NTCIR-19 DAGRI

Task type (Pilot)

Data Analytics for aGRicultural Information
https://sites.google.com/view/dagri/

Yasutomo Kimura

Otaru University of Commerce

Hiroki Sakaji

Hokkaido University

Kazuma Kadowaki

The Japan Research Institute, Limited; Shizuoka University

Tsuyoshi Morioka

AIREV, Inc.

Hayato Aida

Stockmark Inc.

Kosuke Takahashi

Stockmark Inc.

Akio Kobayashi

National Agriculture and Food Research Organization

Masahiro Otomo

National Agriculture and Food Research Organization

Junichi Ishihara

National Agriculture and Food Research Organization

Motivation

- Japanese agriculture is facing severe aging:
  - Average farmer age: 68 (2023)
  - 3/4 will retire within 20 years
- Intergenerational knowledge transfer is an urgent issue
- **AI-based agricultural applications** are expected
- Prefectural technical documents are important resources, but their availability is mostly restricted to **non–machine-readable PDFs**.
- The complexity of diagrams and tables presents **challenges for AI integration**

Goal

- Convert diagrams/tables into machine-readable formats
- Build a structured agricultural knowledge database
- Develop a question-answering system for agricultural guidance
- Propose a new framework for community-based knowledge inheritance
- Contribute to sustainable agriculture and future farmers

Proposed task
Data Analytics for aGRicultural Information (DAGRI)

Subtasks:

- Subtask 1: Conversion from PDF to CSV (PDF2CSV)
- Subtask 2: Table-based Question Answering (TableQA)
- Subtask 3: Figures and Tables Question Answering (FiT-QA)

Subtask 1: Conversion from PDF to CSV (PDF2CSV)

Input: PDF file containing tabular data

Output: CSV file (or alternatively, a JSON file)

Evaluation: We use GriTS (Grid Table Similarity) as the evaluation metric for table structure recognition.

経営類型	家族労働力	品目・栽培型及び規模		経営・技術の特徴
小麦	人 2		a 小麦(長崎W2号) 800 合計 800 経営耕地面積 水田8ha(自作地4ha、借入地4ha)	(1)土地利用型作物の複合経営 (2)作業の一部は委託
経営目標	1 農業総収入 7,589 千円 2 農業経営費 5,894 千円 3 農業所得 1,695 千円		4 1日当たり農業所得 33,567 円 5 1人当たり年間労働時間 202 時間	

経営類型	家族労働力	品目・栽培型及び規模		経営・技術の特徴
小麦	人		a	(1)土地利用型作物の複合経営(2)作業の一部は委託
	2	小麦(長崎W２号)	800	
		合計	800	
		経営耕地面積	水田８ｈａ自作地4ha、借入地4ha)	
経営目標	1 農業総収入 7,589 千円 4 1日当たり農業所得 33,567 円 2 農業経営費 5,894 千円 5 1人当たり年間労働時間 202 時間 3 農業所得 1,695 千円			

Subtask 2: Table-based Question Answering (TableQA)

Input:

- Tables contained in the PDF
- A question sentence

Output:

- An answer sentence

Evaluation: We are currently considering BERTScore, ROUGE, and other metrics.

Question: What are the key considerations when selecting rice varieties?
質問: 水稲の品種の選定の重要事項を教えてください。
Answer: Select from recommended varieties. Renew seeds annually. Choose heat-tolerant varieties.
回答: 奨励品種から選定する。毎年種子更新に努める。高温耐性品種を選定する。

Subtask 3: Figures and Tables Question Answering (FiT-QA)

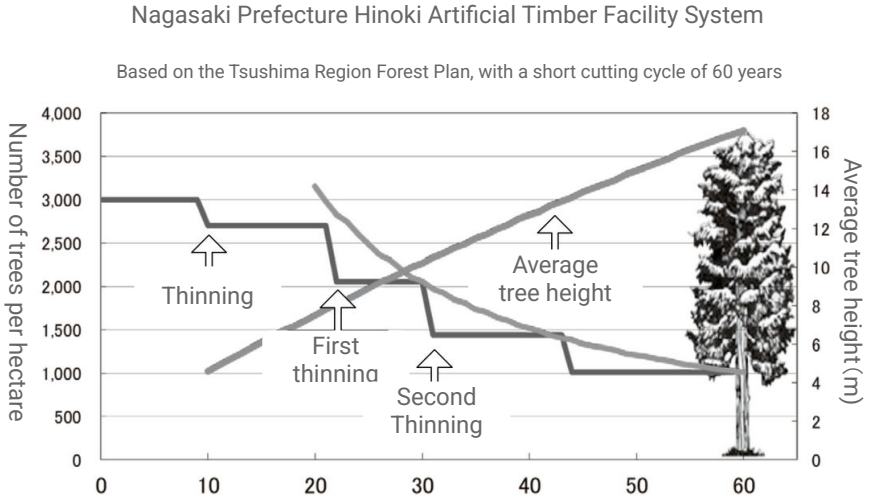
Input:

- Figures and tables contained in PDF documents
- A question

Output:

- Answer sheet in JSON format

Evaluation: We are currently considering BERTScore, ROUGE, and other metrics



Average tree height	5	8	10	13	15	17
Average diameter at breast height			14	18	22	24
形状比	71	73	71	72	68	71
見込み林齢	10	20	30	40	50	60
除間伐 回数	Thinning	First	Second	Third		
除間伐 林齢	10	22	31	44		
除間伐 間伐本数	300	648	616	431		
除間伐 間伐率	10%	24%	30%	30%		
除間伐 残存本数	2,700	2,052	1,436	1,005		
収量比数		0.51	0.52	0.53	0.65	
旧Ry		0.67	0.70	0.73	0.84	

Schedule

- November, 2025: 1st Orientation Meeting of DAGRI
- January – March 2026: Distribution of Preliminary Test Dataset
- January – June 2026: Preliminary Test (Dry Run)
- April, 2026: 2nd Orientation Meeting of DAGRI
- June 2026: Distribution of Formal Test Dataset
- June – July 2026 (July 1–15): Formal Test (Formal Run)
- July 18–29, 2026: Human Evaluation
- August 1, 2026: Return of Evaluation Results to Participants
- Partial release of task overview paper
- September 1, 2026: Deadline for Draft Papers from Task Participants
- November 1, 2026: Final Submission Deadline for Task Papers and Task Overview Papers
- December 8–10, 2026: NTCIR-19 Conference (NII, Tokyo)