http://dx.doi.org/10.62762/TETAI.2026.882738

RESEARCH ARTICLE

# AgriEvoKG: Empirical Failure Modes and Remedies in LLM-Based Cross-Media Agricultural Knowledge Graph Construction

**Xin Zeng**[1,2,3], **Wang Guo**[1,2,3], **Yisheng Miao**[1,2,3], **Huaji Zhu**[1,2,3] **and Huarui Wu**[1,2,3,*]

[1] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
[2] Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[3] Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100097, China

## Abstract

**Large language models (LLMs) make it possible to build knowledge graphs from professional literature at scale, but how such graphs fail at that scale is rarely measured. We built AgriEvoKG, a cross-media knowledge graph of swine, aquaculture, tomato, maize and rice production, and analysed its 199,418 entities and 277,053 relations together with the web system that serves it. An LLM extracted typed triples from the professional literature of each sector; 74.9% of the relations carry an evidence record of document, page and excerpt, and a multimodal embedding model links 25,158 entities to images and 14,132 to video clips. We identified five failure modes of construction. First, the sector schemas diverged: the two sectors with free relation labels hold 23,727 and 19,139 distinct labels. Second, in three sectors most numeric facts lost the name of the measured quantity or object, because 85.4% to 98.0% of their value triples use a generic relation. Third, evidence anchoring was asymmetric: the LLM copied 87.5% of the values verbatim from their excerpts but only 38.7% of the head entities. Fourth, exact-name fusion merged 6,805 entity names across sectors but only 7 relations, and about half of the cross-sector relations touch value-like names. Fifth, media links concentrated on hub items: 104 of 15,314 media items held 71.6% of the links. On the deployed system, 62.5% of the relations added from 50 new passages used new relation types, and in a case study on 100 photographs graph context confirmed the model's initial identification rather than correcting it. We organise these findings into a taxonomy that pairs each failure mode with its root cause, a remedy and a test, and argue that most of them cannot be seen in the small samples on which construction methods are usually judged.**

**Keywords**: cross-media knowledge graph, large-language-model-based knowledge extraction, failure analysis, provenance, cross-media grounding, knowledge graph update, smart agriculture.

**Submitted:** 19 June 2026
**Revised:** 19 October 2026
**Accepted:** 26 October 2026
**Published:** 02 November 2026

**Vol.** 3, **No.** 4, 2026.
10.62762/TETAI.2026.882738

***Corresponding author:**
✉ Huarui Wu
wuhr@nercita.org.cn

**Citation**
Zeng, X., Guo, W., Miao, Y., Zhu, H., & Wu, H. (2026). AgriEvoKG: Empirical Failure Modes and Remedies in LLM-Based Cross-Media Agricultural Knowledge Graph Construction. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 3(4), 239–266.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

## 1 Introduction

Agriculture is among the most knowledge-intensive industries, and its knowledge is among the most dispersed. Decisions on breeding, disease and pest control and cultivation management depend on expert knowledge scattered across technical books, extension manuals, industry reports, photographs and field videos. Smart agriculture aims to bring this knowledge to the point of decision through data-driven systems [1]. Knowledge graphs have become a central tool for this purpose, because they turn dispersed text into a queryable network of entities and relations [2, 3]. Since 2022, large language models (LLMs) have raised expectations further, because they can answer questions over such a network in natural language [4]. However, an LLM without domain grounding fabricates agronomic facts [5], and its answers cannot be traced to a source that a farmer or an extension officer can check.

Agricultural knowledge is inherently cross-media. A disease is recognised from lesions that only an image can show, and an operation such as transplanting or mechanised harvesting is best conveyed by video. The accompanying text carries the causes, thresholds and control measures. Multimodal knowledge graphs, which attach images and other media to symbolic entities, were introduced for this reason [6, 7]. General-purpose resources link encyclopedic entities to images, videos and audio from the open web [6, 8–10]. More recent methods build such graphs without manual annotation [11] or for retrieval-augmented generation [12–14]. These resources contain little of the specialised knowledge, recorded in professional books, that livestock, aquaculture and crop production require.

Agricultural knowledge graphs, by contrast, have been built almost entirely from text, either by manual curation [15] or by extraction from web encyclopedias and scientific articles [16, 17]. In parallel, LLMs have made automatic graph construction practical. Recent methods extract entity-relation graphs from document collections for retrieval [18, 19] and address schema definition, incremental construction and entity canonicalisation [20–22].

Three gaps remain. First, no existing graph jointly covers livestock, aquaculture and crop production with image and video grounding, although these branches of production share pathogens, inputs and management practices and are often managed on the same farm. Second, few resources record where each fact came from or offer a path for adding knowledge after release. Without an evidence record a user cannot check a fact, and without a way to add knowledge the graph falls behind as new varieties, diseases and regulations appear. Third, little is known about LLM-extracted graphs at the scale of hundreds of thousands of facts drawn from professional books. Reports on automatic construction usually give precision on small samples [20, 23] and rarely describe schema drift, the attachment of numeric facts or the connectivity of the result, although these properties decide whether a graph can ground reasoning.

This paper uses AgriEvoKG, a system that builds, updates and uses a cross-media knowledge graph for five agricultural sectors, to study how LLM-based construction fails at this scale. For each sector, an LLM extracted typed triples from the professional literature together with the supporting document, page and excerpt; the triples of a sector form its sector file. We fused the five sector files into one graph by exact entity name, which we call exact-name fusion. A multimodal embedding model links the entities to images and video clips, and a lexical prior derived from file names and folder labels reranks the retrieved candidates. A web system serves the graph, lets editors add knowledge from text and supports image diagnosis.

Our analysis of the sector files and the deployed system identifies five failure modes of construction. The sector schemas diverged, and most numeric facts in three sectors lost the name of the measured quantity or object. Evidence records anchored the tail of a triple better than its head, and exact-name fusion merged many shared entity names but almost no shared relations. Media links concentrated on a small number of hub items, that is, media items linked to at least 100 entities each. The deployed system adds three further failures: editor ingestion introduced new relation labels and overwrote sectors, control knowledge was sparse, and in a diagnosis case study graph context confirmed the model's initial identification instead of correcting it. We organise these failures into a taxonomy of root causes, remedies and tests. Our contributions are:

- An empirical analysis of the sector files and the deployed system that identifies five failure modes of LLM-based construction and three of deployment. A taxonomy pairs each failure mode with its root cause, a remedy and a test that can be recomputed

on the data (Section 6).

- A joint cross-media graph of swine, aquaculture, tomato, maize and rice production with 199,418 entities and 277,053 relations. Of these, 207,421 (74.9%) carry an evidence record and 150,187 (54.2%) carry the full record of document, page and excerpt. Relations inferred from the graph rather than extracted from text carry an inference marker or no evidence record, so they are kept apart from extracted ones.
- An entity-to-media grounding pipeline that keeps at most two images and one 10 s video clip per entity. It links 25,158 entities to images and 14,132 to video clips, and in four of the five sectors every clip link points to the matching segment.
- A deployed web system that combines graph browsing with linked media, editor ingestion with a per-batch cap and a sampled preview, and a three-stage diagnosis workflow. The workflow uses the graph as context during reasoning, applies a vision-dominant decision rule, discloses a fallback to general experience when the graph holds no relevant knowledge, and afterwards shows a conclusion subgraph with reference media as evidence.

Section 2 reviews related work. Section 3 describes the system, Section 4 presents the graph and Section 5 evaluates both. Section 6 organises the failures into a taxonomy, Section 7 discusses lessons and limitations, and Section 8 concludes.

## 2 Related Work

### 2.1 Automatic knowledge graph construction with LLMs

Automatic knowledge graph construction has traditionally been a pipeline of named entity recognition, relation extraction and entity linking, each stage trained on annotated data [23]. LLMs have replaced most of these stages with prompting. Trajanoska et al. [24] showed that a general LLM extracts triples from text with no task-specific training, and Carta et al. [25] prompted an LLM iteratively to extract entities, relations and a schema in separate passes. Text2KGBench [26] measures whether such output conforms to a given ontology and stays faithful to its source sentences. Subsequent methods improve the extracted graph itself. EDC [20] extracts open triples, defines their schema and canonicalises them to control schema drift. iText2KG [21] constructs graphs incrementally from document streams and matches new entities and relations against the existing graph to avoid duplicates. KGGen [22] clusters extracted entities to reduce sparsity.

Other methods build graphs from documents for retrieval. GraphRAG [18] extracts entities, relations and community summaries from a document collection so that an LLM can answer global questions, and LightRAG [19] simplifies this design for fast incremental indexing. Agent frameworks such as ReAct [27] interleave reasoning with tool calls, and Graph-R1 [28] trains such an agent for graph retrieval with reinforcement learning.

Our construction follows this line of work in prompting an LLM with a schema and deduplicating its output. iText2KG is the closest to our ingestion path, because it also adds new text to an existing graph. We do not propose a new extraction algorithm. Instead, nearly all of our extracted triples keep an evidence record with their page and excerpt, and inferred relations are kept apart. Unlike the text-only methods above, our graph is grounded in images and videos.

### 2.2 Multimodal and cross-media knowledge graphs

Multimodal knowledge graphs extend symbolic graphs with visual and other media [7]. MMKG [6] attached web images to the entities of FB15k, DB15k and YAGO15k, and Richpedia [8] collected and filtered images for Wikidata entities at scale. VisualSem [9] curated a high-quality graph with images and multilingual glosses, and TIVA-KG [10] extended coverage to text, image, video and audio for ConceptNet concepts. Construction has since moved toward annotation-free methods. VaLiK [11] captions images with vision-language models and links the captions to graph entities. MMGraphRAG [12] builds document-level graphs in which images become entities connected to the surrounding text. VAT-KG [13] builds a knowledge-intensive graph over video, audio and text for retrieval-augmented generation, and query-driven variants construct a local multimodal graph online for each question [14].

Linking entities to media is itself a retrieval problem. Cross-media retrieval [29] and multimodal entity linking methods such as KGMEL [30] match textual and visual representations in a shared space. Multimodal completion models such as MKGFormer [31] use the linked media for link prediction. Like VaLiK and KGMEL, our grounding pipeline links entity descriptors to images and

**Table 1.** Comparison of AgriEvoKG with representative multimodal and agricultural knowledge graphs. ✗, none; LLM, large language model; VLM, vision-language model.

| Resource | Domain | Media | Construction | Provenance | Update |
|---|---|---|---|---|---|
| MMKG [6] | general | text, image | automatic | ✗ | ✗ |
| Richpedia [8] | general | text, image | semi-automatic | not reported | ✗ |
| TIVA-KG [10] | general | text, image, video, audio | automatic | not reported | ✗ |
| VaLiK [11] | general | text, image | automatic (VLM) | not reported | ✗ |
| MMGraphRAG [12] | documents | text, image | automatic (LLM) | document | ✗ |
| VAT-KG [13] | general | video, audio, text | automatic | not reported | ✗ |
| AGROVOC [15] | agriculture | text | manual | not applicable | curated releases |
| AgriKG [16] | agriculture | text | automatic (web) | not reported | ✗ |
| PlantConnectome [17] | plant science | text | automatic (LLM) | article | not reported |
| AgriEvoKG (this work) | five sectors | text, image, video | automatic (LLM) | page, excerpt | editor ingestion |

video clips, here through a multimodal embedding model [32]. Our setting differs in that the media come in topic folders whose hand-written file names carry much of the reliable signal. We therefore rerank the embedding-based candidates with a lexical prior and keep only a few links per entity.

### 2.3 Agricultural knowledge graphs and language models

Semantic technologies have a long history in agriculture. AGROVOC [15] is a manually curated multilingual thesaurus maintained by the Food and Agriculture Organization of the United Nations. Drury et al. [1] surveyed the field, and subsequent work examined how LLMs and knowledge graphs can be unified [33]. Automatically constructed graphs followed. AgriKG [16]… Domain language models take a third route. ShizishanGPT [36] connects an LLM to tools and knowledge bases; evaluation studies compare fine-tuning with prompting for LLM-based graph construction [35], and provenance verification pipelines assess the reliability of the extracted facts [34].

On the visual side, PlantVillage [37] and PlantDoc [38] provide labelled leaf images for disease classification. They are label sets rather than graphs and carry no causal or management knowledge. To our knowledge, no previous resource links crop, livestock and aquaculture knowledge to images and videos in one graph, and none was built from professional books and industry reports with page-level provenance for most of its relations. Both properties matter for a decision-support system whose advice an agronomist must check. Table 1 places AgriEvoKG among the resources reviewed here.

### 2.4 Graph-grounded reasoning with LLMs

Retrieval-augmented generation [39] reduces hallucination by conditioning an LLM on retrieved text. Graph-based variants retrieve structured neighbourhoods instead. Think-on-Graph [40] lets the LLM explore the graph step by step, Reasoning on Graphs [41] plans relation paths before retrieval, and GraphRAG [18] retrieves community summaries. Other methods encode the graph for the model. MR-MKG [42] extends graph-based retrieval to multimodal graphs and improves an LLM's multimodal reasoning with a graph encoder, and SSQR [43] quantises graph structure into discrete tokens for an LLM. Vision-language models such as LLaVA [44], Qwen-VL [45] and GPT-4V [46] supply the visual perception these methods require.

Our diagnosis workflow does not train the model. Unlike the graph-based variants above, it passes only the names of retrieved entities to the model as context, and a vision-dominant decision rule stops the graph from overriding the image. After the answer, the workflow shows a conclusion subgraph whose relations keep their source pages in the database. We use the workflow as the downstream application for evaluating the graph, not as a new reasoning method.

## 3 The AgriEvoKG System

### 3.1 Architecture and graph model

AgriEvoKG has three parts (Figure 1). An offline construction pipeline (Sections 3.2 and 3.3) extracts triples from each sector's literature with an LLM and imports them into one Neo4j graph [47]. An offline grounding pipeline (Section 3.4) links the entities to images and video clips. An online web system

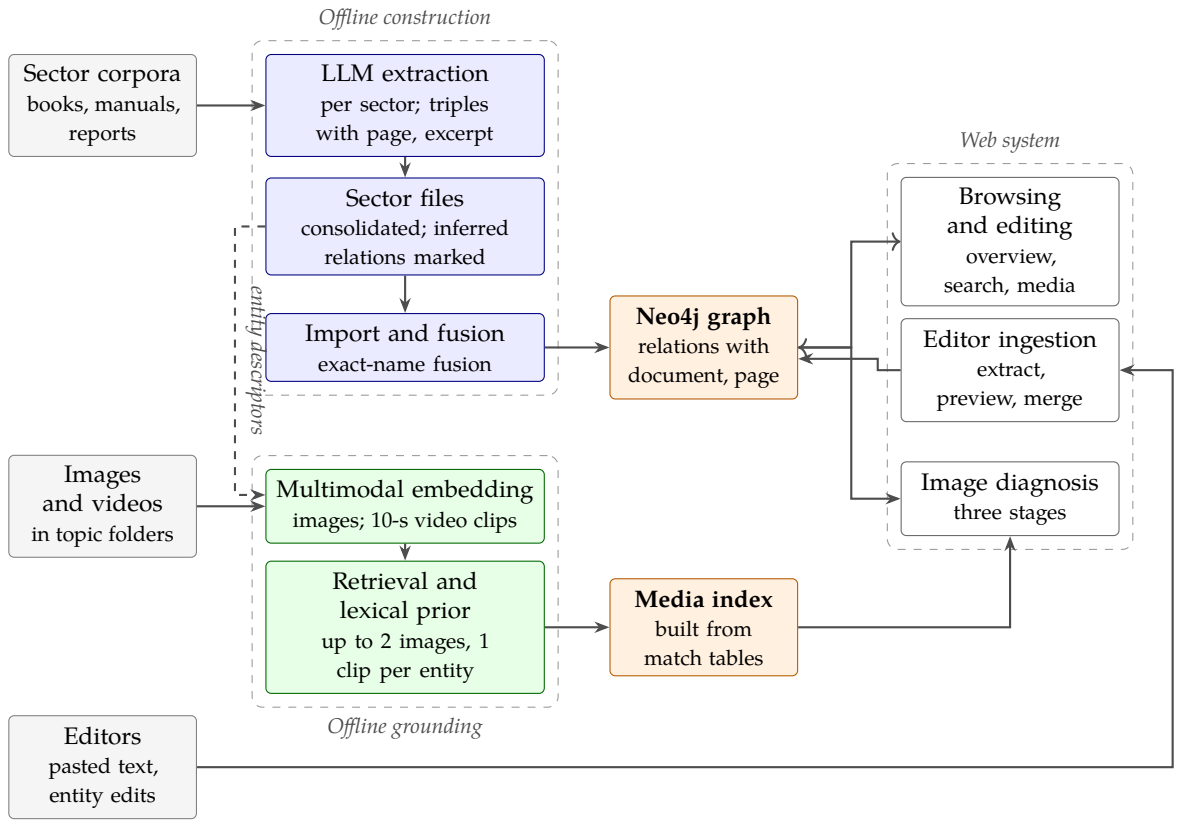



**Figure 1.** Architecture of AgriEvoKG. The construction pipeline turns each sector corpus into a sector file of triples with evidence records and imports the five files into one graph. The grounding pipeline links the entities to images and video clips, and the web system serves the graph and the media links, accepts editor ingestion and runs the diagnosis workflow.

(Sections 3.5 to 3.7) serves the graph and media links, accepts new knowledge from editors and runs the diagnosis workflow.

In our graph model, a cross-media knowledge graph is a tuple $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{M}, \mathcal{A}, \pi)$. Entities $\mathcal{E}$ are identified by name, and all but 215 of them carry one or more entity types assigned during extraction. Relations $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ carry relation types from $\mathcal{R}$, and scored links $\mathcal{A} \subseteq \mathcal{E} \times \mathcal{M} \times \mathbb{R}$ join entities to images and video clips $\mathcal{M} = \mathcal{M}^{\text{img}} \cup \mathcal{M}^{\text{clip}}$. The partial map $\pi$ on $\mathcal{T}$ assigns to a relation the document, page and excerpt it was extracted from, where one exists. A sector function $\sigma : \mathcal{E} \to S$, with $S = \{\text{swine}, \text{aquaculture}, \text{tomato}, \text{maize}, \text{rice}\}$, assigns every entity the sector in which it first entered the graph. A relation is attributed to the sectors of both its endpoints, so sector-level relation counts overlap. In this paper, a triple is an extraction record in a sector file and a relation is an edge of the graph.

The import (Section 3.3) guarantees three properties. (P1) Every entity takes part in at least one relation, because only triples are imported. (P2) One entity exists per name. (P3) At most one relation exists per head, relation type and tail, where the relation type is the normalised relation label. The ingestion path (Section 3.5) preserves P2 because it also merges by name. It preserves P3 only within its own normalisation: it turns punctuation in a label into underscores instead of removing it, so a label that the import had normalised can re-enter as a second relation type.

Manual creation and editing of entities in the interface do not check for existing names and can break P2, and neither path enforces P1.

### 3.2 Corpus and extraction

For each sector, we collected professional books, technical manuals, annual industry reports and journal articles (Table 2). For tomato, rice and aquaculture, a PDF counted as born-digital when at least 30% of its first 20 pages held 20 or more Chinese characters, and as scanned otherwise. The rice corpus, for example, held 211 documents: 16 born-digital PDFs, 14 EPUB books and 181 scanned PDFs. Scanned pages were rendered at double resolution and read with PaddleOCR [48]. For aquaculture, OCR was applied only to pages whose text layer held fewer than 50 characters. For swine and maize, we read only the text

**Table 2.** Sector files and their triples.

| Sector | Documents | Triples | Extracted |
|---|---|---|---|
| Tomato | 107 | 50,476 | 50,476 |
| Rice | 83 | 50,552 | 50,552 |
| Aquaculture | 44 | 54,666 | 54,666 |
| Swine | 21 | 68,501 | 25,361 |
| Maize | 24 | 52,865 | 26,373 |
| Total | 279 | 277,060 | 207,428 |

*Note:* The 277,060 triples yield 277,053 relations, because 7 maize triples repeat a relation of an earlier sector and are merged during fusion (Table 8). Documents are those that contribute at least one triple.

```
{
 "head": {"name": "yield per mu", "type": "Parameter indicator"},
 "relation": "national rice yield per mu in 2014 was",
 "tail": {"name": "454.0 kg", "type": "Quantitative value"},
 "evidence": [{
  "document": "China Rice Industry Development Report 2015",
  "page": 5,
  "excerpt": "yield per mu 454.0 kg, an increase of 6.2 kg",
  "location": "PAGE_5, 'yield per mu 454.0 kg, an increase of
               6.2 kg'"
 }]
}
```

**Figure 2.** A triple from the rice sector file (translated). Its relation label carries the indicator, scope and year, and its evidence record gives the document, page and excerpt. The mu is a Chinese unit of area (1 mu $\approx$ 666.7 $m^2$), so 454.0 kg per mu corresponds to about 6.81 t $ha^{-1}$.

layer with PyMuPDF. Eight of the 32 maize documents have almost no text layer and therefore yielded no triple.

Page filters differed by sector. Tomato pages with fewer than 80 characters were dropped, as were pages with front-matter keywords but no tomato keyword. Aquaculture pages were kept only when they held at least 50 characters and one of 27 sector keywords. The remaining pages formed extraction windows of four consecutive pages for tomato, five for rice and six for aquaculture. For swine and maize, a window holds three pages with more than five characters of text, so it spans more page numbers when nearly empty pages lie between them. Windows were cut at 7,500 characters for rice and at 7,000 for tomato and aquaculture. All 83 rice documents that contribute triples are scanned.

For each window, an LLM returned triples in JSON: Doubao Seed 1.6 Flash (`doubao-seed-1-6-flash-250828`) for rice, Doubao Seed 2.0 Mini (`doubao-seed-2-0-mini-260215`) for tomato and aquaculture, and Doubao Seed 2.0 Lite (`doubao-seed-2-0-lite-260215`) for swine and maize. All three were called through the Volcano Engine API at a temperature of 0.1, and malformed JSON was repaired before parsing. Each triple has a typed head, a relation label, a typed tail and an evidence record (Figure 2). The record names the document by the title in its file name. The model supplied the page, taken from page markers in the window, and an excerpt of at most 100 characters. For swine and maize, the record gives the page range of the window instead, and the excerpt is limited to 50 characters. Rice and tomato records also hold a location hint. The tomato prompt forbade invented knowledge, whereas the aquaculture prompt asked the model to shorten entity names to their essential words.

**Sector schemas.** Each sector had its own schema, written when the sector was processed (Table 3). The swine and maize schemas were closed. They fixed 10 entity classes each and 14 and 13 relation types, respectively, and their prompts forbade new relations. Their files hold 11 and 12 entity types and 21 and 22 relation labels. Beyond the schemas, the files contain an "unclassified" type and, in maize, a type "yield component" used twice. They also contain the labels of the inferred relations and a few stray labels of extracted triples, two in swine and three in maize, such as "associated". The aquaculture schema was closed as well: it fixed 13 entity classes and 11 relation labels, and triples with other labels were discarded. The aquaculture file also holds 13 labels of an earlier schema version, in 814 triples. Of its 62 entity types, 59 lie outside the schema, but they cover only 1.2% of the type assignments.

The rice and tomato schemas left more room. For rice, the prompt named seven core classes and let the model write relation labels as free text, preserving the source wording. Tomato was extracted in two configurations. Of its 107 documents, 32 used a 13-class tomato schema that admitted only 13 fixed relation labels, and the other 75 used the seven rice classes with free labels. The model used this room to coin new entity types, so the tomato and rice files contain 273 and 250 distinct non-empty entity types. The free relation labels diverged further. The rice and tomato files hold 23,727 and 19,139 distinct relation labels, and 79.0% and 83.1% of these labels occur only once. The seven core classes cover 81.8% of the type assignments in tomato and 98.4% in rice, and the classes of the tomato schema account for 17,383 of the remaining 18,336. Section 4 analyses the consequences of this divergence.

**Table 3.** Entity classes of the sector schemas. The last two columns give the numbers of distinct entity types and relation labels in each sector file.

| Sector | Entity classes | Entity types | Relation labels |
|---|---|---|---|
| Tomato | 75 documents: the seven rice core classes; 32 documents: pedigree; gene and molecular marker; agronomic trait; fruit trait; biotic stress; abiotic stress; trait expression; growth stage; breeding and propagation technique; cultivation practice; chemical substance; quantitative value; location and institution | 273 | 19,139 |
| Rice | germplasm; trait; practice; parameter indicator; quantitative value; environmental factor; growth stage (core classes; further types coined by the model) | 250 | 23,727 |
| Aquaculture | farmed object; taxon; pathogen; disease; symptom; nutrient and component; physicochemical and biochemical indicator; production and industry indicator; region; farming stage and mode; technique and input; quantitative value; institution | 62 | 24 |
| Swine | clinical and anatomical sign; disease and pathological state; feed, nutrition and drug; value and dose; management and intervention; production and health indicator; environment and facility parameter; pathogen and vector; physiological stage; organ and anatomical part; unclassified | 11 | 21 |
| Maize | germplasm and variety; input and cultivation technique; disease, pest, weed and pathogen; stress and resistance; value and parameter; phenotype and agronomic trait; growing environment and zoning; gene and molecular marker; quality and nutrient; institution and breeder; yield component; unclassified | 12 | 22 |

**Triple-form constraint.** A numeric value is meaningful only together with the quantity it measures and the object it describes. A value such as "454.0 kg" is of little use unless a relation names this quantity and object. The rice prompt enforces such relations with the triple-form constraint. It asks for triples with evidence, and for a concise core verb when both ends are textual. Whenever the tail is a quantitative value, the relation label must include the conditions of the value, such as variety, environment and period. The excerpt "yield per mu 454.0 kg, an increase of 6.2 kg" then yields the triple in Figure 2, whose label names the indicator, scope and year.

The rice relation labels are long and specific by design, because the relation, not the number, carries the meaning of a numeric fact. The other schemas did not adopt this constraint. The tomato schema of the 32 documents and the aquaculture schema prescribe the generic chain "has indicator" followed by "value is". The swine and maize schemas require every value with a unit to be attached through the generic relation "quantitative parameter". Section 4 quantifies the effect of these generic relations.

**Consolidation.** For tomato and aquaculture, consolidation removed records with empty or placeholder names such as "unknown", names shorter than two characters, self-loops and exact duplicates. For rice, whose self-loops had been dropped during extraction, it removed empty fields and duplicates. In each case, a name kept its first type, and the evidence records of duplicates were merged into a list, of which the import stores only the first. In aquaculture, 59,120 raw triples yielded 54,666 after removing 1,361 invalid records, 4 self-loops and 3,089 duplicates. In tomato, 51,308 yielded 50,476 after removing 181 invalid records and 651 duplicates. For swine and maize, extraction produced 29,355 and 30,908 raw triples, and the released files hold 25,361 and 26,373 extracted triples. Of these, 98.4% and 93.3% can be traced to the extraction output of their source document, with the same head, relation label, tail, excerpt and page range. The others cannot be traced to a document.

**Inferred relations.** Only the swine and maize files contain triples not extracted from text, which become

inferred relations in the graph. The tomato, rice and aquaculture files hold extracted triples only. In swine, 23,123 such triples carry the marker “machine inference” in place of a page and a fixed statement that they were derived from the topology of the swine graph. They state that an input or practice relieves a sign (13,867) or that two conditions require differential diagnosis or share symptoms (5,475). Others state that a condition indirectly induces a symptom (3,571) or causes a local lesion (210). A further 20,017 swine triples, which carry no evidence record, state that two conditions commonly co-occur, in 16,469 cases two clinical signs. In maize, 26,492 triples are marked as computed by graph algorithms. They link varieties with similar agronomic traits (9,237), parallel technical schemes (7,347), varieties and the traits they require (5,237), co-occurring breeding traits (1,880), regionally matched recommendations (1,600) and regional cultivation parameters (1,191).

The sector files record the marker and the relation label of each inferred triple but not the rules that produced it. The import stores the marker, or the absence of an evidence record, with the relation, so inferred relations remain distinguishable. The extracted swine and maize triples, in turn, cite a placeholder title instead of the document title, together with a page range and an excerpt. The traced triples come from 21 swine and 24 maize documents. We keep the inferred relations because the deployed system uses them, but report them separately.

### 3.3 Cross-sector fusion

We imported the five sector files into Neo4j in the order tomato, swine, maize, rice and aquaculture. The order matters because an entity belongs to the sector where it first entered the graph (Section 3.1). The import merges entities by exact name and relations by head, relation type and tail (P2 and P3). Normalisation keeps only Chinese characters, Latin letters and digits, and the graph holds 41,665 distinct relation types. Each relation stores its relation label, document and page. In the deployed database, the 171,835 relations of tomato, swine and maize also store their excerpt where the sector file has one, or a note that they were inferred. For rice and aquaculture, the excerpts remain in the sector files.

Fusion across sectors is therefore exact-name fusion, the simplest form of entity resolution [49]: a name shared by two sector files becomes one entity. Tomato and maize, for example, share pests such as the cotton bollworm (*Helicoverpa armigera*) and the beet armyworm (*Spodoptera exigua*), and swine and aquaculture share a sign such as diarrhoea and a drug such as florfenicol. Of the 199,418 entities, 6,805 names occur in two or more sector files: 4,956 in two, 1,278 in three, 408 in four and 163 in all five. In all, 37,146 relations (13.4%) join entities attributed to different sectors. Section 5.2 measures what this fusion merged and what it missed.

### 3.4 Cross-media grounding

Grounding links entities to images and video clips (Section 1). For the five sectors, we assembled 11,936 images and 3,378 videos, sorted into topic folders such as diseases, pests, cultivation and harvesting (Table 4). The hand-written folder names, which we call folder labels, and the file names usually state the topic of a media item, for example the disease shown. A folder label describes the folder as a whole rather than each entity that an item shows. We therefore retrieve media by content and use file names and folder labels only to rerank.

**Embedding.** We embedded the images and videos with the 8B Qwen3-VL-Embedding model [32], which maps text, images and video into one 4,096-dimensional space, in the manner that contrastive models such as CLIP [50] align images and text. A Qdrant vector database [51] stores the vectors, with one collection per sector and media type. Images that fail a file-integrity check and videos larger than 200 MB are skipped, and images are embedded in groups of eight. Each video is cut into segments of 10 s with an overlap of 5 s, so that a link can point to the matching segment. Each segment is re-encoded at 4 frames per second and a height of 112 pixels, sampled by the model at one frame per second and embedded as a clip with its start and end time. For every entity that heads at least one triple, the pipeline writes an entity descriptor $d(e)$ of the form “This is an image of ⟨name⟩ (⟨type⟩). Features: ⟨relation⟩⟨tail⟩, …”, which lists its outgoing relations, and embeds it in the same space. The descriptor of a disease can thus retrieve photographs of its lesions.

**Scoring.** For each entity, the pipeline retrieves the 15 nearest images and the 400 nearest clips, a larger pool because the many clips of one long video would otherwise fill it. The match score of a candidate image or clip $m$ for entity $e$ adds a lexical prior to the visual similarity,

$$s(e, m) = \cos\big(\phi(d(e)), \phi(m)\big) + b(e, m), \qquad (1)$$

where $\phi$ is the embedding model and the prior $b$ takes the first applicable value of

$$b(e,m) = \begin{cases} 100 & \text{if } c(e) \text{ contains } c(f_m) \\ & \text{or is contained in it,} \\ 80 & \text{if } c(e) \text{ belongs to a synonym} \\ & \text{group of which a member} \\ & \text{occurs in } c(f_m), \\ 60\,\rho & \text{if } |c(e)| > 2 \text{ and} \\ & \rho = \mathrm{ratio}(c(e), c(f_m)) \geq 0.65, \\ 20 & \text{if } c(e) \text{ occurs in } c(p_m), \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Here $c(\cdot)$ keeps the Chinese characters of a string, $c(e)$ is short for $c$ applied to the name of $e$, $f_m$ is the file name of $m$ and $p_m$ its stored path, which includes the folders, and ratio is the Ratcliff–Obershelp similarity of two character sequences. All cases require $c(e)$ and $c(f_m)$ to be non-empty, since an empty string is contained in every string; in particular, when the file name contains no Chinese character, the prior is zero whatever the path. The synonym groups map colloquial names to standard ones, for example "heart-boring worm" to the Asian corn borer (*Ostrinia furnacalis*).

After rescoring with Eq. (1), the pipeline keeps at most two images with $s \geq 0.45$ and one clip with $s \geq 0.40$, recording each link's score, rank and, for a clip, start time in a match table per sector. An earlier version used a fuzzy-match threshold of 0.50 and a pool of 60 clips. Because the cosine term never exceeds 1 and a non-zero prior is at least 20, a candidate with lexical support always outranks one without it. The visual term orders candidates with the same prior and decides alone when none has lexical support. The prior applies only to retrieved candidates, however, so a correctly named file outside the visually nearest candidates cannot be selected.

Each sector's entities are matched only against that sector's media, so a name shared by several sector files is grounded once per sector. In the current graph, 25,158 entities (12.6%) have at least one image link and 14,132 (7.1%) at least one clip link. The swine links come from an earlier version that recorded no start times. Their scores, between 0.38 and 2.64, suggest that it gave name matches a much smaller bonus than the prior above (Section 5.4).

**Media index.** The web system loads the match tables into a media index keyed by normalised names. A link is indexed under its entity when it has lexical support ($s \geq 10$) or when its folder label equals, contains or is contained in the entity name after normalisation. Otherwise, it is indexed under its folder label only. Links supported by visual similarity alone thus stay in the match tables but are shown only for entities whose name matches their folder label, a conservative rule that trades coverage for precision.

At query time, an entity name is normalised and expanded with its core term, obtained by deleting every character for "disease" and "symptom". A synonym table further maps signs to diseases, for example fever to classical swine fever and high fever. The diagnosis workflow ranks matches in tiers: 300 for an exact match, 290 for an exact match of sector and name and 180 for a core-term match. Containment is tested only for names that obtain none of these three tiers, and then scores 200, containment of the core term 180 and reverse containment 160 or 150; the numeric values therefore order candidates within one lookup and do not compare containment with core-term matches. Media of the query's sector gain a further 50. The node panel of the interface (Section 3.7) uses only the exact, sector-and-name and core-term tiers and only media of the entity's sector. Within a tier, the match score $s$ is not used, so ties follow the order of the match tables. The system shows the first clip of the highest tier, played from its start time, and the first two images.

### 3.5 Editor ingestion

A knowledge graph for agriculture is never complete: new varieties are registered, new diseases spread, and regulations on inputs change. Work on growing knowledge graphs has mostly addressed how to update embeddings as facts are added [52]; here the concern is how new text enters the graph. AgriEvoKG lets editors add knowledge after release through the web system (Figure 3). Editors sign in with a shared account and receive a token that grants the editor role for two hours. Guests can browse the graph and run the diagnosis but cannot change it.

**Extract.** The editor selects a sector, sets a cap $K$ on the number of relations (10 by default; the input form accepts up to 50) and pastes a passage (Figure 4(a)). The vision-language model of the diagnosis workflow (Section 3.6) extracts relations together with their entities, and the entities and relations extracted from one passage form a batch. The prompt asks for at most $K$ relations to keep each batch small enough to inspect, but a batch that exceeds the cap is not truncated. It

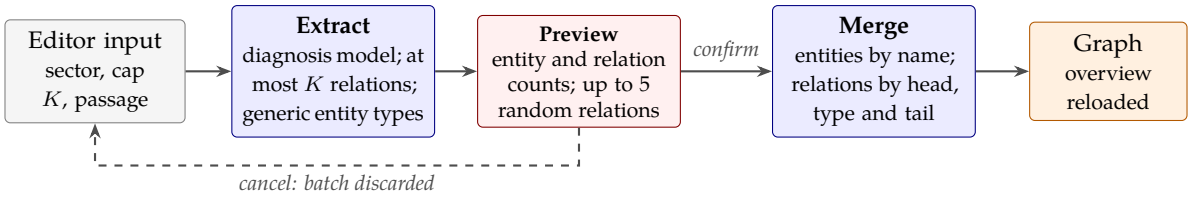
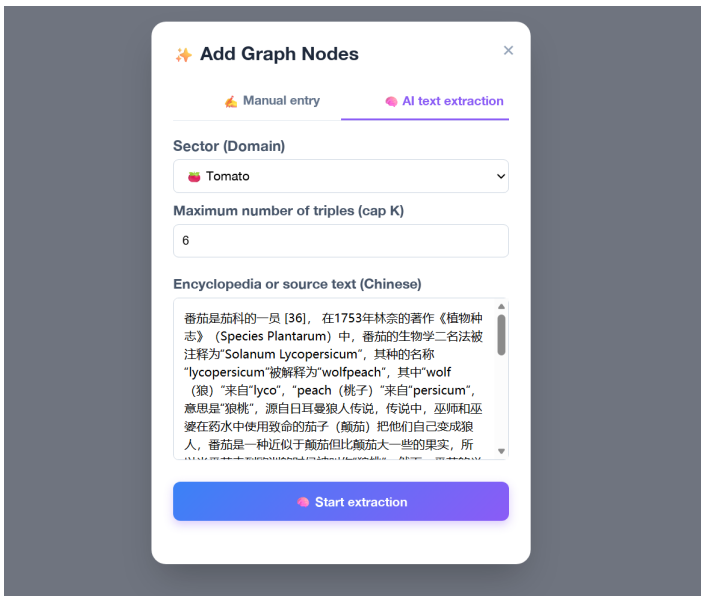
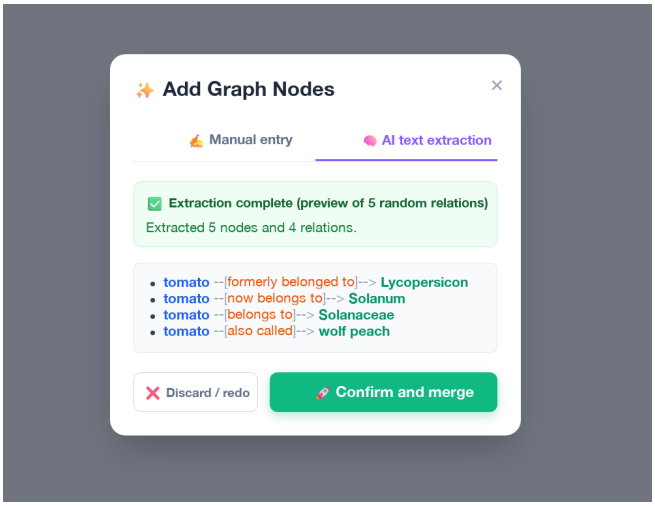



**Figure 3.** Workflow of editor ingestion. A batch is written to the graph only after the editor has inspected its size and a random sample of its relations.




(a) Extract step: sector, cap $K$ and passage.




(b) Preview step: entity and relation counts, sampled relations and a request for confirmation.

**Figure 4.** Screens of the ingestion path in the web interface. English translations are overlaid on the interface labels and the extracted relations; the pasted passage remains in Chinese.

also suggests generic entity types instead of the sector schema, namely disease, crop, symptom, plant part and control method, but does not enforce them and leaves relation labels free.

**Preview.** The interface reports the numbers of extracted entities and relations and shows up to five random relations (Figure 4(b)). This preview is a spot check, not a review of every relation, and keeps the human cost to seconds per batch. The editor confirms the whole batch or cancels and discards it, and can repeat the extraction with a corrected passage.

**Merge.** On confirmation, entities are merged by name and receive the extracted type and the selected sector. Relations are merged on head, relation type and tail, so a relation already in the graph is not duplicated, and a relation whose endpoint is missing is skipped. The interface then reloads the overview (Section 3.7), which is cached for three minutes, so new entities are best found by search. Relations added this way store neither the extracted relation label nor a document or page, and each entity and relation is written in its own transaction (Section 7). Editors can also create, edit and delete individual entities, and these operations are not logged.

New media bypass the ingestion path and enter the graph by rerunning the grounding pipeline on the new media. The video step logs processed files, so only new videos are embedded. The web system loads the new match tables when it restarts.

### 3.6 Graph-assisted image diagnosis

The diagnosis workflow is the main consumer of the graph and the test of its usefulness. It takes a field image $I$ and a question $q$, such as "what disease is this" or "what operation is being performed", and answers in three stages (Figure 5). All stages use the same Doubao vision-language model through the Volcano Engine API.

**Stage 1: perception.** The model receives the image and the question and returns in JSON a short

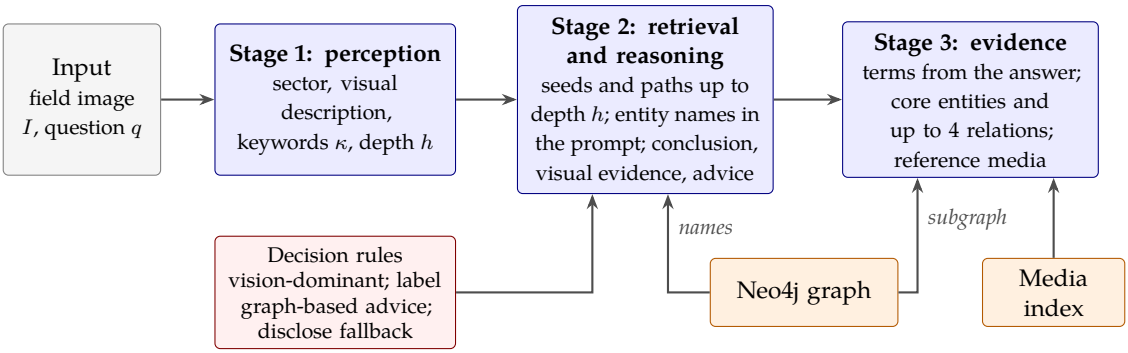



**Figure 5.** The three-stage diagnosis workflow. The graph supplies entity names during reasoning and a conclusion subgraph after it.

description of the visual features, a keyword string $\kappa$ and a retrieval depth $h$ between 1 and 3 (2 by default). The keyword string has the form "sector term", such as "rice mechanised-harvesting", and only $\kappa$ and $h$ are passed on. We call the term in $\kappa$ the initial identification. The prompt asks the model to identify the sector first and to give a specific disease or species name rather than a loose description such as "fish with black spots". If the output cannot be parsed, the keywords become the sector and the term "abnormal", or "agriculture disease" when no sector is recognised.

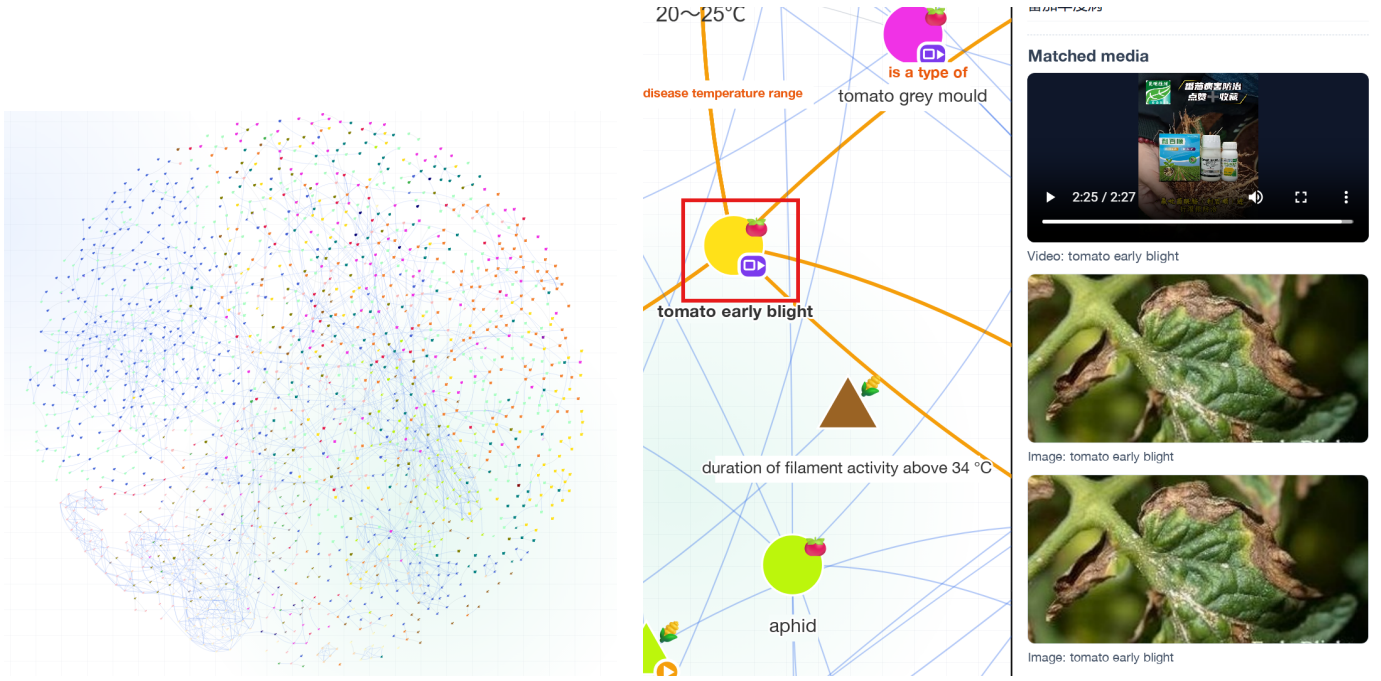
**Stage 2: retrieval and reasoning.** The keywords select up to 35 seed entities of the identified sector whose name contains a keyword or is contained in it, or whose entity type contains one. For each seed, the workflow retrieves up to 18 paths of length 2 to $h$ (length 1 when $h = 1$), longest first. It admits no entity from other sectors when fewer than 10 entities of the identified sector were retrieved, two when fewer than 20 and five otherwise. The query returns at most 500 paths. The names, types and sectors of all retrieved entities enter the system prompt as background knowledge, although the interface draws only the first 120. The model reasons in a separate thinking block, then writes three sections: the conclusion, the visual evidence, and control measures and recommendations, which we call the advice.

Three decision rules govern this stage. First, the vision-dominant rule requires that the model base the identification on the image and not be redirected by unrelated graph entities. We adopted this rule after a graph-dominant version let retrieved entities mislead visual diagnoses. Second, recommendations should draw on the graph where it holds relevant knowledge and must then be labelled as recommended by the graph. Third, when the graph holds no relevant knowledge, the model may use its own knowledge but must state that the advice rests on general agricultural experience. An earlier version without this fallback declined to give advice whenever the graph lacked a control relation.

**Stage 3: evidence.** Stage 3 links the answer to the graph and to media so that the user can inspect its evidence. A further model call extracts two or three terms from the answer, the first being the sector. If it returns none, the Stage 1 keywords are used. Candidates for core entities are the first 50 matching entities, minus names that consist of a number, with or without a unit. Core entities are the candidates whose name equals a term, or the sector followed by a term, and otherwise up to two candidates whose name contains a term. If no rule applies, the first candidate becomes the core entity.

The workflow expands each core entity by one hop and keeps at most four relations within the identified sector, chosen by a relation prior that favours treatment, prevention, drugs, causes, induction, symptoms, manifestations and handling. The interface displays the first 25 entities of this conclusion subgraph next to the answer, or, when no core entity is found, the first 10 entities retrieved in Stage 2. Its relations store their document and page, which the interface does not display. Finally, the terms and the names of the subgraph entities query the media index for at most one clip and two images, reference media that let the user compare the field image with documented cases.

The graph thus plays two roles. During reasoning, it supplies the names of entities of the identified sector as context, but passes neither relations nor source pages to the model. After reasoning, it supplies evidence the user can inspect: a small subgraph around the core entities and media linked to them. Section 5.6 evaluates both roles.




(a) Overview with capped node degrees.

(b) Node panel of tomato early blight (framed), with its linked video clip and images. A video icon marks the node.

**Figure 6.** The web interface. English translations are overlaid on the interface labels and on the entity and relation names.

### 3.7 Web interface

The web system serves the graph from a Neo4j backend. Per sector, its overview starts from the entities whose names occur in the media index and from the 80 entities of highest degree. It adds up to four direct relations around each entity of the first group and up to three around each high-degree entity. It then fills or prunes the view to 1,500 to 1,700 nodes and drops nodes without media whose names contain digits, units or number words. The interface draws at most 10,000 relations and at most 18 per node, except for relations that attach an otherwise unconnected node (Figure 6(a)).

Keyword search retrieves up to three hops around matching entities, and any node can be expanded by one to three hops. A selected node opens the node panel with its properties and linked media, and node icons mark entities whose names have image links, clip links or both in the media index (Figure 6(b)). A selected relation shows its head, relation type and tail, but not the document and page stored with it. Dashboards report entity, relation and media counts per sector, and videos in formats that browsers cannot play are transcoded on first request and cached. All screenshots come from the deployed system, whose interface is in Chinese.

## 4 The Constructed Graph

The graph that the system of Section 3 built is large but sparsely connected, and its five sectors differ in provenance, schema and media coverage. All graph counts refer to the sector files (Table 2), imported as described in Section 3.3. The media counts refer to the media collections, and the media links to the match tables.

**Scale and balance.** Table 4 and Figure 7 show how the graph divides among the sectors. It contains 199,418 entities and 277,053 relations, and no entity is isolated. Tomato (58,715 entities) and rice (50,621) are the largest sectors, and their files cite the most documents (107 and 83). The number of relations per entity ranges from 1.02 for rice to 3.04 for swine. Swine and maize reach high values because they include 43,140 and 26,492 inferred relations among existing entities, so these values reflect inference as much as extraction.

**Provenance.** Table 5 and Figure 8 show that most relations carry an evidence record, whose completeness differs between sectors. More than half (54.2%) can be traced to a document, a page and an excerpt. A further 18.7% can be traced only to a page range and an excerpt, because the swine and maize files replaced the document title with a placeholder. These two files give page ranges, mostly of three pages,

**Table 4.** Statistics of the graph.

| Sector | Entities | Relations | Relations per entity | Images | Videos | With image links | With clip links |
|---|---|---|---|---|---|---|---|
| Tomato | 58,715 | 78,047 | 1.33 | 2,803 | 646 | 8,700 (14.8%) | 1,891 (3.2%) |
| Rice | 50,621 | 51,760 | 1.02 | 2,802 | 962 | 5,563 (10.2%) | 3,861 (7.1%) |
| Aquaculture | 42,291 | 54,433 | 1.29 | 1,031 | 162 | 1,965 (4.4%) | 1,746 (4.0%) |
| Swine | 24,077 | 73,164 | 3.04 | 2,100 | 652 | 5,879 (23.4%) | 4,399 (17.5%) |
| Maize | 23,714 | 56,795 | 2.39 | 3,200 | 956 | 3,192 (12.2%) | 2,341 (9.0%) |
| Total (unique) | 199,418 | 277,053 | 1.39 | 11,936 | 3,378 | 25,158 (12.6%) | 14,132 (7.1%) |

*Note:* Relations count in the sectors of both endpoints, so the sector counts overlap, and the total row gives unique counts. The link shares of each sector divide by the distinct entity names of its sector file, which equal its new plus merged entities in Table 8: tomato 58,715, rice 54,732, aquaculture 44,198, swine 25,089 and maize 26,072. The shares of the total row divide by all 199,418 entities.

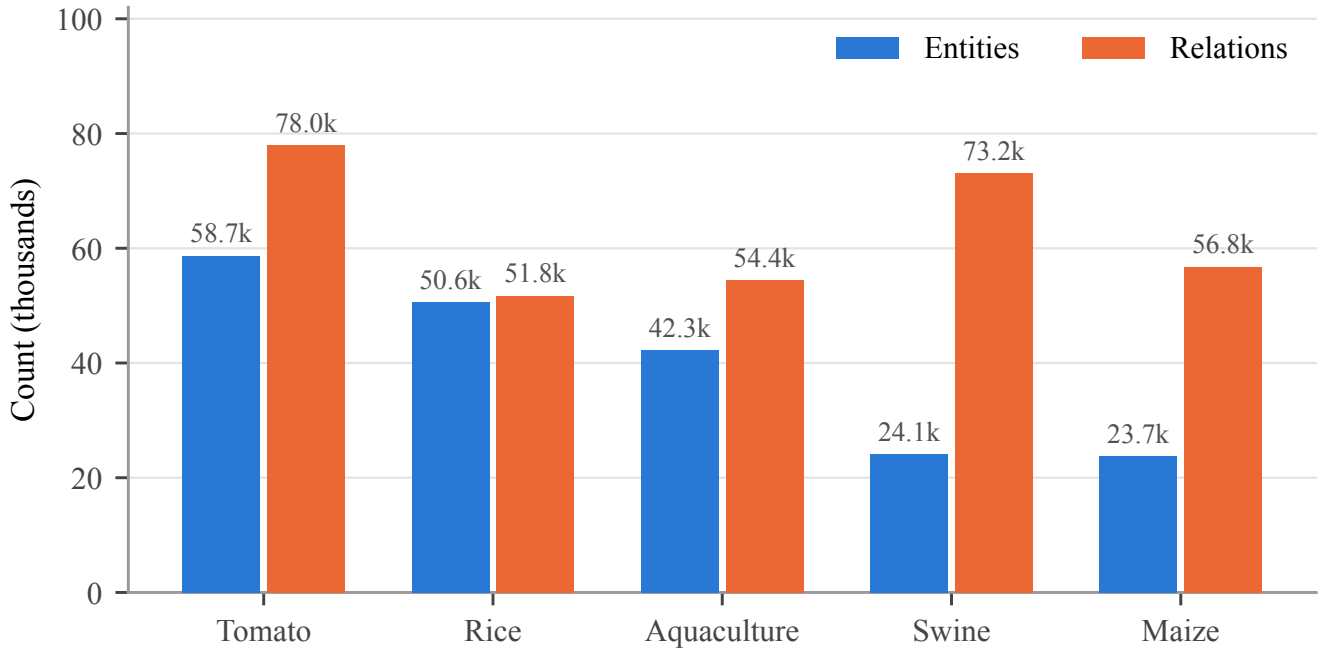



**Figure 7.** Entities and relations per sector, counted as in Table 4.

whereas the other three record single pages. The relations without an evidence record, 25.1% of the graph, all come from swine and maize, where they make up 63.0% and 50.1% of the triples. Without them, these sectors would contribute 25,361 and 26,366 relations, which is why we report them separately.

**Schemas and numeric facts.** The sector schemas diverged (Table 3), with two consequences. First, type assignments are unreliable where missing types were filled in by default. Whenever the model returned untyped fields, the aquaculture extraction typed the head as a farmed object and the tail as a quantitative value. In all, 78.2% of the aquaculture triples carry this pair of types, and 36,299 tails without a digit, such as genus names in taxonomy triples, are typed as quantitative values.

Second, whether a numeric fact keeps the name of the measured quantity depends on whether the schema adopted the triple-form constraint (Section 3.2). Table 6 counts each sector's value triples, whose tail begins with a digit, and the share with a generic relation. In rice, which adopted the constraint, only 2.0% of the value triples use a generic relation such as "value is" or "has". In swine, maize and aquaculture, 85.4% to 98.0% do, so the relation no longer says what was measured. When the head is the measured object, the indicator is lost, as in "a fish species, value is, about 4 cm". When the head is the indicator, the object is lost: the entity "body length" holds 397 "value is" relations from the aquaculture file, whose values, from different species, can no longer be told apart.

**Connectivity.** The graph is sparsely connected. It falls into 36,128 connected components, of which

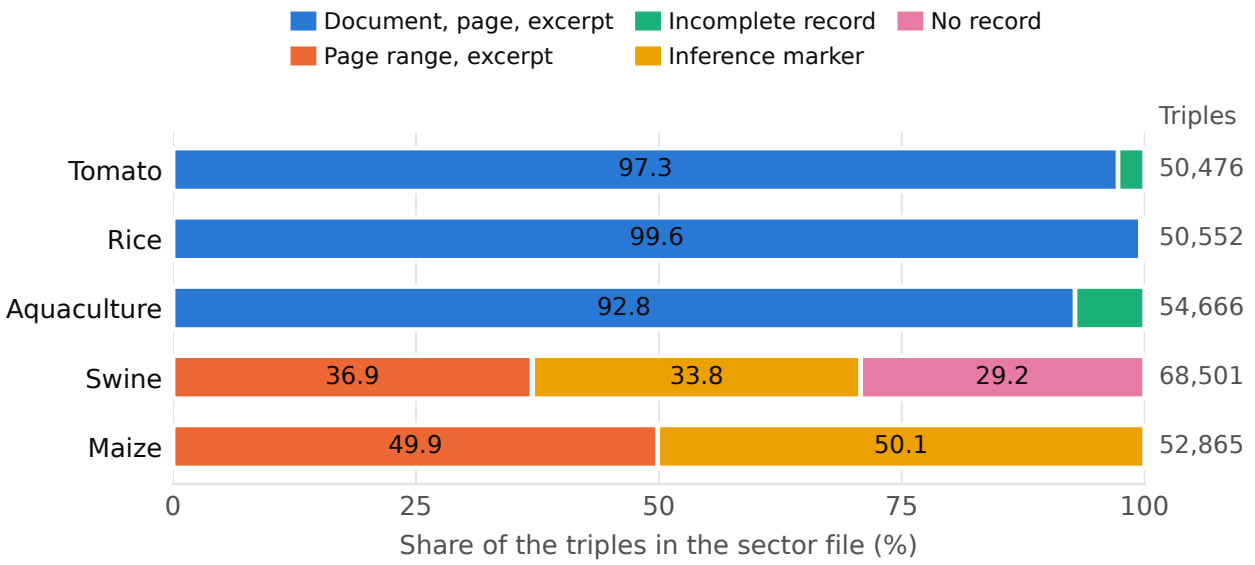



**Figure 8.** Provenance of the triples in each sector file.

**Table 5.** Relations of the graph by provenance. Each relation takes the most complete provenance of its triples, and a page without a digit counts as missing.

| Provenance | Relations | Share (%) |
|---|---|---|
| Evidence record: document, page and excerpt | 150,187 | 54.2 |
| Evidence record: page range and excerpt, no document title | 51,673 | 18.7 |
| Evidence record: page or excerpt missing or unknown | 5,561 | 2.0 |
| Inference marker | 49,615 | 17.9 |
| None | 20,017 | 7.2 |
| Total | 277,053 | 100.0 |

the largest holds 111,791 entities (56.1%) and 28,442 consist of two entities only. Table 6 gives each sector's share in the largest component. Most entities (74.8%) have only one relation. The mean degree is 2.78, and the maximum, 2,693, belongs to "tomato". A two-hop retrieval from a typical seed entity therefore reaches few entities beyond the high-degree entities of its sector, which bounds the context that the diagnosis workflow can retrieve (Section 3.6).

**Cross-media coverage.** The media collections hold 15,314 items (11,936 images and 3,378 videos) and cover the sectors unevenly (Figure 9). Swine has the highest coverage (23.4% of its file's entities with image links, 17.5% with clip links), although maize has more media for a similar number of entities. Unlike the other sectors, swine was matched with an earlier version of the grounding pipeline (Section 3.4). Aquaculture has the lowest image coverage (4.4%, and 4.0% for clips), because its collection is the smallest (1,193 items) while its entity set is the third largest. Tomato has the most entities with image links (8,700) but the lowest clip coverage (3.2%). These differences show where further media collection is most valuable.

Coverage is also bounded by the descriptors used for grounding. Only the 90,393 entities (45.3%) that head a triple receive one, so the 25,158 entities with image links are 27.8% of the entities that can be grounded at all. The descriptors are short for most entities, whose median number of relations is one, and very long for high-degree entities such as "tomato", whose descriptor lists 2,359 relations. Section 5.4 measures how many of these links the media index shows.

## 5 Evaluation

Whereas Section 4 described the graph, this section evaluates it and the system in two parts. The first part needs only the sector files and the import of Section 3.3. It measures how well triples are anchored in their excerpts (Section 5.1), what exact-name fusion joined and missed (Section 5.2) and how much control knowledge diagnosis can draw on (Section 5.3). The second part runs on the deployed system. It measures the media links that users see (Section 5.4), how the ingestion path handles new text (Section 5.5) and the effect of the graph on image diagnosis (Section 5.6). All measures were computed automatically from the sector files, match tables and system responses, so no manual rating enters the numbers.

**Table 6.** Value triples and connectivity by sector. Value triples have a tail that begins with a digit, and generic relations, such as "value is", name no indicator.

| Sector | Value triples | Share of triples (%) | With generic relation (%) | In largest component (%) |
|---|---|---|---|---|
| Tomato | 15,094 | 29.9 | 23.0 | 48.4 |
| Rice | 26,413 | 52.2 | 2.0 | 54.0 |
| Aquaculture | 9,062 | 16.6 | 96.7 | 65.3 |
| Swine | 8,964 | 13.1 | 98.0 | 59.9 |
| Maize | 10,322 | 19.5 | 85.4 | 59.1 |

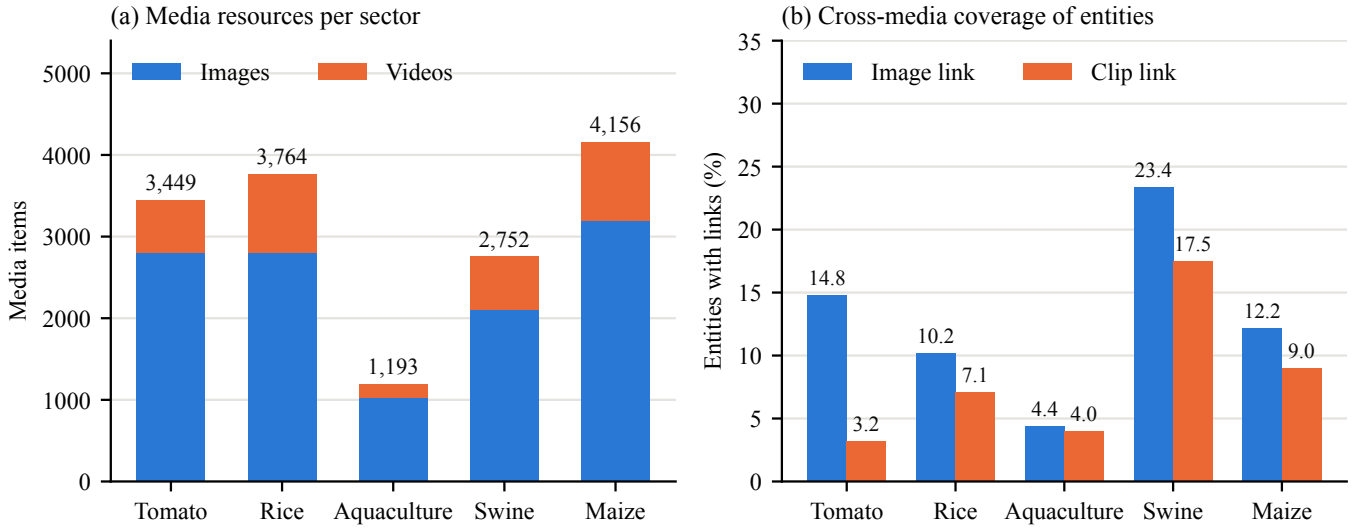



**Figure 9.** Cross-media resources. (a) Images and videos in each sector's collection. (b) Share of each sector file's entities with an image or clip link.

### 5.1 Evidence anchoring of extracted triples

An excerpt can support a triple only if it mentions what the triple is about. As an automatic lower bound on support, we checked whether the head and tail names of every extracted triple occur verbatim in its excerpt. We compared them after Unicode normalisation (NFKC), case folding and removal of whitespace, punctuation and symbols such as the degree sign. Table 7 reports the result for the 201,978 extracted triples that carry an excerpt. The model kept to the excerpt limit of 100 characters, or 50 for swine and maize, in all but 173 of these triples.

The excerpts anchor the tail of a triple much better than its head: the tail occurs in 70.2% of the excerpts, the head in only 38.7% and both in 26.0%. Values fare best. The value of 87.5% of the value triples occurs verbatim, which indicates that the model copied quantities rather than generating them. Rice has the highest share of triples with both names in the excerpt (41.5%) and tomato the lowest (15.0%).

We inspected sampled triples to see why the head is often missing. The excerpt usually holds the sentence that states the tail, whereas a heading or an earlier sentence of the window names the head. The model then composed a descriptive head, such as "autumn open-field tomato" or "tomato greenhouse cultivation". In the 17.2% of triples with neither name in the excerpt, the model had typically reformulated both names, for instance by reordering a planting density and its unit. In other cases, the excerpt was in traditional characters or English, while the names were in simplified Chinese. The check therefore underestimates support, but it locates the triples that only the full page can verify.

### 5.2 Exact-name fusion across sectors

The import of the five sector files is a sequence of incremental updates with the merge semantics of the ingestion path, so it shows at scale what exact-name fusion does. Table 8 follows the graph through the five imports. Each import after the first merged 1,012 to 4,111 of its entity names into existing entities, and these merges produced the 37,146 cross-sector relations. The relations themselves hardly overlap: of the 277,060

**Table 7.** Anchoring of extracted triples in their excerpts. Each column gives the share of triples, or of value triples in the last column, whose named part occurs verbatim in the excerpt.

| Sector | Triples | Head (%) | Tail (%) | Both (%) | Neither (%) | Value (%) |
|---|---|---|---|---|---|---|
| Tomato | 49,099 | 26.5 | 60.2 | 15.0 | 28.3 | 84.9 |
| Rice | 50,335 | 50.1 | 82.5 | 41.5 | 8.9 | 87.6 |
| Aquaculture | 50,864 | 29.6 | 67.6 | 16.7 | 19.4 | 90.1 |
| Swine | 25,307 | 46.5 | 75.9 | 32.8 | 10.4 | 92.0 |
| Maize | 26,373 | 49.7 | 64.5 | 28.5 | 14.3 | 84.2 |
| All | 201,978 | 38.7 | 70.2 | 26.0 | 17.2 | 87.5 |

**Table 8.** Growth of the graph over the five imports. Parentheses give the value-like names among the merged entities, and a dash marks counts that do not apply to the first import.

| | | | Entities | | Relations | | | Graph after import | |
|---|---|---|---|---|---|---|---|---|---|
| | Sector | Triples | New | Merged | New | Existing | Cross-sector | Entities | Relations |
| 1 | Tomato | 50,476 | 58,715 | – | 50,476 | – | – | 58,715 | 50,476 |
| 2 | Swine | 68,501 | 24,077 | 1,012 (799) | 68,501 | 0 | 4,638 | 82,792 | 118,977 |
| 3 | Maize | 52,865 | 23,714 | 2,358 (1,343) | 52,858 | 7 | 10,784 | 106,506 | 171,835 |
| 4 | Rice | 50,552 | 50,621 | 4,111 (2,366) | 50,552 | 0 | 13,464 | 157,127 | 222,387 |
| 5 | Aquaculture | 54,666 | 42,291 | 1,907 (1,129) | 54,666 | 0 | 8,260 | 199,418 | 277,053 |

triples, only 7 repeated an earlier sector's relation. Even when the relation label is ignored, only 76 triples of a later file joined a head and a tail that an earlier file had already joined. The sectors share entity names, such as pests, drugs and clinical signs, but contribute complementary facts.

The shared names, however, are mostly values. Between 57.0% and 79.0% of the names merged by each import are value-like names, which begin with a digit, such as "10%" or "7 to 10 days". Across the graph, value-like names make up 3,808 of the 6,805 shared names (56.0%), and 18,290 of the cross-sector relations (49.2%) touch an entity with such a name. An entity with a value-like name joins unrelated facts, as when a 10% dosage in swine and a 10% fruit-set rate in tomato meet in it.

Exact-name fusion also errs in the other direction: variant names of one concept stay apart. We searched each sector's names that are not value-like for three patterns, with at least two characters in the shorter name, and found 2,250 variant pairs involving 4,274 entities. In 566 pairs, the names differ only in punctuation, character width or spacing. In 1,407 pairs, one name adds the sector name as a prefix, as in "tomato top-cracked fruit" and "top-cracked fruit". The remaining 277 differ by the suffix for disease or syndrome, as in "magnesium deficiency syndrome" and "magnesium deficiency". This suffix pattern also pairs 38 viruses with the diseases they cause, so such patterns can propose merges but not decide them. The media index of Section 3.4 deletes the same characters in media lookup, so such pairs can share media while staying separate in the graph.

Exact-name fusion also leaves the type and sector of a merged entity unreconciled. First, 5,402 entities (2.7%) received more than one entity type, because two sectors typed the same concept differently or two

**Table 9.** Disease entities with control, cause and symptom relations, classified by the terms of the Stage 3 relation prior. The last column gives the share with only one relation.

| Sector | Diseases | Control (%) | Cause (%) | Symptom (%) | Any category (%) | One relation only (%) |
|---|---|---|---|---|---|---|
| Tomato | 771 | 20.0 | 10.0 | 38.8 | 50.2 | 64.1 |
| Rice | 175 | 17.1 | 16.6 | 25.7 | 45.1 | 53.1 |
| Aquaculture | 264 | 15.2 | 0.0 | 33.3 | 45.5 | 43.6 |
| Swine | 492 | 37.0 | 27.4 | 23.8 | 58.9 | 44.7 |
| Maize | 453 | 25.4 | 0.2 | 44.8 | 69.3 | 24.3 |
| All | 2,155 | 24.2 | 11.2 | 34.9 | 55.2 | 47.9 |

concepts share a name. Second, each entity keeps the sector of the file that introduced it, so the per-sector counts of Table 4 depend on the import order. Every name that tomato shares with another sector, for instance, is attributed to tomato.

### 5.3 Control knowledge available to diagnosis

The diagnosis workflow is meant to draw its recommendations from the graph wherever the graph holds relevant knowledge, and to fall back to general experience when the graph holds no control relation (Section 3.6). How often this fallback is needed depends on how many diseases have control relations. We took as disease entities the names of at least three characters that end in the character for disease and do not begin with a digit, and classified their relations with the terms of the relation prior of Stage 3 (Table 9).

Control knowledge is sparse. Only 24.2% of the 2,155 disease entities have a control relation, 11.2% a cause relation and 34.9% a symptom relation, and 47.9% have only one relation. Swine has the best control coverage (37.0% of its diseases), and aquaculture (15.2%) and rice (17.1%) the worst. For about three in four disease entities, the conclusion subgraph can offer no control measure, so the fallback to general experience is the normal case. Cause relations are nearly absent in maize and aquaculture for another reason: their closed label sets express causation in words that the relation prior does not recognise, such as "harms or causes disease" and "pathogen is". A relation prior written for the labels of one sector misses knowledge that other sectors express with other labels.

### 5.4 Media links seen by users

Section 4 reports how many entities have links, but the media index shows a link without lexical support only when its folder label matches the entity (Section 3.4). This experiment measured how many links are shown to the user and how they spread over the media items. We applied the indexing rule of the web system to the five match tables and recorded, for each link, whether the media index shows it for its own entity. Table 10 reports the results per sector, where lexical support means $s \geq 10$. One tomato link is not indexed because its file no longer exists.

The links reached few media items and concentrated on a small number of them. Of the 15,314 items, only 1,851 (12.1%) were linked to any entity, namely 14.2% of the images and 4.7% of the videos. Within each sector, the ten most linked items held 51.3% to 63.7% of the links. Across all sectors, 104 items were each linked to at least 100 entities; we call them hub items. Together they held 41,967 of the 58,597 links (71.6%). The most linked item, a tomato photograph of injury by the herbicide 2,4-D (2,4-dichlorophenoxyacetic acid), was linked to 3,756 entities, or 23.3% of the tomato links. Hubness is a known property of nearest-neighbour retrieval in high-dimensional spaces [53]. The underlying similarities were also weak. In the four sectors matched with the current version of the grounding pipeline, links without lexical support had a median similarity of 0.46 to 0.48, slightly above the thresholds of 0.45 and 0.40. Most descriptors (68.0%) list a single relation and probably carry too little information to separate the items, so many entities share the same nearest items.

Lexical support was rare, and the media index showed few links for their own entity. Only 4.1% of the links had $s \geq 10$. Outside swine, this share ranged from 2.6% in maize to 12.9% in aquaculture, whose folder labels and file names are mostly species names that are also entities. The index showed only 4.6% of the links for their own entity and stored 55,888 of the remaining links under their folder label only. It held media under the names of 1,755 entities, whereas 25,158 entities had image links and 14,132 clip links. Other entities receive media only when their name matches a folder label at query time. The same rule hides hub items from the user. The index showed only 615 (1.5%) hub links for their own entity, and none of the 3,756 links of the tomato hub item. The conservative rule of Section 3.4 hides nearly all hub links (98.5%), but also nearly every link without lexical support.

Swine had no link with lexical support for a different reason: an earlier version of the grounding pipeline produced its match table. This table has no start-time column, so its 4,399 clip links play from the start of their videos. Its scores range from 0.38, below the thresholds, to 2.64, far below the prior of Eq. (2). The 59 swine links scoring above 1 match the entity name to a file name or folder label, exactly or approximately, such as "Hainan pig" to a photograph of the Huainan pig. The threshold of 10 nevertheless treats them as visual matches, and the index shows only 17 of them, those whose folder label matches the entity. Swine becomes comparable with the other sectors only after the current version of the pipeline is run on it.

### 5.5 Editor ingestion

Editors extend the graph through the ingestion path (Section 3.5). We tested whether this path attaches new knowledge to the existing graph and

**Table 10.** Media links in the match tables and the media index. Lexical support means a match score $s \geq 10$.

| Sector | Links | Media items linked (%) | Top-10 items (%) | With lexical support (%) | Shown for their entity (%) |
|---|---|---|---|---|---|
| Tomato | 16,103 | 9.0 | 56.6 | 4.3 | 4.9 |
| Rice | 13,734 | 15.7 | 51.3 | 6.2 | 6.2 |
| Aquaculture | 5,243 | 13.7 | 51.9 | 12.9 | 12.9 |
| Swine | 15,390 | 18.2 | 53.7 | 0.0 | 0.2 |
| Maize | 8,127 | 6.9 | 63.7 | 2.6 | 4.4 |
| All | 58,597 | 12.1 | 26.7 | 4.1 | 4.6 |

**Table 11.** Editor ingestion of ten passages per sector with the cap $K = 10$.

| Measure | Tomato | Rice | Aquaculture | Swine | Maize | Total |
|---|---|---|---|---|---|---|
| Extracted entities | 89 | 96 | 96 | 91 | 101 | 473 |
| Extracted relations | 73 | 81 | 70 | 73 | 80 | 377 |
| Entities merged into existing ones | 60 | 66 | 44 | 39 | 74 | 283 |
| New entities | 29 | 30 | 52 | 52 | 27 | 190 |
| New relations | 70 | 77 | 68 | 73 | 77 | 365 |
| Duplicate relations absorbed | 0 | 1 | 0 | 0 | 0 | 1 |
| Relations skipped (missing endpoint) | 3 | 3 | 2 | 0 | 3 | 11 |
| Isolated entities created | 2 | 1 | 0 | 0 | 0 | 3 |
| Extraction time per batch (s) | 28.4 | 35.9 | 32.9 | 32.9 | 37.1 | 33.4 |
| Moves of existing entities to the sector | 8 | 15 | 15 | 5 | 21 | 64 |

what else the merge changes. For each sector we collected ten passages of technical guidance that agricultural departments and extension agencies published between March 2025 and September 2026. The documents include control plans, pest and disease forecasts, and guidance on heat, typhoons and floods, and none is part of the corpus. Each passage is continuous, unedited and 305 to 654 Chinese characters long. The 50 passages come from 38 documents, at most two from any one. Per sector, four passages concern pests and diseases, three cultivation or husbandry, two disasters and the environment, and one varieties or other topics.

We ingested the passages in the order of Table 11 with the default cap of $K = 10$ and confirmed every batch. To leave the deployed database untouched, we used a separate copy of the graph, built from the sector files with the counts of Section 4. The runs called the extraction and merge functions of the web system directly. Before each merge, we classified every extracted entity and relation against the graph. After the merge, the change in counts agreed with this classification for all 50 batches.

The path added knowledge in small batches that the merge accepted almost entirely (Table 11). The 50 batches extracted 473 entities and 377 relations, or 7.5 relations per batch, and no batch exceeded the cap. The merge kept 365 relations, absorbed one duplicate and skipped 11. Each skipped relation had an endpoint that the model had not listed as an entity and that the graph lacked. Only three new entities remained isolated. Extraction took 33.4 s per batch on average, so the editor’s cost lies in checking the preview, not in waiting.

The new knowledge attached to the existing graph. Of the 473 extracted entities, 283 (59.8%) merged into existing entities. Most were crops, crop types, diseases and pests such as tomato, early rice, sheath blight and the fall armyworm (*Spodoptera frugiperda*), and only two were value-like. Of the 365 new relations, 305 (83.6%) touched at least one pre-existing entity. Aquaculture and swine contributed the most new entities, 52 each. Their passages name current drugs, pathogen variants, devices and management schemes absent from the graph, such as a variant strain of porcine reproductive and respiratory syndrome virus (PRRSV) or batch breeding of sows.

The merge integrated entity names but neither relation labels nor sectors. Of the 365 new relations, 228 (62.5%) carried one of 183 relation types that the graph did not hold before these runs, such as “recommended treatment” or “follow-up disinfection drug”. They arise because the extraction prompt suggests entity types but leaves relation labels free. Merging also

moved 58 existing entities to the sector of the batch, 64 times in all. The tomato import, for example, had attributed the entity maize to tomato. A passage on pig production in cold weather moved it to swine, and a maize passage then moved it to maize. Sheath blight likewise moved from maize to rice and back. The sector attribute therefore records the last batch that mentioned an entity, not the sector where it originated.

### 5.6 Graph-assisted image diagnosis: a case study

This case study asked whether the graph makes the diagnosis workflow (Section 3.6) more accurate and how the workflow attributes advice to the graph. One hundred photographs are too few for firm statistical conclusions, so we report the comparison as a case study and analyse its errors in detail. We compared the deployed workflow (S2) with the model alone (S1) on 100 labelled photographs from the media collections, 20 per sector, using the folder label as the gold label. Drawing the photographs from the collections does not reveal the answer, because the workflow consults the media index only after the answer, to show reference media. We used folders whose labels state a disease, pest, disorder, species or breed, and for rice also a field operation. Their labels were drawn round-robin from a seeded shuffle, giving 5 (maize) to 19 (tomato) labels per sector.

We skipped a photograph whose file name named a label other than its folder label or that duplicated one already drawn. We also skipped it when optical character recognition found a label of the sector in the image, or when the labelled object or operation was not visible on inspection. These rules exposed labelling errors in the collections. All 200 photographs in the folders for tomato spotted wilt are tomato mosaic virus images from a public dataset. Two rice folders for brown spot hold images of wheat brown rust, and the Huai pig photographs are video frames captioned with the breed name. The rules therefore removed tomato spotted wilt and the Huai pig. We further dropped three rice operations rarely shown by their folders, namely chemical control, fertilisation and intermittent irrigation, and the Bamei pig, whose frames carry the breed name in a caption that recognition missed. The final set holds 30 disease, 13 pest, 10 disorder, 20 species, 11 breed and 16 operation cases.

Each photograph came with a question that names only the task, asking for the disease, pest or abnormality, the operation, or the species or breed shown. Both configurations used the same model and question. We obtained S1 by passing an empty entity list to Stage 2.

**Table 12.** Image diagnosis of 100 labelled photographs, 20 per sector. Parentheses give the number of cases per task, and a dash marks a value that does not apply.

| Measure | Model alone (S1) | Deployed workflow (S2) |
|---|---|---|
| Accuracy (%) | 30.0 | 24.0 |
| Fallback to general experience disclosed (%) | 94.0 | 34.0 |
| Advice labelled as recommended by the graph (%) | 0.0 | 88.0 |
| Conclusion subgraph contains the gold entity (%) | 21.0 | 22.0 |
| Entities in the Stage 2 prompt (median) | 0 | 78.5 |
| Value-like names among them (%) | – | 13.0 |
| Latency per case (s) | 20.0 | 30.5 |
| *Accuracy by task* (%) | | |
| Disease (30) | 20.0 | 33.3 |
| Disorder (10) | 10.0 | 20.0 |
| Pest (13) | 46.2 | 30.8 |
| Species (20) | 60.0 | 35.0 |
| Breed (11) | 0.0 | 0.0 |
| Operation (16) | 31.3 | 6.3 |

We ran S2 on the deployed database by calling the workflow endpoint by endpoint, with the order and parameters of the web interface. Stage 3 ran on the answers of both configurations.

A case counted as correct when the gold label, its core term or a listed synonym occurred in the conclusion section of the answer after normalisation. An operation counted as correct when its key terms occurred, for example “unmanned” and “harvest”. Before matching, we removed the names of other labels that contain the gold label, such as African swine fever for swine fever. We fixed the synonyms with the test set, and the operation key terms after a two-case pilot and before the main run. Besides accuracy, we recorded how the answers attributed their advice, whether the conclusion subgraph contained the gold entity, the entities in the Stage 2 prompt and the latency (Table 12).

In this case study, the graph did not make the diagnosis more accurate. S2 answered 24 cases correctly, against 30 for S1. Only S1 was right in 14 cases and only S2 in 8, and the difference was not significant (exact McNemar test, $p = 0.29$). The effect depended on the task (Table 12, lower part). S2 was more accurate on diseases and disorders but less so on pests, fish species and rice operations, and neither configuration identified a pig breed. All accuracies are lower bounds, because the matching is strict and the folder labels

contain errors. An answer of "bacterial leaf spot" for bacterial spot, for instance, counts as wrong. Eight runs returned an empty answer, six of S1 and two of S2, and counted as wrong. S2 was also slower, taking 30.5 s per case against 20.0 s for S1. Stage 3 added 12.9 s for S2 and 17.7 s for S1.

The answers of S2 followed the initial identification of Stage 1. In 78 of the 100 cases, S2 concluded with it. When the initial identification contained the gold label (20 of the 100 cases), S2 was correct in 19 and S1 in 14. When it did not (80 cases), S2 was correct in only 5 and S1 in 16. The retrieved entities confirmed the initial identification rather than correcting it, and a wrong initial identification led S2 away from the answer that the model alone would have given. For example, S2 took a black cutworm (*Agrotis ipsilon*) moth for *Spodoptera litura* and spider-mite (*Tetranychus* spp.) damage on tomato for powdery mildew, both terms proposed by Stage 1. For the rice operations, S2 also adopted the generic wording of its initial identification, such as "mechanised harvesting", where S1 named the combine harvester.

To see how the answers failed, we sorted the 70 wrong answers of S1 and the 76 of S2 into five categories with fixed rules on the conclusion text. An answer was empty, named the crop or animal of another sector, named another label of the same sector in the test set, named only the broader category of the gold label, or named a disease, pest, disorder, species, breed or operation outside the test labels. For S2, these categories held 2, 4, 14, 11 and 45 answers, and for S1, 6, 1, 15, 10 and 38. Most errors were plausible but wrong identifications rather than failures to answer, and their form depended on the task. S2 named a similar species for 12 of its 13 wrong fish answers, such as blue mackerel scad for mackerel or common carp (*Cyprinus carpio*) for grass carp (*Ctenopharyngodon idella*). For 9 of its 11 wrong breed answers, it named another breed, such as the Taihu pig for the Dingyuan pig, or only a local black pig. For 12 of its 15 wrong rice operations, it named another operation of the test set or only a broad term such as mechanised harvesting. For 4 of its 9 wrong pest answers, it named the genus, cutworm, instead of the black or yellow cutworm. Four S2 answers placed the photograph in another sector, for instance maize grain damage taken for blossom-end rot of tomato. Of the 76 wrong answers of S2, 58 contained a Stage 1 term, so most S2 errors carried the initial identification through to the conclusion. The gold labels themselves are folder labels, and some, such as grain damage or a local pig breed, are hard to recognise from a single photograph.

S2 applied the second and third decision rules, but its attributions did not reflect what the graph contains.

Operation recognition: mechanised harvesting of rice. Left: input image and Stages 1 and 2; centre: conclusion, visual evidence and recommendation; right: conclusion subgraph and reference media.




**Figure 10(a).** Three cases from the deployed diagnosis interface. English translations are overlaid on the interface labels and the graph; the answers remain in Chinese and are translated in Section 5.7.

Species identification: common goldfish, with the disclosure that feeding advice rests on general experience.

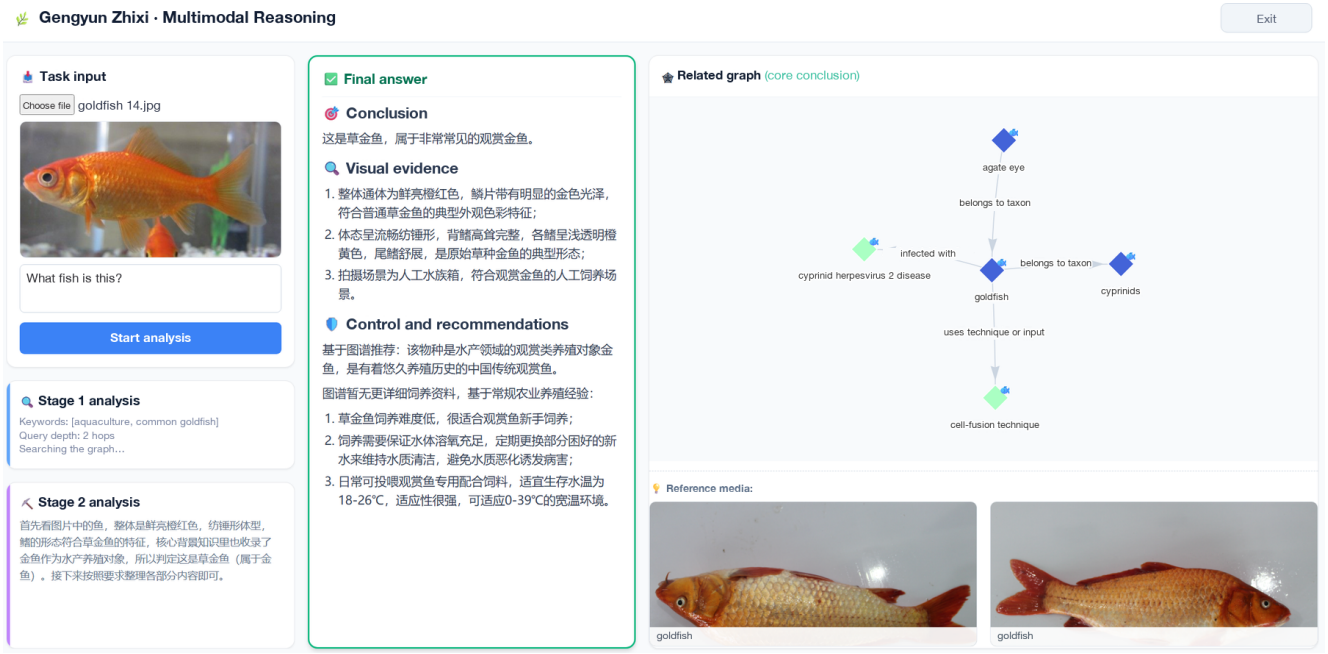



**Figure 10(b).** Three cases from the deployed diagnosis interface (continued).

S1, which saw no graph, disclosed a fallback to general experience in all 94 non-empty answers. S2 labelled its advice as recommended by the graph in 88% of the cases and disclosed a fallback in only 34%. Yet the model of S2 saw only entity names, a median of 78.5 per case, of which 13.0% were value-like. The attribution "recommended by the graph" thus signals that a graph context was present, not that the graph contains the advice. The conclusion subgraph contained the gold entity in 21% (S1) and 22% (S2) of the cases, so the evidence shown after the answer depends on the answer rather than on the configuration.

### 5.7 Case studies

Single cases show what the user sees in the interface. Figures 10(a)–10(c) show three cases from the deployed system, each of a different task. They illustrate both roles of the graph, as context during reasoning and as evidence after it. In each case, we compare the advice in the answer with what the conclusion subgraph records.

In the first case (Figure 10(a)), the question asks which operation is performed on rice. Stage 1 proposed the keywords "rice" and "mechanised harvesting". The conclusion identifies mechanised harvesting of mature rice. Its visual evidence is a contiguous mature paddy that has partly been cut, a combine harvester driven by an operator, and a scene consistent with harvest operations. The conclusion subgraph centres on the rice combine harvester, with a named harvester model and two fitted devices that measure yield and grain moisture. The reference media show a semi-mechanised and an unmanned harvesting scene. The recommendation on cutting and placing practices that reduce grain loss is labelled as recommended by the graph, but no relation of the conclusion subgraph states it. This label reflects the model's own attribution (Section 7).

In the second case (Figure 10(b)), the question asks which fish is shown. Stage 1 proposed "aquaculture" and "common goldfish". The conclusion identifies a common goldfish (*Carassius auratus*) from its colour, body shape and fins. The conclusion subgraph places goldfish under the cyprinids and links it to the variety "agate eye", to the disease caused by cyprinid herpesvirus 2 (CyHV-2) and to a cell-fusion technique. The answer attributes the species facts to the graph. It states that its feeding recommendations rest on general aquaculture experience, because the graph holds no feeding relation for goldfish, and no relation in the subgraph concerns feeding. Section 5.3 shows that this situation is common.

In the third case (Figure 10(c)), the question asks what is wrong with a tomato flower. Stage 1 proposed "tomato" and "deformed flower". The conclusion identifies a deformed (fasciated) flower, a non-infectious abnormality of flower development, from fused flowers. The answer states that the graph

Disorder diagnosis: tomato deformed flower, with the causes recorded in the subgraph and the disclosure for the control measures.

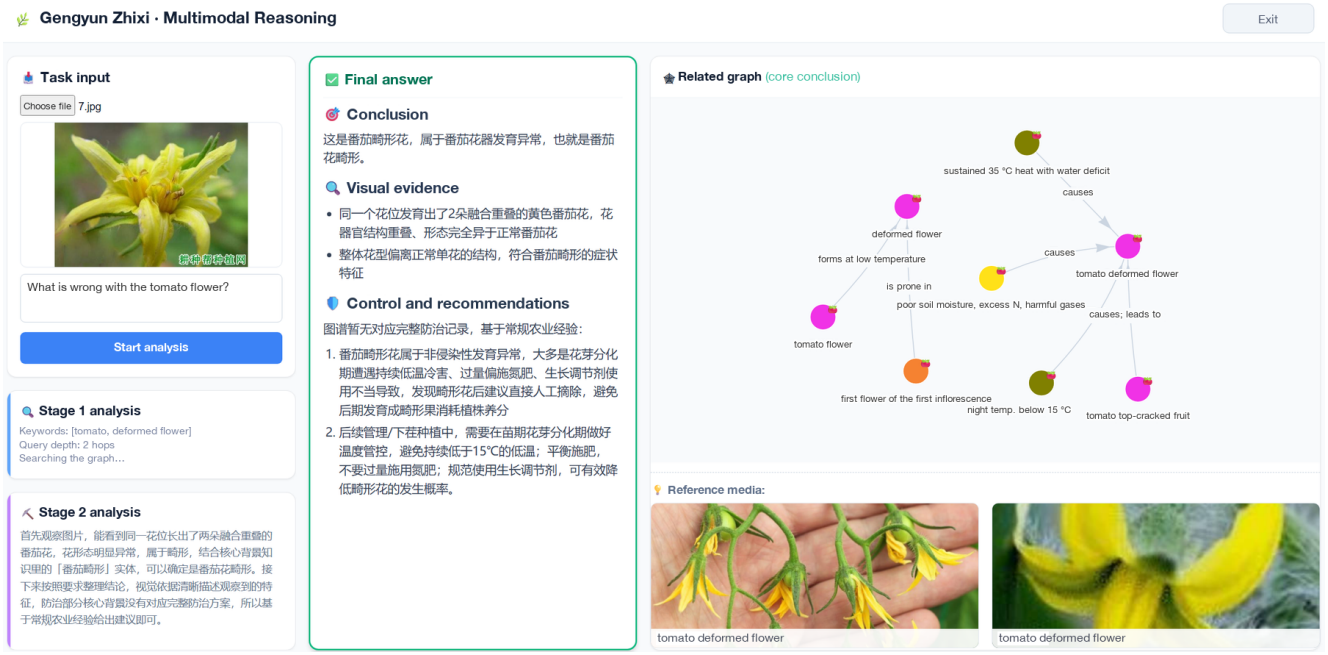



**Figure 10(c).** Three cases from the deployed diagnosis interface (continued).

lacks complete control measures and that its measures rest on general experience. It recommends removing deformed flowers and avoiding low temperature and excess nitrogen during flower-bud differentiation. The conclusion subgraph lets the user check this advice against the recorded causes, which are sustained temperatures of 35 °C with water deficit, night temperatures below 15 °C, and improper soil moisture with excess nitrogen and harmful gases. It also records that the first flower of the first inflorescence is the most susceptible and that deformed flowers lead to cracked fruit. The reference media show two documented deformed-flower images. The advice agrees with the low-temperature and nitrogen causes, although the model never saw these relations.

## 6 A Taxonomy of Failure Modes

Sections 4 and 5 report the failures of AgriEvoKG one measurement at a time. Table 13 gathers them into a taxonomy. Each row pairs a failure mode with its evidence, its root cause, a remedy and a test. The test is a measure that can be recomputed on the sector files, the match tables or the stored runs, so a remedy counts as effective only when its measure improves. The first five rows are failures of construction. They arise from extracting sector by sector and fusing by name, not from agricultural content, and we therefore expect them in other graphs that LLMs build from professional literature. The last three rows concern the deployed system and depend on its ingestion and diagnosis design.

The taxonomy separates symptoms from causes. Schema divergence and lost quantity names, for example, appear in different measures but share one cause, sector-by-sector schema design, and one remedy, a shared core schema with the triple-form constraint, addresses both. Fusion noise and ingestion drift likewise share a cause: merging by surface form without types or a shared label set. Every test needs only data that the system already produces, so each remedy can be checked by rerunning the measures of Sections 4 and 5.

## 7 Discussion

**Main findings.** We set out to build a cross-media knowledge graph for five agricultural sectors, to keep the evidence of each fact, to let editors update the graph, and to describe the structure and failures of an LLM-built graph of this size. AgriEvoKG meets the first three aims in part: 74.9% of the relations in its graph carry an evidence record, the entities are linked to images and video clips, and editors can add knowledge from new text. The fourth aim produced the main contribution of this paper, the taxonomy of Section 6.

**Why scale matters.** Most of these failures cannot be seen in the small samples on which construction

**Table 13.** Failure modes of AgriEvoKG with their root causes, remedies and tests. C, construction; D, deployment.

| Failure mode | Evidence | Root cause | Remedy | Test |
|---|---|---|---|---|
| C1 Schema divergence | Rice and tomato hold 23,727 and 19,139 relation labels, of which 79.0% and 83.1% occur once; the closed schemas hold 21 to 24 (Table 3) | Schemas written sector by sector, with free relation labels and no shared core | A shared core schema with sector extensions, fixed before extraction | Distinct relation labels per sector and the share used once |
| C2 Lost quantity names | 85.4% to 98.0% of the value triples of swine, maize and aquaculture use a generic relation, against 2.0% in rice (Table 6) | Schemas that attach values through generic relations such as "value is" | The triple-form constraint in every sector, so that the label names the indicator and its conditions | Share of value triples with a generic relation |
| C3 Anchoring asymmetry | Tails occur in 70.2% of the excerpts, heads in 38.7% and values in 87.5% (Table 7) | The excerpt quotes the sentence of the tail, while a heading or an earlier sentence names the head | Store the heading or sentence that names the head with the evidence record | Share of triples whose head occurs in their evidence |
| C4 Fusion noise | 6,805 shared names but 7 shared relations; 49.2% of the cross-sector relations touch value-like names; 2,250 variant pairs stay apart (Section 5.2) | Merging by surface name, without entity types or a filter for values | Canonicalisation that respects entity types, excludes value-like names and joins variants | Share of cross-sector relations through value-like names; variant pairs left apart |
| C5 Hub concentration | 104 hub items hold 71.6% of the 58,597 media links; the median similarity of links without lexical support is 0.46 to 0.48 (Section 5.4) | Short descriptors, 68.0% of which list one relation, and hubs in nearest-neighbour retrieval | Richer descriptors, and a link only if the entity also ranks among the item's nearest neighbours | Share of links held by hub items; top-10 share per sector |
| D1 Ingestion drift | 62.5% of the new relations carry new relation types; 58 entities changed sector 64 times (Table 11) | Free relation labels in the ingestion prompt; the sector is overwritten on merge | Map extracted labels to existing relation types; keep the sector of merged entities | Share of new relations with new types; sector changes per batch |
| D2 Sparse control knowledge | Only 24.2% of the disease entities have a control relation, and cause relations are nearly absent in maize and aquaculture (Table 9) | Little control knowledge extracted; a relation prior written for the labels of one sector | Targeted extraction of control knowledge; a relation prior defined on shared relation types | Share of disease entities with a control relation |
| D3 Confirmed initial identification | S2 concluded with the Stage 1 term in 78 of 100 cases and was correct in only 5 of the 80 cases whose term missed the gold label (Section 5.6) | Retrieval starts from one term and passes only entity names | Retrieve for several alternative diagnoses and pass the relations that discriminate between them | Accuracy on cases whose initial identification is wrong |

methods are usually judged [20, 23], a limitation that Paulheim [54] already noted for partial gold standards in knowledge graph evaluation. A sample-based evaluation asks whether each triple is correct, but four of the five construction failures are properties of the graph as a whole. A triple with a label used only once, or with the generic relation "value is", can be correct and well supported. The failure appears only when 79.0% of the 23,727 rice labels occur once, or when 397 "value is" relations from different species meet at "body length". A merge of the name "10%" is invisible within one sector file and appears only after fusion, when 18,290 cross-sector relations touch value-like names while the five sectors share only 7 relations. A single media link may look plausible, but 104 items hold 71.6% of the 58,597 links. Only the anchoring asymmetry shows at the level of single triples, and even this one requires heads and tails to be checked separately, which a judgement of whole triples does not do. Profiling the whole graph with simple automatic measures, as in Sections 4 and 5, should therefore accompany sample-based precision whenever an LLM builds a graph of hundreds of thousands of facts.

**Inferred relations and provenance.** A quarter of the relations come from graph inference or lack an

evidence record. These relations raise the counts of swine and maize but cannot be checked against text, and the sector files record their inference markers but not the rules that produced them. The inference marker allowed our analysis to treat inferred relations apart from extracted ones. A release of the graph should keep it, and the interface should display it, which it does not do yet.

**Context and evidence in diagnosis.** A graph can give the model context before the answer and give the user evidence after it. The diagnosis case study is too small for firm conclusions, but it shows how a graph context can fail. The entity names kept the prompt short but did not improve identification. They confirmed the initial identification of Stage 1, which helped when it was right and misled the model when it was wrong. In our 100 cases, the workflow was not more accurate than the model alone (24% against 30%), despite the vision-dominant decision rule. A graph context can correct an initial identification only if it offers alternatives, so retrieval should start from several alternative diagnoses and pass the relations that discriminate between them. As evidence, the graph is shown to the user only after the answer, as a conclusion subgraph without source pages. The label "recommended by the graph", which appeared in 88% of the cases, remains the model's own attribution. The graph also holds control relations for only 24.2% of its disease entities, and the relation prior of Stage 3 does not recognise how the maize and aquaculture schemas express causation. The fallback to general experience is therefore the normal case, and its disclosure tells the user that the advice does not come from the graph.

**Limitations.** We did not judge the correctness of relations or media links manually. The anchoring check of Section 5.1 is a lower bound on support rather than a measure of correctness, and the measurements of Sections 5.4 to 5.6 count behaviour rather than quality. Exact-name fusion merges homonyms and value-like names and leaves synonyms apart, and the single sector attribute depends on the import order. Relation labels were not unified after extraction either. In rice and tomato, whose prompts left labels free, 55.8% and 37.2% of the distinct relation labels occur only in value triples, because each such label states the conditions of its value. These labels were not mapped to a common set of value relations, which explains much of the gap between these two sectors and the other three in Table 3.

Provenance is incomplete. The swine and maize files give a placeholder instead of the document title, although most of their triples can be traced to a document, and 25.1% of the relations have no evidence record. The deployed database holds excerpts only for tomato, swine and maize, so rice and aquaculture relations can be checked against their excerpts only in the sector files. In grounding, the lexical prior reranks only the visually nearest candidates, and the media index hides links without lexical support unless their folder label matches. Entities that occur only as tails, 54.7% of all entities, have no descriptor and are not grounded, and the swine match table was produced with an earlier version of the grounding pipeline and records no clip start times.

The deployed system has further limits. Relations added by editors carry no document or page, use generic entity types and free relation labels, and are written item by item rather than in one transaction. They also overwrite the sector of an existing entity, and the ingestion path does not detect conflicting values. The interface does not yet display the source of a relation, and the single shared editor account leaves edits unattributed. During development we observed that blurred or damaged photographs were diagnosed less reliably than sharp ones, and the system does not compensate for this. The diagnosis test used 100 photographs from the media collections, took their folder labels, which contain errors, as gold labels and scored the answers by strict string matching. Its accuracies are therefore lower bounds, only large differences between the configurations would be significant, and we treat it as a case study rather than a benchmark. Finally, the corpus and the interface are in Chinese.

**Future work.** The remedies of Table 13 define the next steps, and their tests show whether each step works. Beyond them, the ingestion path should record the editor, the source text and the time of every relation and commit each batch in one transaction, the interface should display the document, page and excerpt of every relation, and the current version of the grounding pipeline should be rerun on swine. A larger diagnosis test with independently verified labels is needed before the effect of graph context can be judged. We also plan to derive an explicit rule base of agronomic conditions and actions from the control and condition relations of the graph, to train the retrieval policy of the diagnosis workflow with reinforcement learning in the manner of Graph-R1 [28], and to extend the corpus to English sources so that the approach can be evaluated across languages.

## 8 Conclusion

We have built AgriEvoKG, a system that constructs a cross-media knowledge graph for swine, aquaculture, tomato, maize and rice production and serves it on the web for browsing, editor ingestion and image diagnosis. The graph's 199,418 entities and 277,053 relations come from LLM extraction over the professional literature of each sector and, for swine and maize, from graph inference. Of the relations, 74.9% carry an evidence record, and the entities are linked to images and 10 s video clips. Our analysis identified five failure modes of LLM-based construction: sector schemas diverged; numeric facts lost the name of the measured quantity or object where no constraint required it; evidence records anchored values far better than head entities; exact-name fusion merged entity names but almost no relations; and media links concentrated on a small number of hub items. Three further failure modes concern deployment. Editor ingestion introduced new relation types and overwrote sectors, the graph holds control relations for only 24.2% of its disease entities, and in a case study of 100 photographs graph context confirmed the model's initial identification rather than correcting it. Table 13 pairs each failure mode with its root cause, a remedy and a test, and most of the failures become visible only when the whole graph is profiled. We expect the same checks to apply to other fields whose knowledge is recorded in books, images and video.

## Data Availability Statement

The sector files, match tables, ingestion passages and diagnosis runs analysed in this study are available from the corresponding author on reasonable request.

## Funding

This work was supported by the National Major Agricultural Science and Technology Project under Grant 0304.

## Conflicts of Interest

The authors declare no conflicts of interest.

## AI Use Statement

Doubao Seed 1.6 Flash, Seed 2.0 Mini and Seed 2.0 Lite (ByteDance) were used to extract triples from the source literature. A Doubao vision-language model (ByteDance) supported editor ingestion and image diagnosis. Qwen3-VL-Embedding (Alibaba) provided embeddings for entity-to-image and entity-to-video matching. The extracted triples and diagnostic responses are model outputs; the documents, photographs and videos are source inputs. The production procedures, evaluation and limitations are described in Sections 3-7. Doubao (ByteDance) was also used to help check and revise the language of the manuscript. All AI tools were used in accordance with their respective terms of use. The authors remain responsible for the manuscript's content, originality and integrity.

## Ethical Approval and Consent to Participate

Not applicable. The study involved no human participants and no experiments on live animals; ethics review was therefore not required. The animal photographs and videos were taken from existing media collections; the ingestion test used published technical guidance documents.

## References

[1] Drury, B., Fernandes, R., Moura, M.-F., & de Andrade Lopes, A. (2019). A survey of semantic web technology for agriculture. *Information Processing in Agriculture, 6*(4), 487-501. [CrossRef]

[2] Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys (Csur), 54*(4), 1-37. [CrossRef]

[3] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems, 33*(2), 494-514. [CrossRef]

[4] Pan, J. Z., Razniewski, S., Kalo, J. C., Singhania, S., Chen, J., Dietze, S., ... & Graux, D. (2023). Large language models and knowledge graphs: Opportunities and challenges. *Transactions on Graph Data and Knowledge, 1*(1), 2:1-2:38. [CrossRef]

[5] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys, 55*(12), 1-38. [CrossRef]

[6] Liu, Y., Li, H., Garcia-Duran, A., Niepert, M., Onoro-Rubio, D., & Rosenblum, D. S. (2019). MMKG: Multi-modal knowledge graphs. In *The Semantic Web: ESWC 2019* (Lecture Notes in Computer Science, Vol. 11503, pp. 459-474). Springer. [CrossRef]

[7] Zhu, X., Li, Z., Wang, X., Jiang, X., Sun, P., Wang, X., ... & Yuan, N. J. (2024). Multi-modal knowledge graph construction and application: A survey. *IEEE Transactions on Knowledge and Data Engineering, 36*(2), 715-735. [CrossRef]

[8] Wang, M., Wang, H., Qi, G., & Zheng, Q. (2020). Richpedia: A large-scale, comprehensive multi-modal knowledge graph. *Big Data Research, 22*, 100159. [CrossRef]

[9] Alberts, H., Huang, N., Deshpande, Y., Liu, Y., Cho, K., Vania, C., & Calixto, I. (2021, November). Visualsem: a high-quality knowledge graph for vision and language. In *Proceedings of the 1st workshop on multilingual representation learning* (pp. 138-152). [CrossRef]

[10] Wang, X., Meng, B., Chen, H., Meng, Y., Lv, K., & Zhu, W. (2023). TIVA-KG: A multimodal knowledge graph with text, image, video and audio. In *Proceedings of the 31st ACM International Conference on Multimedia* (pp. 2391-2399). [CrossRef]

[11] Liu, J., Meng, S., Gao, Y., Mao, S., Cai, P., Yan, G., ... & Shi, B. (2025, October). Aligning vision to language: Annotation-free multimodal knowledge graph construction for enhanced llms reasoning. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 981-992). IEEE. [CrossRef]

[12] Wan, X., & Yu, H. (2025). Mmgraphrag: Bridging vision and language with interpretable multimodal knowledge graphs. *arXiv preprint arXiv:2507.20804*. [CrossRef]

[13] Park, H., Seo, J., Jang, M., Park, H., Baek, H. D., Chang, G., ... & Kim, S. (2025). VAT-KG: Knowledge-Intensive Multimodal Knowledge Graph Dataset for Retrieval-Augmented Generation. *arXiv preprint arXiv:2506.21556*. [CrossRef]

[14] Bu, C., Chang, G., Chen, Z., Dang, C., Wu, Z., He, Y., & Wu, X. (2025, July). Query-driven multimodal GraphRAG: Dynamic local knowledge graph construction for online reasoning. In *Findings of the association for computational linguistics: ACL 2025* (pp. 21360-21380). [CrossRef]

[15] Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., & Keizer, J. (2013). The AGROVOC linked dataset. *Semantic Web, 4*(3), 341-348. [CrossRef]

[16] Chen, Y., Kuang, J., Cheng, D., Zheng, J., Gao, M., & Zhou, A. (2019, April). AgriKG: An agricultural knowledge graph and its applications. In *International conference on database systems for advanced applications* (pp. 533-537). Cham: Springer International Publishing. [CrossRef]

[17] Lim, S. C., Itharajula, M., Møller, M. H., Sunil, R. S., Fo, K., Chuah, Y. S., ... & Mutwil, M. (2025). PlantConnectome: A knowledge graph database encompassing >71,000 plant articles. *The Plant Cell, 37*(7), koaf169. [CrossRef]

[18] Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., ... & Larson, J. (2024). From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*. [CrossRef]

[19] Guo, Z., Xia, L., Yu, Y., Ao, T., & Huang, C. (2025, November). LightRAG: Simple and Fast Retrieval-Augmented Generation. In *EMNLP (Findings)* (pp. 10746-10761). [CrossRef]

[20] Zhang, B., & Soh, H. (2024, November). Extract, define, canonicalize: An llm-based framework for knowledge graph construction. In *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 9820-9836). [CrossRef]

[21] Lairgi, Y., Moncla, L., Cazabet, R., Benabdeslem, K., & Cléau, P. (2024, November). itext2kg: Incremental knowledge graphs construction using large language models. In *International Conference on Web Information Systems Engineering* (pp. 214-229). Singapore: Springer Nature Singapore. [CrossRef]

[22] Mo, B., Yu, K., Kazdan, J., Mpala, P., Yu, L., Kanatsoulis, C., & Koyejo, S. (2026). Kggen: Extracting knowledge graphs from plain text with language models. *Advances in Neural Information Processing Systems, 38*, 30092-30115. [CrossRef]

[23] Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2023). A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys, 56*(4), 1-62. [CrossRef]

[24] Trajanoska, M., Stojanov, R., & Trajanov, D. (2023). Enhancing knowledge graph construction using large language models. *arXiv preprint arXiv:2305.04676*. [CrossRef]

[25] Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., & Tiddia, S. G. (2023). Iterative zero-shot llm prompting for knowledge graph construction. *arXiv preprint arXiv:2307.01128*. [CrossRef]

[26] Mihindukulasooriya, N., Tiwari, S., Enguix, C. F., & Lata, K. (2023, October). Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text. In *International semantic web conference* (pp. 247-265). Cham: Springer Nature Switzerland. [CrossRef]

[27] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*. [CrossRef]

[28] Luo, H., E, H., Chen, G., Lin, Q., Guo, Y., Xu, F., ... & Luu, A. T. (2026). Graph-R1: Towards agentic GraphRAG framework via end-to-end reinforcement learning. In *Proceedings of the 43rd International Conference on Machine Learning* (PMLR 306, pp. 82220-82240). https://proceedings.mlr.press/v306/luo26c.html

[29] Peng, Y., Huang, X., & Zhao, Y. (2018). An overview of cross-media retrieval: Concepts, methodologies, benchmarks, and challenges. *IEEE Transactions on Circuits and Systems for Video Technology, 28*(9), 2372-2385. [CrossRef]

[30] Kim, J., Lee, G., Kim, T., & Shin, K. (2025, July). Kgmel: Knowledge graph-enhanced multimodal entity linking. In *Proceedings of the 48th international*

*ACM SIGIR conference on research and development in information retrieval* (pp. 3015-3019). [CrossRef]

[31] Chen, X., Zhang, N., Li, L., Deng, S., Tan, C., Xu, C., ... & Chen, H. (2022, July). Hybrid transformer with multi-level fusion for multimodal knowledge graph completion. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 904-915). [CrossRef]

[32] Li, M., Zhang, Y., Long, D., Chen, K., Song, S., Bai, S., ... & Lin, J. (2026). Qwen3-vl-embedding and qwen3-vl-reranker: A unified framework for state-of-the-art multimodal retrieval and ranking. *arXiv preprint arXiv:2601.04720*. [CrossRef]

[33] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 3580-3599. [CrossRef]

[34] Amaral, G., Rodrigues, O., & Simperl, E. (2024). ProVe: A pipeline for automated provenance verification of knowledge graphs against textual sources. *Semantic Web*, *15*(6), 2159-2192. [CrossRef]

[35] Ghanem, H., & Cruz, C. (2025). Fine-tuning or prompting on LLMs: evaluating knowledge graph construction task. *Frontiers in big Data*, *8*, 1505877. [CrossRef]

[36] Yang, S., Liu, Z., Mayer, W., Ding, N., Wang, Y., Huang, Y., ... & Feng, Z. (2024, November). Shizishangpt: An agricultural large language model integrating tools and resources. In *International Conference on Web Information Systems Engineering* (pp. 284-298). Singapore: Springer Nature Singapore. [CrossRef]

[37] Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*. [CrossRef]

[38] Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., & Batra, N. (2020). PlantDoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (pp. 249-253). [CrossRef]

[39] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, *33*, 9459-9474. https://proceedings.neurips.cc/paper_files/paper/2020/hash/6b493230-Abstract.html

[40] Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., ... & Guo, J. (2024, May). Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In *International Conference on Learning Representations* (Vol. 2024, pp. 3868-3898).

[41] Luo, L., Li, Y. F., Haffari, R., & Pan, S. (2024, May). Reasoning on graphs: Faithful and interpretable large language model reasoning. In *International Conference on Learning Representations* (Vol. 2024, pp. 14400-14423).

[42] Lee, J., Wang, Y., Li, J., & Zhang, M. (2024). Multimodal reasoning with multimodal knowledge graph. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics* (*Volume 1: Long Papers*) (pp. 10767-10782). [CrossRef]

[43] Lin, Q., Zhao, T., He, K., Peng, Z., Xu, F., Huang, L., ... & Feng, M. (2025, July). Self-supervised quantized representation for seamlessly integrating knowledge graphs with large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (*Volume 1: Long Papers*) (pp. 13587-13602). [CrossRef]

[44] Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. In *Advances in Neural Information Processing Systems* (Vol. 36, pp. 34892-34916). [CrossRef]

[45] Bai, J., Bai, S., Yang, S., Wang, S., Tan, S., Wang, P., ... & Zhou, J. (2023). Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*. [CrossRef]

[46] OpenAI. (2023). *GPT-4V(ision) system card*. Retrieved October 12, 2026 from https://openai.com/index/gpt-4v-system-card/

[47] Neo4j, Inc. (n.d.). *Neo4j graph database*. Retrieved October 12, 2026 from https://neo4j.com/

[48] Du, Y., Li, C., Guo, R., Yin, X., Liu, W., Zhou, J., ... & Wang, H. (2020). Pp-ocr: A practical ultra lightweight ocr system. *arXiv preprint arXiv:2009.09941*. [CrossRef]

[49] Christophides, V., Efthymiou, V., Palpanas, T., Papadakis, G., & Stefanidis, K. (2020). An overview of end-to-end entity resolution for big data. *ACM Computing Surveys* (*CSUR*), *53*(6), 1-42. [CrossRef]

[50] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PmLR. https://proceedings.mlr.press/v139/radford21a.html

[51] Qdrant. (n.d.). *Qdrant: Vector search engine*. Retrieved October 12, 2026 from https://qdrant.tech/

[52] Cui, Y., Wang, Y., Sun, Z., Liu, W., Jiang, Y., Han, K., & Hu, W. (2023, June). Lifelong embedding learning and transfer for growing knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 4, pp. 4217-4224). [CrossRef]

[53] Radovanovic, M., Nanopoulos, A., & Ivanovic, M. (2010). Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of machine learning research*, *11*(sept), 2487-2531. https://www.jmlr.org/papers/volume11/radovanovic10a/radovanovic10a.pdf

[54] Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, *8*(3), 489-508. [CrossRef]



**Xin Zeng** is an assistant research fellow in the Big Data and Intelligence Department at the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences. Zeng received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences. Zeng's research focuses on human–computer interaction and multimodal foundation models for agriculture. Zeng also serves as a reviewer for ICASSP and other international conferences and journals. (Email: zengx@nercita.org.cn)



**Wang Guo** is an assistant researcher at the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, a core member of the scientist team for intelligent management in the National Bulk Vegetable Industry Technology System, and a key technical member of the Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs. His research focuses on agricultural big data technology and digital village technology. (Email: guow@nercita.org.cn)



**Yisheng Miao** is a Research Professor at the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences. His main research interests include artificial intelligence in agriculture, particularly agricultural multimodal large models, intelligent vegetable production, and autonomous farming systems. (Email: miaoys@nercita.org.cn)



**Huaji Zhu** is the deputy director of the Big Data and Intelligence Department at the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, the secretary-general of the Intelligent Agriculture Committee of the Chinese Association for Artificial Intelligence, and a chief scientist in the National Key R&D Program. His research focuses on agricultural big data technology and digital village technology. (Email: zhuhj@nercita.org.cn)



**Huarui Wu** received the Ph.D. degree in engineering and is a Professor and Doctoral Supervisor. He is Deputy Director of the Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Intelligent Management Scientist of the National Bulk Vegetable Industry Technology System, and Director of the Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs. He has been selected for national talent programs and has received a Second Prize of the National Science and Technology Progress Award, nine provincial or ministerial science and technology awards, and the titles of Beijing Science and Technology New Star and Beijing Outstanding Talent. His research focuses on agricultural artificial intelligence algorithms and models, cloud management and control platforms, and autonomous unmanned systems. (Email: wuhr@nercita.org.cn)