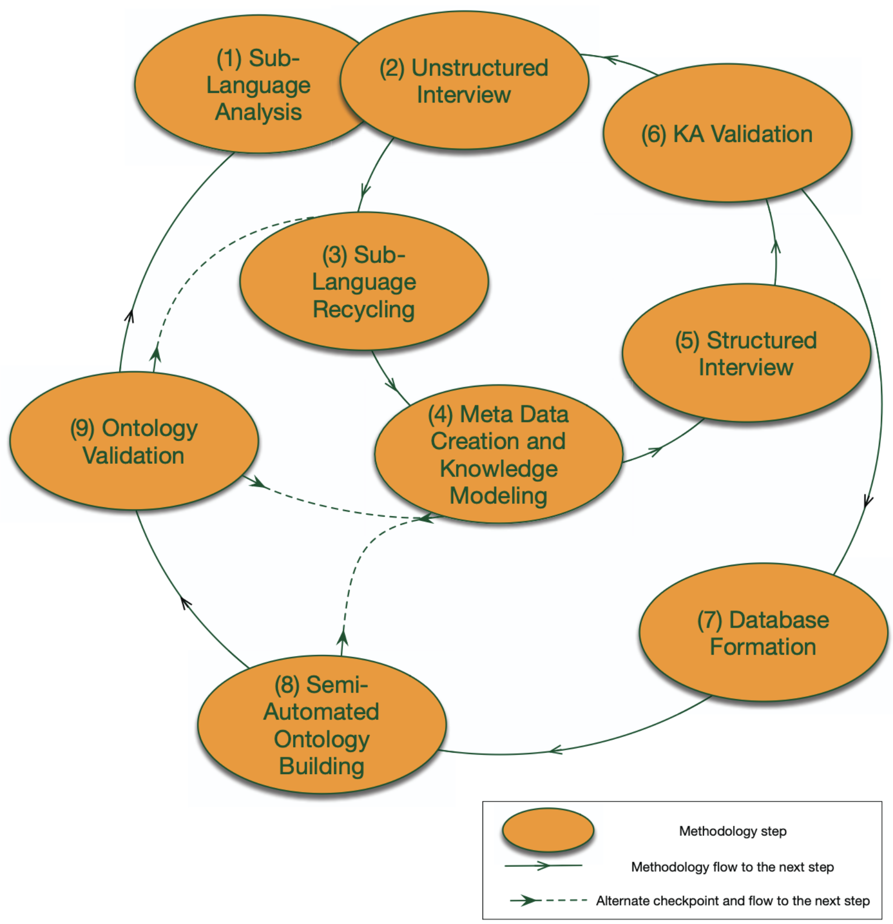

Figure 3.1: *The cycle of the KNARM methodology*[1]

same. The second interview built upon previously established project knowledge, and clarified processes identified during earlier stages of development. It helped to provide crucial knowledge about the project which was invaluable moving forward.

In the initial unstructured interview, a series of 324 competency questions was collected from domain experts in the field of urban agriculture. The experts consisted of professors, PhD candidates, and graduate students at Kansas State University from various departments participating in the UFS project. Each competency question was connected to one or more of the questions in the survey data that the UFS project gathered. For example, one survey question asks, "What was the previous use of the land on your urban agriculture site?" An associated competency question is, "Do farms with industrial land-use history show higher levels of soil or water contamination from toxic trace elements or hazardous waste?" (See Appendix A).

3.2.3 Sub-Language Recycling (3)

Sub-language Recycling is the reuse of existing ontological resources that may be relevant to the ontology in progress. This step exists to reduce development time and effort while keeping consistent with other experts in the field[1]. Sub-language Recycling was done in my project when I took in the work done on the previous ontology. The prototype ontology served as the basis for all of the future work done on my project. The prototype ontology was the only resource included in the sub-language recycling step of this process. The reasons for this were threefold. First, given the fact that it arose from the same project as this ontology, the one prototype ontology was deemed sufficient. Second, while there would have been other agricultural ontologies to use, none of them would have been specifically focused on urban agriculture, limiting their usefulness. Lastly, this ontology is meant to be an application ontology specific to the competency questions posed by the domain experts for the UFS project, rather than a general urban agricultural ontology. While this ontology could serve as a basis for a more generalized urban agricultural ontology, further discussions of that potential will be saved for the Conclusion and Future Work section.

Searching for domain-related databases is also included in the Sub-language recycling step. No urban agriculture databases, or general databases that may assist in answering the competency questions, were found in the course of this research as of the conclusions of this study. There do exist generalized agricultural databases, such as the USDA's National Agricultural Statistics Service, but those data sources don't have the capability of filtering for urban agricultural data, if they even collected that data. As a result, the only data provided for my project is the survey data provided from the UFS project. While in-depth on many aspects of the urban agriculture sites, the data only covers 23 sites. The small size of the dataset makes it difficult to draw conclusions on many of the competency questions asked by the domain experts. Further discussion of the results and difficulties therein are detailed in the Results section later.

3.2.4 Meta-Data Creation and Knowledge Modeling (4)

Meta-Data Creation and Knowledge Modeling is meant to create metadata to further describe the domain of the project and model the knowledge already created using axioms[1]. My project had no need of metadata, and the axioms either already existed or were created in the Semi-Automated Ontology Building step described later.

3.2.5 Structured Interview (5)

The Structured Interview step is designed to have the domain experts answer closed-ended, specific questions to answer important issues that may have arisen in the development of the ontology[1]. There was one structured interview that was held with the domain experts and prototype ontology developer after my project began. The purpose of that interview was fourfold; to clarify questions about the specifics details of the competency questions that existed, to more clearly define previously nebulous terms such as 'biodiversity' in the competency questions, to update the other members of the UFS project on the progress of the ontology, and to validate the ontology as it currently stood, as will be detailed in the next section. By the conclusion of that interview, all the issues with the competency questions that were known at the time were resolved, greatly contributing to the eventual understanding that would be reached about my project. There were a few smaller questions that were discovered after this interview, but those were comparatively simple enough and able to be handled with an email exchange rather than requiring a full meeting.

3.2.6 Knowledge Acquisition Validation (6)

The purpose of the Knowledge Acquisition Validation step is to receive feedback on the ontology development, identifying any issues in the ontology and/or the knowledge base, and allowing for revision before further steps are undertaken[1]. This step occurred at multiple points through the process of developing this ontology through both reviews of the ontology with my supervisor as well as reviews with domain experts, such as in the meeting discussed in the previous section. The meetings with the domain experts primarily served to validate the concepts, definitions, and relations that were present in the ontology so that it would be most useful for the overall UFS project. The meetings with my supervisor served to revise the structure of the ontology and how it might better fit the needs of the UFS project. These meetings lead to shifts in design priority for the ontology, most notably the shift from categorizing the sites as Individuals in Protege to categorizing the sites as separate Classes for the purpose of more easily reasoning over them.

3.2.7 Database Formation (7)

The Database Formation step is meant to build a database to serve as the representation of the knowledge gathered in the above steps[1]. The survey gathered by the UFS project was made available in the form of four csv files which were used extensively throughout this process, along with a fifth csv file of the competency questions. These files weren't collected into a formal relational database as they were small enough to still be human queryable and LLM parseable with ease. Additionally, no project metadata had been created about those files, so no unified databases were established. While my project does feature a Neo4j graph database for the backend of the application portion of my project later, that database was created from the ontology, which is the inverse of how the Database Formation step operates. As such, it will be discussed in the Application section below.

3.2.8 Semi-Automated Ontology Building (8)

Semi-Automated Ontology Building is the step that involves the building of the ontology itself[1]. The Semi-Automated Ontology Building step of the KNARM methodology was undoubtedly the step most revisited in the process of building the ontology. Ultimately over the course of my project, three ontologies were generated by ChatGPT-5.4, each building on the last after some manual revision and further refinement of the overall goal of the ontology. For all of these generations, the input consisted of the previous ontology, the .csv of competency questions, and the .csvs of survey data. The prompting method used for each was zero-shot prompting. I did not experiment with prompting techniques because I already established prompts that work with the methodology[24]. Furthermore, the literature is currently divided on what is the best method for prompting an LLM. Aggarwal et. al. found that zero-shot prompting scored best in their metrics[23], while Saeedizade and Blomqvist found their CQbyCQ method worked best[22]. The purpose of this work isn't to find which model or method would be the best, as is the purpose of many papers in the literature review section of this paper, but to use a model and method that would work.

The manual work that needed done in the ontologies were largely consistent across each generated ontology. ChatGPT consistently added either redundant classes into the ontology when equivalent classes already existed or classes that were redundant with those it had added in that generation. For example, in one of the first ontologies generated, ChatGPT created a Yield class despite a Yield class already existing in the ontology. I argue that this may be due to slight differences in the requests of competency questions; one may be primarily categorized under one survey question, but rely on a factor from a different survey question resulting in a duplicate class being made. Regardless of the reason why, the removal of these redundancies and consolidating them into one class was one of the largest sources of manual work done on the ontologies. A second bulk of manual adjustments were revising things from the prototype ontology that had since become redundant, unused, or unnecessary in the new direction my ontology was taking. Through each subsequent generation and discussion with the domain experts, the ontology for my project grew more focused on answering specifically these competency questions and less all-encompassing as the prototype's was, resulting in more and more parts of the ontology that didn't serve that focus.

3.2.9 Ontology Validation and Evolution (9)

The final step of KNARM is meant to test if the ontology is both accurate to the answers of the domain and consistent with itself[1]. The discussion of the ontology's accuracy is in the Results section.

3.3 Application

The application portion of my project serves as a user-friendly, accessible way of bringing the formal representation of this research dataset and the statistical methods applied on it to the users. Towards that end, the application consists of two parts; a Neo4j graph database backend, and a React frontend. The application, just as the ontology, was generated by ChatGPT then any necessary edits were made by hand.

Neo4j was the backend of choice for this application. It's one of the most widely used graph database applications, which makes it both adequate for modeling knowledge graphs and ontologies as well as easily interoperable with the React front end. The ontology created in the previous section, along with the survey data and competency questions, were loaded into a Neo4j database that was created by ChatGPT-5. The LLM closely modeled the ontology into the database, adding some metadata to nodes and connecting edges where necessary to include the quantitative data from the original sources that wouldn't have been able to be modeled in the ontology. The end result is a database that includes 1698 nodes and 11,043 edges between them. Additionally, the competency questions were also translated into

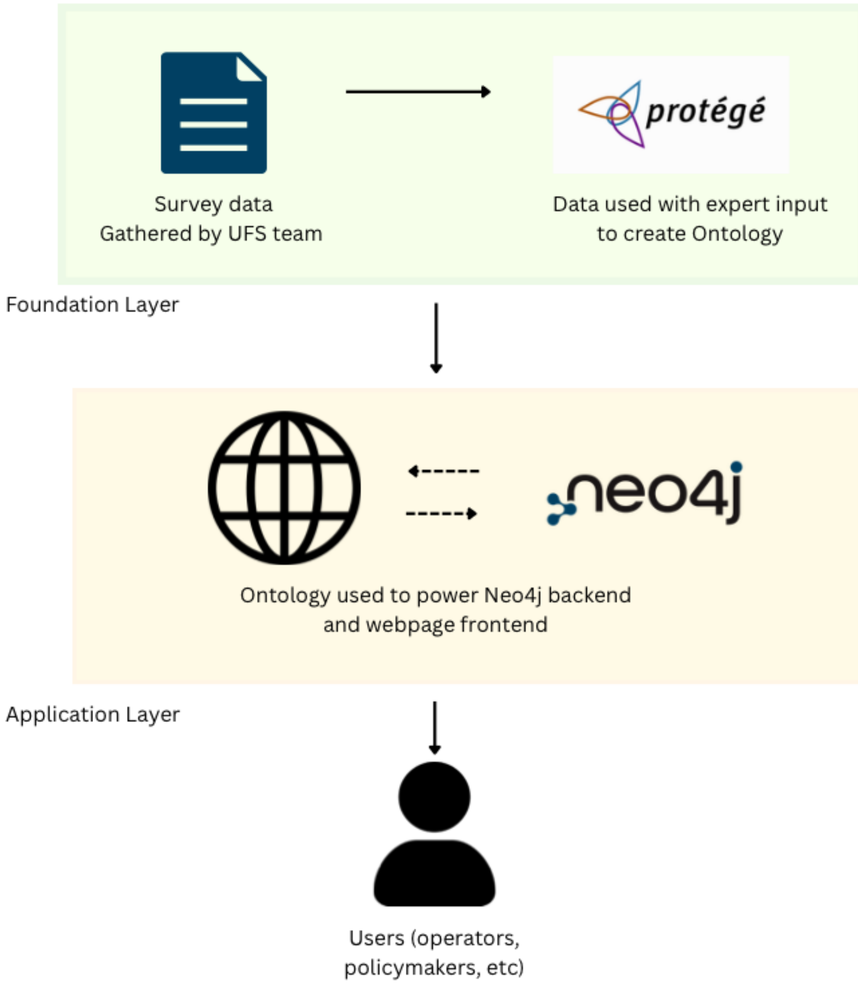

Figure 3.2: *A system diagram for UrbanFoodKG and the application*

Cypher queries via LLM generation. These queries required a large number of regenerations and manual adjustments for reasons that will be further discussed in the Results section.

The front end of this application received the least focus during my project, as the reasoning and ontology building are where the reasoning of my project occurs. The end goal of the UFS project was to provide something that urban agriculture operators could use to answer questions they may have, starting with the competency questions provided by the domain experts but eventually leading into a more open-ended system, though only the former has been completed as of now. That desire led to a web-based application, as it would allow for ease of access to all the distributed operators and any updates to the system would be easier to deploy and able to instantly be accessible by any operators using the application. React was specifically chosen for the previous experience I had with it, the depth of resources and tutorials available for development, and interoperability it has with Neo4j. The structure of the React web app consists of three parts; an Express server which connects to the Neo4j database, the user interface, and the competency questions, stored as both the Cypher queries mentioned above and as a simplified JSON form for use in the UI. Future versions of the front end may allow for users to input more free-form questions, which will be discussed more in the Conclusion and Future Work section of this report.

The methodology above provided the blueprint for the project that was followed throughout the entire process. The results of the development process will be discussed in the following section.

Chapter 4

Results

This section will begin by reviewing the results of the LLM-enhanced ontology building process. Then, the results of the development of the knowledge graph backend will be discussed. Lastly, the results of integrating the knowledge graph with the frontend application and the statistical approaches used for answering the competency questions will be shown.

4.1 Ontology Building

As detailed above, this ontology was built with the KNARM methodology in mind, with heavy usage of LLMs in the construction process, starting from a prototype ontology. The ontology resulting from this process had three major revisions as my project progressed, becoming more sharply focused on the competency questions with each revision. The final resulting ontology consists of four sections; two categories specifically related to the UFS project, and two categories that would be applicable for more general uses of this ontology.

The first of the general categories is *transdisciplinary components*, which has three subclasses; *government support*, *demographics*, and *environmental and sustainability components.* The first two only contain one relevant factor each, *demographics* containing the farm operator's education level and *government support* only containing if farms had such support, as those were the only statistics that the answerable competency questions queried. These

Ontology metrics:

Metrics

Axiom	1,872
Logical axiom count	502
Declaration axioms count	454
Class count	417
Object property count	36
Data property count	0
Individual count	0
Annotation Property count	5

Class axioms

SubClassOf	269
EquivalentClasses	163
DisjointClasses	7
GCI count	0
Hidden GCI Count	0

Object property axioms

SubObjectPropertyOf	0
EquivalentObjectProperties	0
InverseObjectProperties	0
DisjointObjectProperties	0

Figure 4.1: *Metrics that detail the scope of the ontology*

sections have the greatest possibilities for expansion in terms of classes contained, further discussion of which will be in the Conclusion and Future Works section later. The *environmental* subtree of the ontology contains information relating to both the farm's impact on the environment, and the environment's impact on the operations of the farm. An example of the farm's impact on the environment would be the impact on the local soil under the *soil erosion control method* class, and an example of the environment's impact on the farm would be the *pollution sources* class which may impact the farm's food safety.

The second of the general categories is the *urban ag production model* class. This class contains information immediately relevant to the operations of an individual farm. There are two subtrees of classes that contain detailed and expansive axioms in the knowledge graph which were generated based on data, questionnaires, and competency questions; *ag productivity components* and *infrastructure and practices.* The *ag productivity components details* aspects of a site immediately relevant to growing a crop, such as what medium the crops are being grown in (native soil, imported potting soil, hydroponics, etc). *Infrastructure and*

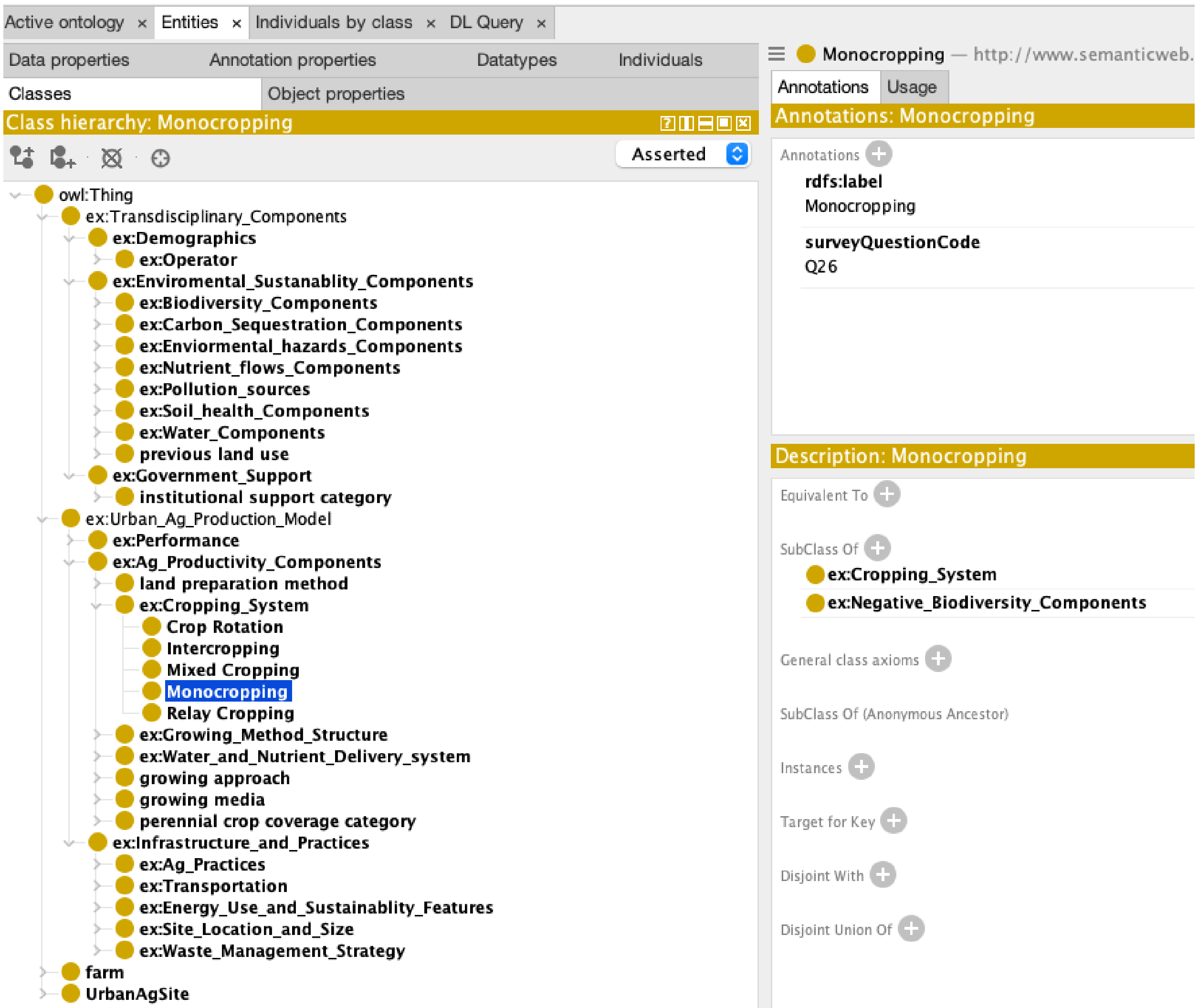

Figure 4.2: *A drop down list of many of the classes in the ontology*

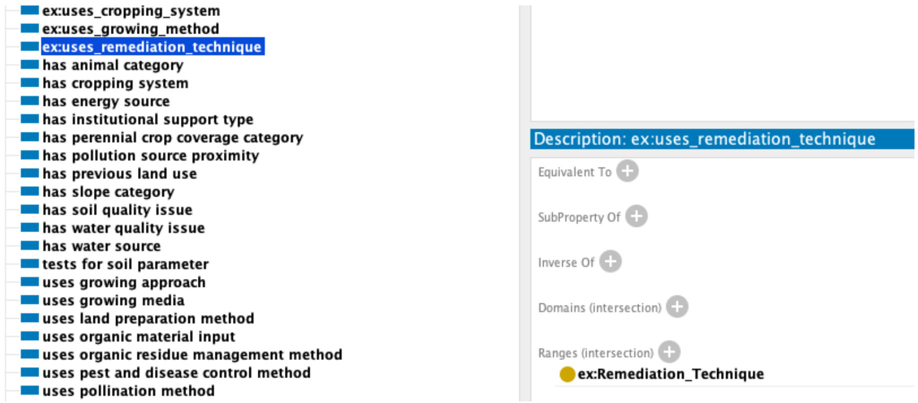

Figure 4.3: *An example of an object property and range*

practices details site-wide operational details, such as the adoption of remediation measures at the site. Additionally, there is one tree and one subtree that are not axiomatically optimized due to a lack of data making their competency questions impossible to answer. The *performance* subtree describes economic aspects of a farm, namely yield. It was created in an earlier draft of the ontology before the issues with the lack of yield data were recognized, but given the category's prominence in the competency questions this category remains in the ontology as a foundation for future work. The *transportation* class, as a subclass *infrastructure and practices*, is unused for the same reasons as *performance*, though unlike *performance*, *transportation* is much less prominent as a potential field of future research.

The first of the project-relevant categories is the *farm* category, which was defined to be equivalent to a specific site and defined through on specific axioms based on the data available through the surveys and competency questions. For example, the subclass *farm that has previous land use vacant land use* is used to answer the question 'Do farms on previously vacant land show higher risks of soil compaction, low organic matter, or illegal dumping?'. To determine which sites should belong in these categories, object properties are used. Object properties are properties of objects, in this case classes, in an ontology. Every property has a range, the scope of values that belong to the property. For example, the **uses remediation technique** property has a range consisting of the *remediation technique* class. The sites were encoded this way to assist in reasoning, namely seeing how the *UrbanAgSite* classes cluster under the given farm classes as dictated by domain expert interests.

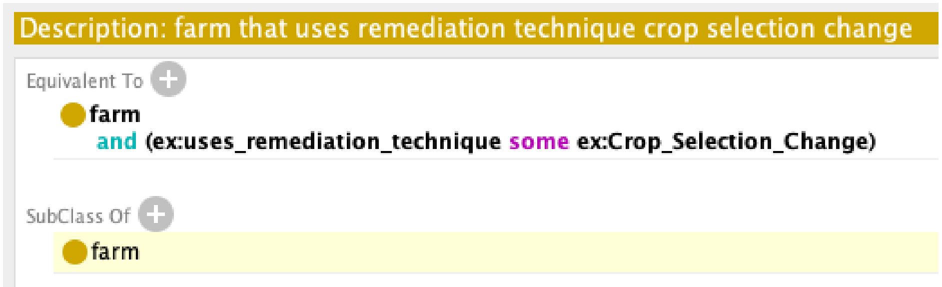

Figure 4.4: *An example of how range interacts with farm to produce classes for reasoning*

Figure 4.5: *All of the properties of Site_30UVITaHR8JEeBj*

The second project-relevant category is the *UrbanAgSite* tree, which contains each site the survey data covers as its own class. This data is encoded with the object properties described previously, so every site is equivalent to the totality of the properties of the site. Using the HermiT reasoner, these farms can then be inferred to be subclasses of a given farm category.

In 4.5, Site_30UVITaHR8JEeBj and all of its associated properties, i.e. **'uses organic material input' some Manure**, are shown. When used with the HermiT reasoner and the *farm* classes, you can see in 4.6 that not only is Site_30UVITaHR8JEeBj a subclass of *farm that uses organic material input manure*, that it is clustered with all of the other farms that use manure as well.

These two subtrees are the foundational modules that the final knowledge graph builds

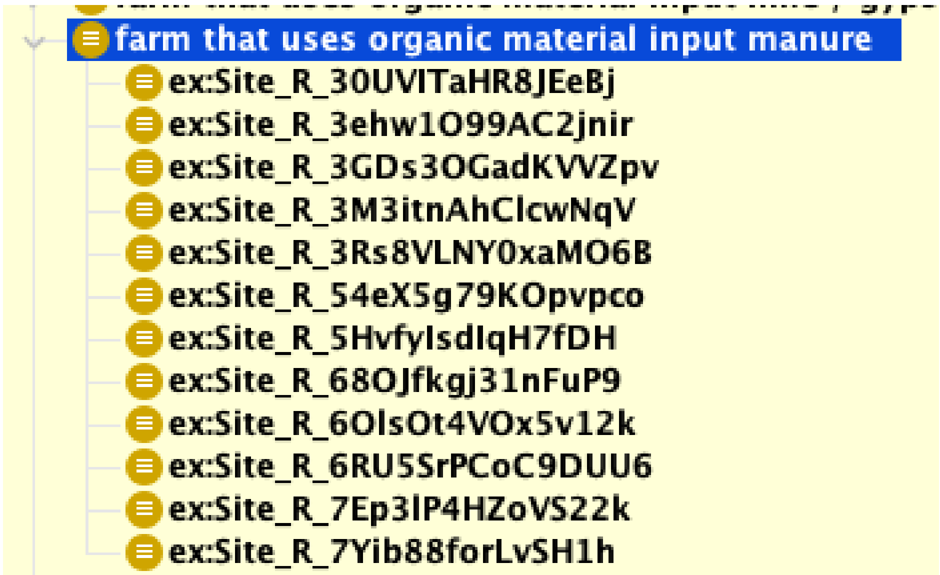

Figure 4.6: *The cluster of classes that use manure*

upon described in the next section.

4.2 Knowledge Graph

The knowledge graph portion of my project consists of the transformation from the ontology to the Neo4j database. The usage of the Neo4j database makes it easier to query aspects of the data compared to the Protege schema, and allows for future flexibility in the project compared to Protege.

The Neo4j database consists of three parts; the farm nodes, the relationships, and the research concept nodes. The farm nodes are similar in scope to the site classes of the ontology in that they are meant to represent each urban agriculture site surveyed. Unlike the site classes, the farm nodes contain more site metadata that isn't relevant in the ontology, such as the source of the information in the node. The relationships in the database are equivalent to the object properties in the ontology and have the same role of collecting the information about each site. The research concept nodes detail the general nodes discussed previously; they contain the possible values for a relationship.

There is a difference in the way that the relationships are managed between Neo4j and

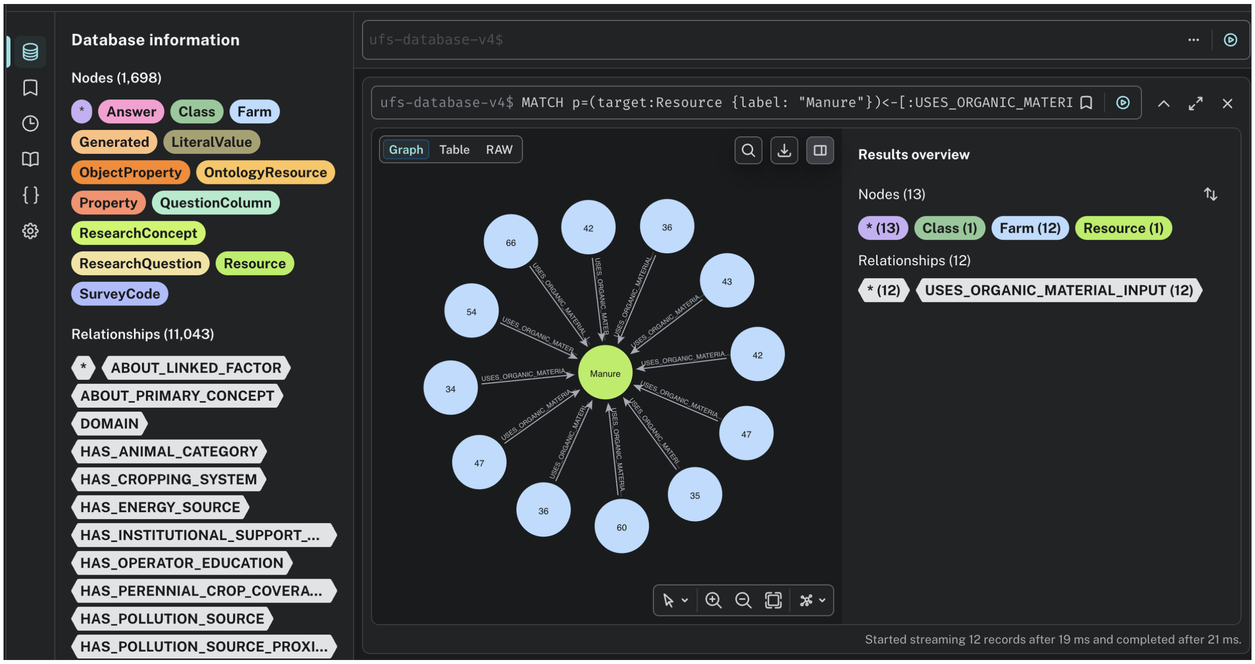

Figure 4.7: *An example of 4.6 in the Neo4j database*

the ontology. In the ontology, the object properties belong to each site class. In the graph database, the relationships are between the farm nodes and the research concept nodes. This has the effect of making the relationships much less strict. The proper use of the vocabulary is based on the literature used and the domain experts, and is the only thing preventing a farm from having a nonsensical relationship like ***HAS_PREVIOUS_LAND_USE*** → "No college".

The Cypher queries based on this knowledge graph and these relations, similarly to the ontology itself, required several generations and hand-adjustments to reach their final state. Part of this is due to discoveries that the failure of these queries revealed, such as some relations not being properly categorized in the database. Others were a result of LLM hallucinations that required more human reworking and more specific prompting to fix, like the assumptions of relations that didn't exist. Many of those relations were to do with the last major source of adjustment required, which was creating averages for certain statistics (i.e. biodiversity), as the Cypher queries assumed an explicit relation for those averages existed, when they were derived statistics.

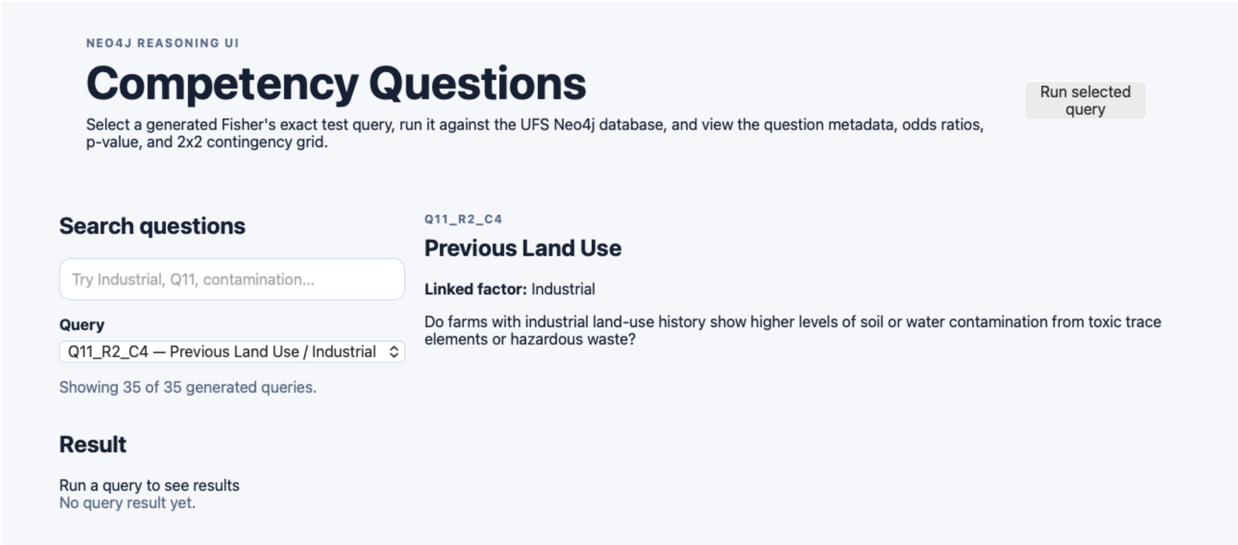

Figure 4.8: *The web front end of the application*

4.3 Application

The application consists of two portions; the knowledge graph back end, the construction of which was discussed above, and the user facing front end.

Currently, this front end exists as a simple self-hosted React web app. There is only one webpage, which allows a user to choose which of the 35 answerable competency questions they wish to see the answer to. These questions are answered with Fisher's exact test. Fisher's test uses an analysis of contingency tables, here a 2x2 grid, which displays the frequency of the outcome given its position on both axes. Using a Fisher's test, the odds ratio of a given test can be calculated, which is the odds of farms with the question's factor appearing in the group of farms where the outcome being tested for is present, where 1 is average odds, <1 indicates that the two have a negative correlation, and >1 means the two have a positive correlation. In the case where no farms may fill a given factor, the Haldane-Anscombe correction is used, which adds 0.5 to all cell counts. The outcome of a Fisher's test is compared against a null value, in this case always being that there is no relation between the factors being measured. Additionally, the p-value is calculated to determine the likelihood of the results being random.

Fisher's was chosen for two reasons. First, most of the competency questions are either directly stated as or can be easily reframed as seeking the outcome of one specific action or the effect of one specific property. For example, the competency question "Do farms using imported potting mix show lower contaminations compared to those relying on native soils?" can be described with the contingency table "Farms with imported spotting mix / Farms with native soils" and "Farms with higher than average contaminants / Farms with lower than average contaminants." The odds ratio is then simply the measure of "Farms that have above average contaminants and imported soil mix" divided by "Farms that have above average contaminants and native soil." The second reason is that Fisher's exact test works well with smaller datasets, which is essential when there are only 23 sites in the dataset, which prevents the use of other statistical tests like the chi-squared test.

The queries derived from the competency questions use the relations in the knowledge graph described in further detail in the above section to form the Cypher queries. In the previous example, the basis of the query are the ***HAS_SOIL_CONTAMINATION*** and ***USES_GROWING_MEDIA*** relationships. In cases like this, where the query is relatively simple, most of the work of the query is in normalizing values, such as letter capitalization or replacing hyphens with underscores, to ensure that no relationship or factor is excluded by mistake, and in applying the Fisher's exact test on top of the data. In some cases, the average of a given statistic like biodiversity or sustainability is required to answer the query and is also calculated. Plans for future expansion of this application will be discussed in the Conclusion and Future Works section.

The application currently answers 35 queries from the 324 competency questions. These queries are those that were possible to answer with the data currently available; namely, those questions that did not ask about yield or didn't have a temporal dimension, as those were the largest two gaps in the data as previously mentioned. Additionally, some of the queries better fit being split into two different queries, namely those asking about the influence of operator education and institutional support on a given factor. While Fisher's test can be extended beyond the 2x2 table used for the majority of the questions, increasing the size of the table would spread the lack of data that already plagues my project even further. As such, it was deemed easier to simply split the question in two.

Result

1 row

Q11_R5_C4

questionID	Q11_R5_C4
primaryConcept	Previous Land Use
linkedFactor	Agricultural
question	Do farms established on previously agricultural land face risks of soil degradation, nutrient runoff, or pesticide residues?
factorRelationship	HAS_PREVIOUS_LAND_USE
factorValues	Agricultural
outcomeRelationship	HAS_POLLUTION_SOURCE
outcomeValues	Pesticide Drift
oddsRatio	1.714
oddsRatioHaldaneAnscombe	1.667
pValue	1

Fisher's exact test 2x2 grid

Factor group	Pollution Source: Pesticide Drift	No Confirmed Pollution Source: Pesticide Drift
Agricultural	2	7
All other farms	2	12

Figure 4.9: *An example of a question being answered*

Chapter 5

Conclusion and Future Works

In my project I built an ontology, knowledge graph, and application for the purpose of assisting in urban agriculture using LLM-enhanced methods. The ontology successfully answered the competency questions data allowed and serves as a basis for a future generalized expansion. The ontology also served as the schema for a knowledge graph encapsulated in a Neo4j graph database, enabling further reasoning and inference. The application serves as a front-end for users to use a narrow set of queries to gain a further understanding of urban food systems. The intended end users of this application would be a combination of urban agricultural operators, policy and decision makers, and the general public. Operators may use the application to better guide their practices in their own operations, and policymakers may use the applications to better craft urban agricultural policy to support those operators or their public needs.

No project is without flaws, neither in the process of completing the project nor in the final project. The flaws in the final project serve as potential avenues for future work which have been discussed above. The flaws in the process cannot be undone, but they can serve as important lessons for future work. The lesson that is the most beneficial to learn would be to ensure that you understand the end goal before setting out. While I knew that the end goal of my project was an ontology specific to answering the UFS competency questions, understanding exactly what needed to be changed in the prototype ontology took more cycles of LLM and human revision than would have been necessary had a clearer understanding of the goal been had at the beginning. As an expansion of that point, further and earlier discussions with domain experts would have assisted in the lessening of the time spent revision, as well as further understanding the end goal by refining what competency questions could be answered from the data available.

In the future, there are two avenues that could be explored to expand my project. The first involves the collection of more data. While the breadth of the data is undoubtedly impressive, there is only so much that can be drawn from one year's worth of data from 23 sites. Gathering data over multiple years would be the easiest way to expand the dataset, as the operators of all the sites are already familiar with the UFS project and are likely willing to answer any further surveys given that they have already answered one. This wouldn't be able to solve all of the problems that were present in the currently existing dataset, as there would still only be a limited number of sites for any category, like those on formerly industrial land. To rectify that, the UFS group would need to gather information from more urban ag sites, which may require expanding the scope of the UFS project beyond the state of Kansas and the Kansas City Metro Area, and beyond Kansas State University. The survey could be improved by including additional questions. Some of the competency questions could be better answered through the collection of further data, such as asking the operators for a self-reporting of how their yields might compare to a similar site based on their own experience and judgment. While this wouldn't be an extremely rigorous method of measuring yield, it is one of the few ways that further data on the topic could be gathered without expanding the sites the survey measures.

Part of the above questions and gaps in the research have been addressed by other members of the UFS project, or may be in the future. The second avenue for future work is what I would do in the future to expand the application portion of my project. The first route of application expansion is allowing the system to handle user-generated questions, rather than the strict competency questions it currently answers. Depending on how much flexibility is desired of the system, a tightly-controlled system of users choosing from dropdown boxes would be comparatively simple to implement. Such a system would require further discussion with the UFS team for what questions, beyond the competency questions, urban agriculture operators may want answered, and would likely require periodic updates if new questions present themselves. Allowing for a more freeform system of user inputs presents two obvious choices for implementation. An algorithmic method of natural language processing would be another way of allowing further flexibility. Relying on an LLM to automatically convert the user's question into Cypher queries may be more effective than trying to do so purely algorithmically. The best solution might involve a mix of strategies, which is another area of exploration. In addition to adjusting how user inputs are handled, it may be necessary to modify the specific statistical method being used. Although Fisher's test can address the competency questions the available data supports, it has not been evaluated on those competency questions that were excluded due to insufficient data.

A second avenue for extending the application would involve establishing baseline values for key statistics, especially yield. Since most of the competency questions focus on yield, identifying a baseline would substantially broaden the range of questions my project could address. The ideal way to achieve this would be to incorporate additional urban agriculture and food systems datasets, either those overlooked in the initial data collection or those that become available after the project ends. Because I cannot estimate the likelihood of this outcome, it may be necessary to draw on a broader agricultural dataset to answer yield-related questions. However, this too has some shortcomings that could potentially affect the results and conclusions. Because of the differences between conventional and urban agriculture and agricultural practices. Certain statistical approaches, such as resampling, might offer some help, but with an already limited dataset, these techniques would be risky. Beyond yield, other variables, such as soil health indicators (NPK levels, soil organic matter, and related measures), also lack robust baseline statistics, making the corresponding competency questions difficult or impossible to answer.

Beyond the application, the next area of future work would be in generalizing and broadening the scope of the ontology. The ontology created for my project is narrow in scope and specific to the needs of answering the competency questions posed by the domain experts in the UFS project. While its very existence provides a foundation for future work in the field, adding more generalized classes to the ontology will greatly assist in that effort. Given the two largely unused UFS datasets for my project, the operator information and community health, and some data points from the operations dataset that were not used, there is ample room for expanding the ontology to include them in addition to any other classes related to urban agriculture. Most of those datasets will likely expand the *transdisciplinary components* subtree of the ontology with things such as the influence of the farms on local food access, as seen through the operations data, and the community outreach programs a farm does and the demographic data of the farms' operators, as seen through the operator information dataset. The more generalized ontology would make my project more than simply a foundation for others to build upon.

Given the above changes to the ontology, there would be necessary follow on changes to the knowledge graph. If multiple years data were included, that would necessitate a 'year' category of metadata for the relevant statistics, for example. If other factors were included in the ontology that may affect biodiversity or sustainability, such as the data on potential agricultural chemical runoff or further soil health data , then they would also need to be included in the calculations for their averages. Other data points may have similar effects on averages that have not been tabulated yet.

These future improvements to this application may make this application more effective and better serve the operators, policymakers, and the public at large. This application aims to help the domain experts and users better through its continued development and become one small part in improving the broader urban food systems ecosystem for a better society.

Bibliography

[1] Hande Küçük McGinty. *Knowledge Acquisition and Representation Methodology (KNARM) and Its Applications.* Ph.D., University of Miami, United States – Florida, 2018. URL https://www.proquest.com/docview/2050044221/abstract/63AD8976B0734662PQ/1.

[2] Madara Dobele and Andra Zvirbule. The Concept of Urban Agriculture – Historical Development and Tendencies. *Rural Sustainability Research*, 43(338):20–26, August 2020. ISSN 2256-0939. doi: 10.2478/plua-2020-0003. URL https://reference-global.com/article/10.2478/plua-2020-0003.

[3] Carina Júlia Pensa Corrêa, Kelly Cristina Tonello, Ernest Nnadi, and Alexandra Guidelli Rosa. SEEDING THE CITY: HISTORY AND CURRENT AFFAIRS OF URBAN AGRICULTURE. *Ambiente & Sociedade*, 23:e00751, 2020. ISSN 1414-753X, 1809-4422. doi: https://doi.org/10.1590/1809-4422asoc20180075r1vu2020L1AO. URL https://www.scielo.br/j/asoc/a/D9jj4kzfLtzqKwWqbkxVhnc/?lang=en.

[4] Erica Dorr, Benjamin Goldstein, Arpad Horvath, Christine Aubry, and Benoit Gabrielle. Environmental impacts and resource use of urban agriculture: a systematic review and meta-analysis. *Environmental Research Letters*, 16(9):093002, August 2021. ISSN 1748-9326. doi: 10.1088/1748-9326/ac1a39. URL https://doi.org/10.1088/1748-9326/ac1a39.

[5] Florian Thomas Payen, Daniel L. Evans, Natalia Falagán, Charlotte A. Hardman, Sofia Kourmpetli, Lingxuan Liu, Rachel Marshall, Bethan R. Mead, and Jessica A. C. Davies. How Much Food Can We Grow in Urban Areas? Food Production and Crop Yields of Urban Agriculture: A Meta-Analysis. *Earth's Future*, 10(8):e2022EF002748, 2022. ISSN 2328-4277. doi: 10.1029/2022EF002748.

URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2022EF002748. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022EF002748.

[6] Senate Agriculture Committee. Farm Bill 101: Growing Urban Agriculture | The United States Senate Committee On Agriculture, Nutrition & Forestry, August 2023. URL https://www.agriculture.senate.gov/newsroom/dem/press/release/farm-bill-101-growing-urban-agriculture.

[7] Rositsa T. Ilieva, Nevin Cohen, Maggie Israel, Kathrin Specht, Runrid Fox-Kämper, Agnès Fargue-Lelièvre, Lidia Poniży, Victoria Schoen, Silvio Caputo, Caitlin K. Kirby, Benjamin Goldstein, Joshua P. Newell, and Chris Blythe. The Socio-Cultural Benefits of Urban Agriculture: A Review of the Literature. *Land*, 11(5):622, May 2022. ISSN 2073-445X. doi: 10.3390/land11050622. URL https://www.mdpi.com/2073-445X/11/5/622.

[8] Mahbubur R. Meenar and Brandon M. Hoover. Community Food Security via Urban Agriculture: Understanding People, Place, Economy, and Accessibility from a Food Justice Perspective. *Journal of Agriculture, Food Systems, and Community Development*, 2012. URL https://ageconsearch.umn.edu/record/359510.

[9] Sarada Krishnan, Dilip Nandwani, George Smith, and Vanaja Kankarta. Sustainable Urban Agriculture: A Growing Solution to Urban Food Deserts. In Dilip Nandwani, editor, *Organic Farming for Sustainable Agriculture*, volume 9, pages 325–340. Springer International Publishing, Cham, 2016. ISBN 978-3-319-26801-9 978-3-319-26803-3. doi: 10.1007/978-3-319-26803-3_15. URL http://link.springer.com/10.1007/978-3-319-26803-3_15. Series Title: Sustainable Development and Biodiversity.

[10] Guo-yu Qiu, Hong-yong Li, Qing-tao Zhang, Wan Chen, Xiao-jian Liang, and Xiang-ze Li. Effects of Evapotranspiration on Mitigation of Urban Temperature by Vegetation and Urban Agriculture. *Journal of Integrative Agriculture*, 12(8):1307–1315,

August 2013. ISSN 2095-3119. doi: 10.1016/S2095-3119(13)60543-2. URL https://www.sciencedirect.com/science/article/pii/S2095311913605432.

[11] Brian K. Obach and Kathleen Tobin. Civic agriculture and community engagement. *Agriculture and Human Values*, 31(2):307–322, June 2014. ISSN 1572-8366. doi: 10.1007/s10460-013-9477-z. URL https://doi.org/10.1007/s10460-013-9477-z.

[12] Cynthia Zutter and Ashley Stoltz. Community gardens and urban agriculture: Healthy environment/healthy citizens. *International Journal of Mental Health Nursing*, 32(6):1452–1461, 2023. ISSN 1447-0349. doi: 10.1111/inm.13149. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13149. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/inm.13149.

[13] Kansas State University. Game-changing research initiation program. URL https://www.k-state.edu/research/faculty/funding/grip/.

[14] Thomas R Gruber. *Encyclopedia of Database Systems*, chapter Ontology. Springer-Verlag, 2009.

[15] Rajendran Deepa and Srinivasan Vigneshwari. An effective automated ontology construction based on the agriculture domain. *ETRI Journal*, 44(4):573–587, 2022. ISSN 2233-7326. doi: 10.4218/etrij.2020-0439. URL https://onlinelibrary.wiley.com/doi/abs/10.4218/etrij.2020-0439. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.4218/etrij.2020-0439.

[16] Krithikha Sanju Saravanan and Velammal Bhagavathiappan. Innovative agricultural ontology construction using NLP methodologies and graph neural network. *Engineering Science and Technology, an International Journal*, 52:101675, April 2024. ISSN 2215-0986. doi: 10.1016/j.jestch.2024.101675. URL https://www.sciencedirect.com/science/article/pii/S2215098624000612.

[17] Neha Kaushik and Niladri Chatterjee. Automatic relationship extraction from agricultural text for ontology construction. *Information Processing in Agriculture*, 5

(1):60–73, March 2018. ISSN 2214-3173. doi: 10.1016/j.inpa.2017.11.003. URL https://www.sciencedirect.com/science/article/pii/S2214317317300227.

[18] Jiayi Li, Ziyuan Wang, Daniel Garijo, and María Poveda-Villalón. MASEO: A Multi-Agent System for Explainable Ontology Generation. 2022.

[19] Abid Talukder, Maruf Ahmed Mridul, and Oshani Seneviratne. Towards Automated Ontology Generation from Unstructured Text: A Multi-Agent LLM Approach, April 2026. URL http://arxiv.org/abs/2604.23090. arXiv:2604.23090 [cs.AI].

[20] Nadeen Fathallah, Arunav Das, Stefano De Giorgis, Andrea Poltronieri, Peter Haase, and Liubov Kovriguina. NeOn-GPT: A Large Language Model-Powered Pipeline for Ontology Learning. In Albert Meroño Peñuela, Oscar Corcho, Paul Groth, Elena Simperl, Valentina Tamma, Andrea Giovanni Nuzzolese, Maria Poveda-Villalón, Marta Sabou, Valentina Presutti, Irene Celino, Artem Revenko, Joe Raad, Bruno Sartini, and Pasquale Lisena, editors, *The Semantic Web: ESWC 2024 Satellite Events*, volume 15344, pages 36–50. Springer Nature Switzerland, Cham, 2025. ISBN 978-3-031-78951-9 978-3-031-78952-6. doi: 10.1007/978-3-031-78952-6_4. URL https://link.springer.com/10.1007/978-3-031-78952-6_4. Series Title: Lecture Notes in Computer Science.

[21] Yash Tiwari, Owais Ahmad Lone, and Mayukha Pal. OntoRAG: Enhancing Question-Answering through Automated Ontology Derivation from Unstructured Knowledge Bases, May 2025. URL http://arxiv.org/abs/2506.00664. arXiv:2506.00664 [cs.AI].

[22] Mohammad Javad Saeedizade and Eva Blomqvist. Navigating Ontology Development with Large Language Models. In Albert Meroño Peñuela, Anastasia Dimou, Raphaël Troncy, Olaf Hartig, Maribel Acosta, Mehwish Alam, Heiko Paulheim, and Pasquale Lisena, editors, *The Semantic Web*, volume 14664, pages 143–161. Springer Nature Switzerland, Cham, 2024. ISBN 978-3-031-60625-0 978-3-031-60626-7. doi: 10.1007/978-3-031-60626-7_8. URL https://link.springer.com/10.1007/978-3-031-60626-7_8. Series Title: Lecture Notes in Computer Science.

[23] Tanay Aggarwal, Angelo Salatino, Francesco Osborne, and Enrico Motta. Large language models for scholarly ontology generation: An extensive analysis in the engineering field. *Information Processing & Management*, 63(1):104262, January 2026. ISSN 0306-4573. doi: 10.1016/j.ipm.2025.104262. URL https://www.sciencedirect.com/science/article/pii/S0306457325002031.

[24] Yinglun Zhang, Aryan Singh Dalal, Caleb Martin, Srikar Reddy Gadusu, and Hande Küçük McGinty. Olive: Ontology learning with integrated vector embeddings. *Applied Ontology*, 20(1):36–53, 2025.

Appendix A

Competency Question Appendix

A.1 Answered Competency Questions

1. Previous Land Use — Industrial: Do farms with industrial land-use history show higher levels of soil or water contamination from toxic trace elements or hazardous waste?

2. Previous Land Use — Residential: Do farms located on land with pre-1978 residential use show elevated soil lead concentrations?

3. Previous Land Use — Residential: Is the adoption of remediation measures at residential-use sites dependent on post-grad operator education?

4. Previous Land Use — Residential: Is the adoption of remediation measures at residential-use sites dependent on institutional support?

5. Previous Land Use — Commercial: Do farms on former commercial land show higher contamination risks due to chemical storage or waste disposal?

6. Previous Land Use — Commercial: Is institutional support essential for enabling remediation of commercial-use sites?

7. Previous Land Use — Agricultural: Do farms established on previously agricultural land face risks of soil degradation, nutrient runoff, or pesticide residues?

8. Previous Land Use — Agricultural: Does prior agricultural land-use increase exposure risks from agrochemical residues in nearby water and soil?

9. Proximity to Pollution Sources — within 1 mile: Do farms located within 1 mile of known pollution sources show higher levels of soil, water, or air contamination?

10. Proximity to Pollution Sources — within 1 mile: Does biodiversity near farms within 1 mile of pollution sources show greater disruption?

11. Proximity to Pollution Sources — 1-5 miles: Do farms located 1–5 miles from pollution sources experience moderate levels of environmental contamination due to airborne drift or water runoff?

12. Soil Testing Frequency & Parameters — pH: Do farms that monitor pH more frequently adopt more effective remediation or soil amendment practices?

13. Soil Testing Frequency & Parameters — Organic Matter: Does frequent testing for organic matter correlate with greater adoption of composting or organic amendments?

14. Contaminant Remediation Methods — Contaminant Remediation Methods: Are remediation practices (e.g., raised beds, soil replacement, biochar) associated with lower levels of lead, arsenic, or other contaminants in soil tests?

15. Contaminant Remediation Methods — Contaminant Remediation Methods: Are farms with institutional support more likely to adopt effective remediation strategies?

16. Growing media — Imported Soil / Potting Mix: Do farms using imported potting mix show lower contamination compared to those relying on native soils?

17. Growing media — Native Soil: Are farms using native soil more likely to face contamination risks?

18. Growing Media — Water (Hydroponics / Aquaponics): Do hydroponic/aquaponic systems eliminate soil-related contaminant risks entirely?

19. Growing Methods — Open Field: Do open-field systems have higher pesticide/herbicide usage/drift and associated contamination risks?

20. Growing Methods — Raised Bed (framed/open): Do raised beds effectively reduce contamination risk on polluted sites by using imported soil/potting mix?

21. Growing Methods — Soilless Cultivation (Hydroponics, Aquaponics): Is adoption of soilless cultivation strongly dependent on institutional support?

22. Growing Methods — Soilless Cultivation (Hydroponics, Aquaponics): Is adoption of soilless cultivation strongly dependent on post-grad operator education?

23. Growing Methods — Container Gardening (pots, grow bags): Do container systems reduce contamination risk compared to in-ground methods?

24. Growing Methods — In-ground Gardening: Are in-ground gardens more likely to be contaminated due to reliance on natural soils?

25. Growing Methods — Sustainable Farming: Are biodiversity and pollination services significantly higher in sustainable systems?

26. Growing Methods — Sustainable Farming: Is adoption of cultural/biological pest management reducing reliance on synthetic pesticides?

27. Growing Methods — Greenhouse Production: Does greenhouse production reduce chemical drift and pollution risk compared to open-field systems?

28. Growing Methods — Organic Farming: Are organic systems improving biodiversity and pollination services?

29. Cropping System — Mixed Cropping: Does mixed cropping increase biodiversity and reduce pest/disease risk compared to monocropping?

30. Perennial Crop Coverage — >75%: Do farms with over 75% perennial crops maximize soil conservation,carbon sequestration, biodiversity, and provide nesting habitats for animals?

31. Weeds, Pest and disease control inputs — Synthetic Pesticides, herbicides, Fungicides,: Does it reduce biodiversity and pollination insects within the farm?

32. Weeds, Pest and disease control inputs — Cultural Control (crop rotation, companion planting): Do cultural methods reduce chemical inputs?

33. Weeds, Pest and disease control inputs — Cultural Control (crop rotation, companion planting): Do cultural methods promote biodiversity?

34. Pollination methods — Reduce pesticide use: Does reduced pesticide use promote pollinator populations and local biodiversity?

35. Institutional Support — Significant support (funding, training, resources): Would significant institutional support encourage adoption of sustainable practices (solar panel, soil test, water test, renewable energy sources)

A.2 Complete Competency Question List

I Previous Land Use

(a) Industrial

i. Do farms with industrial land-use history show higher levels of soil or water contamination from toxic trace elements or hazardous waste?

ii. Does institutional support improve remediation outcomes and land recovery in industrial-use sites?

iii. Are farms with industrial land-use history associated with higher remediation costs and reduced crop yields?

iv. Do residents near industrial-use farm sites face higher risks of exposure to soil contaminants (ingestion, dust inhalation, skin contact)?

(b) Residential

i. Do farms located on land with pre-1978 residential use show elevated soil lead concentrations?

ii. Are crops from residential-use sites more likely to accumulate contaminants from legacy paint dust, runoff, or household waste?

iii. Does soil testing at residential sites reduce exposure risks and guide safer crop selection (to apply remediation methods)?

iv. Is the adoption of remediation measures at residential-use sites dependent on operator education and institutional support?

(c) Commercial

i. Do farms on former commercial land show higher contamination risks due to chemical storage or waste disposal?

ii. Are yields from commercial-use sites less stable due to contamination issues and remediation needs?

iii. Is institutional support essential for enabling remediation of commercial-use sites?

iv. Do commercial-use farms pose higher community health risks through soil or water pollution?

(d) Agricultural

i. Do farms established on previously agricultural land face risks of soil degradation, nutrient runoff, or pesticide residues?

ii. Does sustainable nutrient cycling in agricultural-use sites reduce fertilizer needs and increase productivity?

iii. Does prior agricultural land-use increase exposure risks from agrochemical residues in nearby water and soil?

iv. Are agricultural-use sites more resilient if soil conservation and crop diversity practices are adopted?

(e) Vacant

i. Do farms on previously vacant land show higher risks of soil compaction, low organic matter, or illegal dumping?

ii. Are vacant-use sites less productive until soil remediation?

iii. Do vacant-use sites present hidden community health risks due to unknown prior contaminant exposure?

iv. Can urban greening or restoration of vacant sites enhance long-term food system resilience?

II Proximity to Pollution Sources

(a) within 1 mile

i. Do farms located within 1 mile of known pollution sources show higher levels of soil, water, or air contamination?

ii. Are farms within 1 mile of pollution sources more likely to experience reduced productivity due to crop damage or remediation barriers?

iii. Does biodiversity near farms within 1 mile of pollution sources show greater disruption?

iv. Do farmers and nearby residents face higher risks of chronic illnesses or poisoning from ingestion, inhalation, or skin contact when farms are located within 1 mile of pollution sources?

(b) 1-5 miles

i. Do farms located 1–5 miles from pollution sources experience moderate levels of environmental contamination due to airborne drift or water runoff?

ii. Are crop yields moderately affected in farms 1–5 miles from pollution sources, requiring higher input use or water filtration?

iii. Do communities near farms 1–5 miles from pollution sources face indirect exposure risks through contaminated irrigation water, dust, or produce?

iv. Are farms within 1–5 miles able to maintain resilience through monitoring, adaptive land-use strategies, or protective structures?

(c) 5-10 miles

i. Do farms located 5–10 miles from pollution sources show lower immediate risks but higher long-term vulnerability to watershed pollution or groundwater contamination?

ii. Is agricultural productivity generally stable for farms 5–10 miles from pollution sources, unless pollutants spread through irrigation water or supply chains?

iii. Are health risks lower for communities near farms 5–10 miles from pollution sources, though cumulative effects (e.g., bioaccumulation) may still occur?

iv. Do farms 5–10 miles from pollution sources demonstrate stronger resilience compared to closer zones, but remain vulnerable to systemic pollution shocks?

III Soil Testing Frequency & Parameters

(a) pH

i. Are soil pH imbalances associated with reduced nutrient uptake and higher fertilizer requirements?

ii. Do farms that monitor pH more frequently adopt more effective remediation or soil amendment practices?

iii. Do farms that regularly test soil pH show better crop yields compared to those that do not?

(b) Organic Matter

i. Are higher levels of organic matter associated with improved soil fertility (NPK) and microbial diversity?

ii. Does frequent testing for organic matter correlate with greater adoption of composting or organic amendments?

iii. Do farms that test for soil organic matter (SOM) show higher resilience to drought and extreme weather?

(c) NPK , Micro nutrients

i. Are excessive NPK applications (without soil testing) associated with greater risks of pollution (toxic elements)?

ii. Do farms that test for NPK and micronutrients achieve higher productivity and reduced input costs?

iii. Is nutrient deficiency (detected via soil testing) linked to lower yields or poorer crop quality in urban farms?

iv. Do farms that track micronutrients (e.g., Zn, Fe, Mn) show improved food safety and nutritional value in harvested crops?

(d) Texture

i. Is soil texture testing linked to improved water management and erosion control practices?

ii. Do farms with clay soils use lower amendment inputs due to higher nutrient and water retention?

iii. Do farms with sandy soils require higher input use due to poor nutrient retention and water conservation capacity?

IV Application of Soil Test Recommendations

(a) Unspecified

i. Do farmers who regularly apply soil test recommendations show improved soil fertility (balanced NPK, pH, organic matter) compared to those who don't?

ii. Is there a relationship between following soil test recommendations and lower contamination levels?

iii. Does education level influence whether farmers apply soil test recommendations?

iv. Do farmers with institutional support more frequently apply soil test recommendations?

V Contaminant Remediation Methods

(a) Unspecified

i. Are remediation practices (e.g., raised beds, soil replacement, biochar) associated with lower levels of lead, arsenic, or other contaminants in soil tests?

ii. Do remediation methods vary depending on the previous land use (industrial vs residential vs vacant)?

iii. Is there a relationship between remediation methods and application of soil test recommendations?

iv. Are farms with institutional support more likely to adopt effective remediation strategies?

v. Does the choice of remediation method correlate with farmers' reported pollution sources?

VI Growing media

(a) Imported Soil / Potting Mix

i. Do farms using imported potting mix show lower contamination compared to those relying on native soils?

ii. Are hidden contaminants present in uncertified potting mixes?

(b) Native Soil

i. Are farms using native soil more likely to face contamination risks?

ii. Does the use of native soil enhance biodiversity when managed sustainably?

iii. How does the fertility of native soils compare to productivity in alternative growing media?

(c) Perlite

i. Are farms using perlite achieving higher yields due to improved aeration in soilless systems?

(d) Rockwool

i. Do farms using rockwool achieve consistently higher yield

ii. Are there significant waste management or sustainability challenges linked to rockwool adoption?

iii. How does rockwool improve water/moisture management compared to other substrates?

(e) Clay Pebbles

i. Is the higher upfront material cost justified by long-term reuse?

VII 205

(a) Compost

i. Do farms using compost demonstrate better soil organic matter content and nutrient cycling?

ii. Does compost application reduce reliance on synthetic fertilizers and improve water retention?

iii. Are contamination risks (pathogens, toxic elements) associated with poorly processed compost?

VIII 138

(a) Coco Coir

i. Does coco coir provide water-holding benefits to peat without the unsustainable extraction footprint?

ii. Are farms using coco coir less prone to nutrient imbalances?

iii. Is salt content in coco coir a risk factor for crop toxicity?

(b) Sand

i. Do sand-based growing systems show higher nutrient leaching (Low NPK)?

ii. Is crop yield significantly lower in sand unless combined with compost or fertilizers?

iii. How do contamination risks compare between sand and native soils?

(c) Water (Hydroponics / Aquaponics)

i. Do hydroponic/aquaponic systems eliminate soil-related contaminant risks entirely?

ii. Are hydroponic/aquaponic farms more likely to meet market demand for high-quality, contaminant-free produce compared to soil-based farms?

iii. How does water quality tests (pH, EC, nutrient balance) influence productivity?

iv. Does dependence on stable energy supply limit resilience in low-resource urban contexts?

v. How does water recirculation efficiency in hydroponics/aquaponics compare to water conservation in soil-based systems?

vi. Are aquaponic systems providing additional food security benefits (fish + plants)?

IX Growing Methods

(a) Open Field

i. Are open-field farms more vulnerable to soil erosion and nutrient runoff compared to protected structures?

ii. Do open-field systems have higher pesticide/herbicide usage/drift and associated contamination risks?

iii. Does biodiversity support (e.g., wildlife habitat) improve resilience/ damage crops in open-field systems?

iv. How does variability in climate affect yield stability in open-field compared to greenhouse/vertical systems?

(b) Protected Structures (High Tunnel/ Hoop houses/ Green houses)

i. Do protected structures reduce pesticide and herbicide use compared to open-field systems?

ii. Are protected systems more energy intensive, and how does that affect resilience under rising energy costs?

iii. Do protected structures consistently deliver higher yields and quality compared to open-field farms?

iv. How much biodiversity trade-off occurs in protected systems compared to open-field methods?

(c) Rooftop gardening

i. Do rooftop gardens contribute to urban ecosystem services (cooling, stormwater management, air quality?

ii. Is rooftop farming vulnerable to pollution from dust and trace elements?

iii. Are rooftop systems more economically viable for niche high-value crops compared to open-field?

(d) Raised Bed (framed/open)

i. Do raised beds effectively reduce contamination risk on polluted sites by using imported soil/potting mix?

(e) Vertical Gardening

i. How does the higher input demand (materials, nutrients) affect sustainability compared to soil-based systems?

ii. Do vertical gardening systems provide higher returns per square foot than other urban growing methods?

iii. Does vertical gardening contribute to improved dietary diversity and urban nutrition access?

iv. Are vertical gardens more resilient to climate variability

(f) Soilless Cultivation (Hydroponics, Aquaponics)

i. Do hydroponic/aquaponic systems eliminate soil-related contaminant risks entirely?

ii. How does water/nutrient quality management (pH, EC) affect yields and food safety?

iii. Is adoption of soilless cultivation strongly dependent on operator education and institutional support?

iv. Are aquaponic systems providing added nutritional security by integrating fish?

(g) Container Gardening (pots, grow bags)

i. Do container systems reduce contamination risk compared to in-ground methods?

ii. How significant are waste management concerns (plastics) in container gardening?

(h) In-ground Gardening

i. Are in-ground gardens more likely to be contaminated due to reliance on natural soils?

ii. How do soil conservation methods (cover crops, mulching) influence productivity in in-ground gardening?

iii. Do in-ground systems provide higher yields at lower cost compared to raised beds and containers?

(i) Sustainable Farming

i. Do farms practicing mixed/intercropping, crop rotation, and relay cropping show more stable yields over time?

ii. Are biodiversity and pollination services significantly higher in sustainable systems?

iii. Do organic amendments improve long-term soil fertility and reduce input dependence?

iv. Is adoption of cultural/biological pest management reducing reliance on synthetic pesticides?

v. How do soil/water conservation and energy-saving practices improve resilience in sustainable systems?

(j) Greenhouse Production

i. Do greenhouse systems consistently deliver higher yields and profitability compared to other protected methods?

ii. Are greenhouse farms more vulnerable to energy price fluctuations due to high energy demand?

iii. Does greenhouse production reduce chemical drift and pollution risk compared to open-field systems?

iv. Are food safety outcomes (reduced pesticide use, controlled environment) significantly better in greenhouses?

(k) Organic Farming

i. Do organic farms consistently achieve premium market prices despite lower yields?

ii. Are organic systems improving biodiversity and pollination services?

iii. Do cultural/mechanical/biological pest management strategies in organic farms reduce agrochemical exposure without reducing yields too drastically?

X Cropping System

(a) Monocropping

i. Do monocropping farms show higher soil nutrient depletion and erosion risks compared to diversified systems?

ii. Is pest/disease incidence significantly higher in monocropping compared to intercropping or rotations?

iii. Are monocropping systems more dependent on chemical fertilizers and pesticides than other cropping systems?

iv. Are weed and pest pressures higher in monocropping systems?

(b) Mixed Cropping

i. Does mixed cropping increase biodiversity and reduce pest/disease risk compared to monocropping?

ii. Are yields more stable in mixed cropping systems despite potentially lower maximum productivity?

iii. Does mixed cropping improve dietary diversity and nutritional outcomes at the household level?

iv. How does labor intensity in mixed cropping compare to economic stability gained from crop diversification?

(c) Intercropping

i. Do intercropping systems improve nutrient cycling and soil fertility compared to monocropping?

ii. Are weed and pest pressures lower in intercropped systems?

iii. Is land-use efficiency significantly higher in intercropping compared to monocropping or rotation?

iv. Do intercropping farms report lower pesticide use and associated health benefits?

(d) Crop Rotation

i. Does crop rotation prevent long-term soil nutrient depletion more effectively than monocropping?

ii. Are pest and disease cycles reduced more effectively under crop rotation compared to continuous cropping?

iii. Do farms practicing crop rotation achieve more stable yields and lower input costs compared to non-rotating farms?

iv. Is crop rotation associated with better soil health indicators (organic matter, microbial activity) and higher resilience?

(e) Relay Cropping

i. Do relay cropping systems reduce bare-soil periods and improve soil conservation compared to monocropping?

ii. Are overall seasonal yields higher under relay cropping compared to sole cropping?

XI Perennial Crop Coverage

(a) ¡10%

i. Do farms with less than 10% perennial crop coverage show minimal soil conservation and biodiversity benefits?

ii. Does low perennial coverage correlate with lower overall productivity and limited income diversification?

iii. Is food supply less diverse in systems dominated by annual crops?

iv. Are farms with ¡10% perennial coverage more vulnerable to climate shocks, pests, or market fluctuations?

(b) ¿75%

i. Do farms with over 75% perennial crops maximize soil conservation,carbon sequestration, biodiversity, and provide nesting habitats for animals?

ii. Is long-term productivity and stable income higher in farms dominated by perennial crops?

iii. Does perennial dominance ensure a consistently diverse and nutritious food supply?

XII Organic Material Inputs

(a) Compost

i. Do farms using compost achieve higher biodiversity, improved soil health, and enhanced ecosystem services?

ii. Does compost use reduce reliance on chemical fertilizers and support perennial crop yields?

iii. Is resilience higher on farms using compost due to buffering against climate and pest shocks?

iv. Does compost contribute to a consistent and nutritious food supply?

(b) Manure

i. Does manure application maximize soil biodiversity and prduction capacity?

ii. Does manure application reduce reliance on chemical fertilizers and cost?

iii. Does it contribute to long-term productivity and stable income?

(c) Biofertilizer

i. Do biofertilizers enhance soil organic matter, structure, and water retention?

ii. Does biofertilizer reduce the need for synthetic inputs and improve crop yield?

(d) Organic Fertilizers

i. Do organic fertilizers increase nutrient content in soil (NPK)?

(e) Digestate / Slurry

i. Does it reduce reliance on chemical fertilizers and improve yields?

(f) Crop Residues / Cover Crops

i. Do crop residues and cover crops add nutrients reducing synthetic chemicals?

ii. Do they support biodiversity and moderate to high productivity?

iii. Does it improve soil moisture consevation and erosion control?

(g) Lime / Gypsum

i. Can they boost productivity and reduce fertilizer costs?

(h) Biochar

i. Does biochar improve soil nutrient availability, reduce erosion, and enhance soil organic matter/moisture?

ii. Does it reduce other input costs (water/ amendments)?

XIII Synthetic inputs

(a) NPK , Micro nutrients

i. Do they increase yields and production cost?

ii. Can overuse lead to nutrient runoff, soil salinity, or environmental contamination?

XIV Weeds, Pest and disease control inputs

(a) Synthetic Pesticides, herbicides, Fungicides,

i. Can these inputs improve crop productivity if used properly?

ii. Does misuse increase soil toxic trace elements concentration in soil test?

iii. Does the use of these products pose any potential health risks to human?

iv. Does it reduce biodiversity and pollination insects within the farm?

(b) Biological Control (natural predators, parasitoids, nematodes, microbes)

i. Do biological controls reduce pest populations and improve yield without chemical inputs?

ii. Does use reduce pesticide exposure for communities?

iii. Does it reduce synthetic pesticide input cost?

(c) Mechanical Control (manual removal, traps)

i. Do mechanical controls reduce pest pressure while protecting yield and crop quality?

ii. Does use reduce pesticide exposure for communities?

iii. Does it reduce synthetic pesticide input cost?

(d) Cultural Control (crop rotation, companion planting)

i. Do cultural methods promote biodiversity, reduce chemical inputs?

ii. Does use reduce pesticide exposure for communities?

iii. Does it reduce synthetic pesticide input cost?

(e) Integrated Management Strategies (IMS)

i. Do IMS approaches minimize environmental impact (potential toxins) ?

ii. Does it improve yield and profit?

iii. Does it reduce synthetic input cost?

XV Pollination methods

(a) Provide floral resources (plant pollinator-friendly plants, allow wildflowers, avoid mowing)

i. Do they enhance food production by reducing pesticide reliance?

(b) Install bee hives/boxes

i. Do bee hives optimize yields while reducing input costs

ii. Does it potentially provide additional income (e.g., honey)?

(c) Promote nesting habitat (nest boxes, dead wood, undisturbed soil)

i. Does creating nesting habitats support pollinator biodiversity and improve yields

(d) Reduce pesticide use

i. Does reduced pesticide use promote pollinator populations and local biodiversity?

ii. Does it improve production capacity and profits?

XVI Animal Categories

(a) Integrated Animals (Livestock: goats, poultry, pigs, etc.)

i. Can livestock integration improve overall productivity and income?

ii. Do they reduce fertilizer costs while providing manure?

iii. Are there risks of overgrazing, manure mismanagement, or conflicts with crops?

iv. Does livestock integration increase risk for human health if not properly managed

(b) Wildlife (deer, rabbits, groundhogs, etc.)

i. Can wildlife reduce input costs (fertilizers)?

ii. Can wildlife cause crop damage?

(c) Aquaculture (fish)

i. Can aquaculture reduce fertilizer needs and reduce cost while supporting income?

ii. Does aquaculture pose risks such as water pollution or disease

(d) Pollinators (bees, butterflies)

i. Does pollinators increase productivity and yield? /profit

ii. Do pollinators contribute to biodiversity?

iii. Can uncontrolled pollinator activity damage crops or structures?

(e) Natural Pest Controllers (birds, bats, predatory insects)

i. Do natural predators support biodiversity and reduce the need for synthetic pest control inputs?

XVII Land preparation

(a) Conventional Tillage

i. Does it increase soil erosion and soil carbon loss if mismanaged.

ii. Does it increase availability of nutrients reducing input needs

iii. Does conventional tillage pose any direct health risks to workers or consumers?(if contaminated)

(b) No-Till Farming

i. Does no-till farming reduce soil erosion and improve yields

ii. Does it decrease land preparation cost

iii. Does it increase input needs (fertilizers)

(c) Minimum Tillage

i. Does minimum tillage reduce soil erosion and improve yields

ii. Does it decrease land preparation cost

iii. Does it increase input needs (fertilizers)

(d) Strip Tillage

i. Does it reduce soil erosion, input needs

ii. Does it reduce fuel and labour cost

XVIII Slope

(a) Flat (0–1% slope)

i. Does flat land have minimal erosion risk and easy water management?

(b) Very steep slope (¿10% slope)

i. Does a very steep slope have a high erosion risk, challenging water management (application method), and reduced productivity?

XIX Soil erosion control methods

(a) Terracing / Contour Farming

i. Does [soil conservation method] reduce soil erosion and prevent nutrient loss?

(b) Cover Cropping

i. Does cover cropping improve soil fertility?(soil test: NPK)

ii. Does cover cropping reduce fertilizer cost?

iii. Does cover cropping reduce irrigation need?

iv. Does cover cropping increase yield and profit?

(c) Mulching

i. Does mulching improve crop yield?

ii. Does mulching reduce waste?

iii. Does mulching reduce irrigation need?

iv. Does mulching improve soil fertility (NPK)?

(d) Windbreaks

i. Does windbreaks reduce potential pollution?

ii. Does windbreak increase crop yield?

(e) Bioswales

i. Does reduce contamination risk?

XX Water source

(a) Municipal Water Supply

i. Does using municipal water supply increase costs?

ii. Does the use of municipal water supply provide reliable, good-quality water?

(b) Rainwater Harvesting

i. Does rainwater harvesting reduce runoff and erosion?

ii. Does rainwater harvesting reduce reliance on municipal water bill

(c) Groundwater (wells, aquifers)

i. Is water source reliable, safe (quality), and supportive of stable crop yields?

(d) Recycled / Treated Wastewater

i. Is this water source costly to access/ treat?

ii. Does it align with reducing wastewater discharge?

iii. Does it contaminated with toxic elements or microbes

(e) Surface Water (ponds, rivers, streams)

i. Does it align with water quality

ii. Does it reduce yield

iii. Does it increase energy cost

iv. Does it reduce water bill

XXI Water conservation methods

(a) Micro Irrigation Systems (e.g., drip irrigation)

i. Is this water source/method effective, safe(contaminant free), and supportive of stable yields?

(b) Rainwater Harvesting

i. Does rainwater harvesting reduce runoff and erosion?

ii. Does rainwater harvesting reduce reliance on municipal water bill

(c) Recycled or Greywater

i. Is this water source costly to access/ treat?

ii. Does it align with reducing wastewater discharge?

(d) Automated Irrigation System

i. Does it reduce water bill

ii. Does it affct for electricity bill

iii. Does it affect for yield and profit

(e) Drought-Tolerant or Native Plants

i. Does it reduce water bill

ii. Does it reduce soil erosion

iii. Does it affect for yield and profit

XXII Water quality issues

(a) Nitrate Levels

i. Would farms using nitrate-rich water sources show reduced crop yields or altered soil nutrient balance?

ii. Would high nitrate levels in irrigation water pose health risks to consumers or communities?

(b) Phosphate Levels

i. Do elevated phosphate levels in irrigation water affect soil fertility/ crop produce?

ii. Could high phosphate runoff increase eutrophication of water sources?

(c) Toxic Trace Elements (e.g., lead, arsenic)

i. Would farms irrigated with water containing trace metals show contaminations

ii. Do toxic trace elements in water pose persistent health risks

(d) Microbial Contaminants (coliform, E. coli)

i. Would irrigation with microbially contaminated water reduce crop safety or yields?

ii. Do microbial contaminants increase health risks

(e) pH Level

i. Would extreme pH in water affect nutrient availability, crop growth, or soil quality?

(f) Salinity

i. Would irrigation with saline water reduce plant growth, productivity, or soil quality?

XXIII Water application methods

(a) Drip Irrigation

i. Would farms using drip irrigation show higher water-use efficiency and increased crop yields

ii. Does drip irrigation reduce contamination risk

(b) Sprinkler Irrigation

i. Would farms using sprinkler irrigation cover large areas effectively but have moderate water efficiency?

ii. Does sprinkler irrigation increase the risk of pathogen spread via water splash and moderately affect resilience?

(c) Soaker Hoses

i. Would using soaker hoses conserve water and reduce surface runoff while maintaining soil moisture?

(d) Hand Watering

i. Would hand watering be effective for small land area while being labor-intensive?

(e) Flood Irrigation

i. Would flood irrigation lead to high water usage, erosion, and nutrient leaching?

ii. Does this method increase contamination risk and reduce overall system resilience?

(f) Automated Irrigation (timers or sensors)

i. Would automated irrigation optimize water use and improve efficiency while supporting consistent yields?

ii. Would automated irrigation incur higher installation and energy costs

XXIV Energy Sources

(a) Electricity from the Grid

i. Would farms using grid electricity have reliable energy for operations, and how does the energy mix affect carbon footprint?

ii. Does reliance on grid electricity involve variable costs and vulnerability to outages affecting resilience?

(b) Solar Panels

   i. Would farms using solar panels reduce electricity costs?

   ii. Does it affect for the yield

(c) Wind Power

   i. Would wind power supplement farm energy needs while providing low emissions?

(d) Diesel or Gasoline Generators

   i. Would farms using diesel/gas generators face high emissions/ pollution?

   ii. Do fuel costs and health risks from fumes affect resilience and sustainability?

(e) Biomass Energy

   i. Would biomass energy provide a low-cost energ/ reduce grid electricity billy

   ii. Does biomass energy reduce farm waste while providing sustainable energy?

XXV Energy efficient and sustainable building structures or appliances

(a) Energy-efficient lighting (LEDs, etc.)

   i. Would farms using LED or energy-efficient lighting reduce emissions?

   ii. Does energy-efficient lighting lower operational costs

(b) Insulated and/or green roofs or walls

   i. Would insulated or green roofs/walls reduce energy/ grid electricity cost?

   ii. Does it increase biodiversity?

   iii. Does it increase mental health of building dwellers

(c) Rainwater harvesting systems

   i. Would rainwater harvesting reduce cost on municipal water?

   ii. Does rainwater harvesting reduce runoff and erosion?

(d) Green building materials (recycled/ locally sourced)

i. Would using green building materials reduce environmental impact and waste?

(e) Smart climate control system

i. Would smart climate control systems optimise energy use and reduce emissions?

ii. Do these systems improve yields and profits

XXVI Organic residue Management methods

(a) Composting

i. Would composting recycle organic waste?

ii. Would composting increase moisture conservation in soil?

iii. Does composting enhance soil fertility, reduce synthetic fertilizer needs?

iv. Does composting increase yield and thereby profits?

(b) Vermicomposting (using worms to decompose organic waste)

i. Would vermicomposting promote microbial activity and improve soil quality parameters?

ii. Does vermicomposting reduce fertilizer cost?

iii. Would composting increase moisture conservation in soil?

iv. Does vermicomposting increase yield and thereby profits?

(c) Mulching

i. Would mulching reduce soil erosion and retain soil moisture?

(d) Incorporating Crop Residues into Soil

i. Would incorporating crop residues improve soil organic matter and nutrient cycling?

ii. Does this practice support following crop productivity, reduce fertilizer costs?

XXVII Waste Management Practice

(a) Disposed at landfill

    i. Would disposing plastic/chemical containers in landfills contribute to soil and land pollution?
    ii. Does landfill disposal involve low direct cost of waste handling
    iii. Does it increase health vulnerability?

(b) Recycled
    i. Would recycling reduce pollution sources, and overall waste?
    ii. Does recycling lower material costs in the farm?
    iii. Does it reduce health vulnerability?

(c) Returned using a take-back program
    i. Would take-back programs reduce waste handling costs?
    ii. Would take back program reduce pollution sources, and overall waste?
    iii. Does it reduce health vulnerability?

(d) Reused or repurposed
    i. Would reusing or repurposing materials reduce waste and resource consumption?

(e) Unspecified
    i. Does reuse maximize material lifecycle, lower costs, and improve sustainability?

(f) Burning / Burying
    i. Would burning or burying plastic/chemical containers release pollutants and contaminate soil/air?
    ii. Does this practice pose high health/environmental risks

(g) Farm reduces waste (reusable, biodegradable, less plastic)
    i. Would adopting reusable, biodegradable, or reduced-plastic practices minimize waste generation?
    ii. Does preventive waste reduction lower costs and improve health

XXVIII Mode of transportation

(a) Personal vehicle (gasoline/diesel)

i. Would using gasoline/diesel vehicles contribute to high emissions and air pollution?

ii. Does this mode provide convenience but remain vulnerable to fuel price fluctuations and traffic delays?

(b) Personal vehicle (electric/hybrid)

i. Would electric/hybrid vehicles reduce emissions compared to fossil fuels?

ii. Does this mode improve safety and efficiency but depend on charging infrastructure and higher initial costs?

(c) Bicycle

i. Would using bicycles result in zero emissions and minimal environmental pollution impact?

ii. Is this mode low-cost and healthy, but limited by cargo capacity and distance?

(d) Public transportation

i. Would public transport lower per-capita emissions?

ii. Is it cost-effective but dependent on schedule and availability?

(e) On foot

i. Does walking produce zero emissions and no cost?

ii. Is it safe and healthy but limited to short distances?

(f) Delivery service

i. Would using delivery services vary in emissions depending on fleet efficiency and distance?

ii. Does it offer convenience for bulk transport but reliability depends on service and logistics?

XXIX Institutional Support

(a) Significant support (funding, training, resources)

i. Would significant institutional support encourage adoption of sustainable practices (solar panel, soil test, water test, renewable energy sources) and reduce negative impacts?

ii. Does it improve productivity, reduce input costs, and enhance system resilience?

(b) No support and unaware of programs

i. Would lack of awareness reduce adoption of sustainable practices and increase environmental pollution risks?

ii. Does this lead to low health/safety improvement and weak system resilience?

iii. Doe this lead to low profit and reduce production capacity?