Results of SemTab 2025

Oktie Hassanzadeh[1,*], Marco Cremaschi[2], Fabio D'Adda[2], Fidel Jiomekong Azanzi[3], Jean Petit BIKIM[3] and Ernesto Jimenez-Ruiz[4]

[1]*IBM Research, USA*

[2]*University of Milan–Bicocca, Italy*

[3]*Department of Computer Science, University of Yaoundé 1, Cameroon*

[4]*City St. George's, University of London, UK*

Abstract

2025 continues the mission of benchmarking semantic table interpretation systems in an era shaped by large language models (LLMs) and retrieval-augmented strategies. This edition introduces two complementary tracks. The first is the large-scale *MammoTab* dataset for cell entity annotation (CEA) aligned to Wikidata. The second is the new *Secu-Table* Track designed to evaluate robustness under noisy, adversarial, and security-domain table conditions. This paper presents an overview of the challenge setup, datasets, evaluation methodology, and the key outcomes of this edition.

Keywords

Tabular data, Knowledge Graphs, Matching, SemTab Challenge, Semantic Table Interpretation, Large Language Models, LLMs

1. Introduction

Tabular data continues to serve as a foundational format across Web datasets, enterprise data lakes, and analytics pipelines. However, semantic interpretation of tabular data remains a difficult challenge in many real-world scenarios. A central task is the annotation of table elements with entities, classes, and relationships from a knowledge graph (KG). This process, known as *Semantic Table Interpretation* (STI), supports data discovery, integration, analytics, and knowledge-based augmentation. STI research has been shaped by both early foundational work on semantic table annotation [1, 2, 3, 4] and more recent advances in machine learning and large language models [5].

Since its inception in 2019, the SemTab Challenge has provided an annual benchmarking platform for evaluating STI systems across multiple subtasks. The challenge has expanded over the years and has been featured at ISWC through all editions from 2019 to 2024 [6, 7, 8, 9, 10, 11]. These efforts have driven the development of diverse STI approaches, including traditional pipelines, embedding-based models, LLM-based systems, and retrieval-augmented frameworks.

This paper presents the 2025 edition of the challenge, summarizing its tracks, datasets, evaluation metrics, and results.

2. The Challenge

The 2025 challenge includes two main tracks. The first is the **MammoTab Track**, which focuses on large-scale Cell Entity Annotation (CEA). The second is the **Secu-Table Track**, which evaluates robustness under noisy or adversarial table conditions and includes CEA, CTA, and CPA subtasks.

2.1. MammoTab Track (CEA)

Task. Participants perform Cell Entity Annotation by linking entity-bearing table cells to Wikidata entity identifiers (QIDs), or marking them as NIL when no matching entity exists.

Dataset. MammoTab is a large-scale dataset composed of more than one million Wikipedia tables annotated using Wikidata [12]. It captures the inherent ambiguity, heterogeneity, and missing context characteristic of web tables.

The Round 1 subset contains about 870 tables with 84,907 annotated cells. It reflects natural complexity, such as alias variation, homonymy, and NIL mentions. These challenges are motivated by earlier SemTab editions [10, 11].

An example of the annotation format is:

```
LYQZQ0T5,1,1,Q3576864
```

indicating that table `LYQZQ0T5`'s cell in row 1 column 1 needs to be mapped to Wikidata entity `Q3576864`.

Evaluation. Each system outputs one entity prediction per cell. Evaluation uses:

$$\text{Precision} = \frac{|\text{Correct Annotations}|}{|\text{Submitted Annotations}|}$$

$$\text{Recall} = \frac{|\text{Correct Annotations}|}{|\text{Ground Truth Annotations}|}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1 is the primary ranking metric, with Precision as tie-breaker.

2.2. Secu-Table Track (Robustness and Multiple Tasks)

Task. The Secu-Table Track evaluates the robustness of systems under noise, incomplete context, perturbations, and adversarial scenarios. It includes:

- **CEA**: Cell Entity Annotation,
- **CTA**: Column Type Annotation,
- **CPA**: Column Property Annotation.

Dataset. The dataset includes 1,554 tables, including 76 gold tables and 1,478 test tables [13]. Noise includes:

- misspellings and alias variants,
- missing or incomplete context,
- added ambiguity or homonymy,
- NIL mentions.

Two knowledge graphs are used: Wikidata for general-purpose tasks and the SEPSES Computer Security KG for security-domain CEA, CTA and CPA. Robust STI has been highlighted in past studies [14, 15].

Formats:

- **CEA**: `filename,row_id,column_id,entity_id`
- **CTA**: `filename,column_id,type_id`
- **CPA**: `filename,col0,col1,property_id`

Evaluation. Same metrics as for MammoTab:

$$\text{Precision} = \frac{|\text{Correct Annotations}|}{|\text{Submitted Annotations}|}$$

$$\text{Recall} = \frac{|\text{Correct Annotations}|}{|\text{Ground Truth Annotations}|}$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A selective mode allows predicting "I do not know" when confidence is low.

3. Results

Four systems participated in the MammoTab Track:

- **ADFr** [16]: This system is built on GRASP [17], a framework for interactive KG reasoning and SPARQL generation. GRASP enables the LLM to retrieve and validate KG entities before producing annotations. It was originally introduced for SPARQL question answering and KG exploration. The system iteratively searches, validates, and assigns entity predictions using evidence from the KG.
- **RAGDify** [18]: This is a multi-stage retrieval and generation pipeline. It cleans table data, retrieves candidates using exact and fuzzy methods, performs LLM-based debate ranking, and applies LLM cross-level verification for NIL detection and consistency.
- **ditlab** [19]: This framework uses iterative refinement by annotating both original and transposed table orientations, improving contextual consistency. It integrates CEA and CTA through multi-stage candidate generation and unsupervised entropy-based scoring.
- **Kepler-aSI**: This system combines KG querying and LLM-based disambiguation. Earlier versions were presented in previous SemTab editions [20, 21].

System	Precision	Recall	F1 Score
ADFr	0.758	0.758	**0.758**
RAGDify	0.603	0.603	0.603
ditlab	0.549	0.549	0.549
Kepler-aSI	0.403	0.157	0.226

Table 1
Performance of systems on the MammoTab Track in SemTab 2025.

Acknowledgments

Declaration on Generative AI

During the preparation of this work, the authors used an LLM to assist with text generation based on website materials and informal notes. All content was reviewed and edited by the authors, who take full responsibility for the final version.

References

[1] Z. Zhang, Effective and efficient semantic table interpretation using tableminer+, Semantic Web 8 (2017) 921–957.

[2] Z. Syed, T. Finin, V. Mulwad, , A. Joshi, Exploiting a Web of Semantic Data for Interpreting Tables, in: Proceedings of the Second Web Science Conference, 2010.

[3] V. Mulwad, T. Finin, Z. Syed, A. Joshi, T2LD: interpreting and representing tables as linked data, in: A. Polleres, H. Chen (Eds.), Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts, Shanghai, China, November 9, 2010, volume 658 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2010. URL: https://ceur-ws.org/Vol-658/paper489.pdf.

[4] V. Efthymiou, O. Hassanzadeh, M. Rodriguez-Muro, V. Christophides, Matching Web Tables with Knowledge Base Entities: From Entity Lookups to Entity Embeddings, in: ISWC, volume 10587, Springer, 2017, pp. 260–277.

[5] Y. Suhara, J. Li, Y. Li, D. Zhang, c. Demiralp, C. Chen, W.-C. Tan, Annotating columns with pre-trained language models, in: Proceedings of the 2022 International Conference on Management of Data, SIGMOD '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 1493–1503. URL: https://doi.org/10.1145/3514221.3517906. doi:10.1145/3514221.3517906.

[6] E. Jiménez-Ruiz, O. Hassanzadeh, K. Srinivas, V. Efthymiou, J. Chen (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2019, co-located with the 18th International Semantic Web Conference (ISWC 2019), volume 2553 of *CEUR Workshop Proceedings*, CEUR-WS.org, Auckland, New Zealand, October 30, 2019, 2019.

[7] E. Jiménez-Ruiz, O. Hassanzadeh, V. Efthymiou, J. Chen, K. Srinivas, V. Cutrona (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2020, co-located with the 19th International Semantic Web Conference (ISWC 2020), volume 2775 of *CEUR Workshop Proceedings*, CEUR-WS.org, Virtual conference (originally planned Athens, Greece), November 5, 2020, 2020.

[8] E. Jiménez-Ruiz, O. Hassanzadeh, V. Efthymiou, J. Chen, K. Srinivas, V. Cutrona, J. Sequeda, M. Hulsebos, N. Abdelmageed, D. Oliveira, C. Pesquita (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2021, co-located with the 20th International Semantic Web Conference (ISWC 2021), volume 3103 of *CEUR Workshop Proceedings*, CEUR-WS.org, Virtual conference, October 27, 2021, 2021.

[9] N. Abdelmageed, J. Chen, V. Cutrona, V. Efthymiou, O. Hassanzadeh, M. Hulsebos, E. Jiménez-Ruiz, J. Sequeda, K. Srinivas (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2022, co-located with the 21st International Semantic Web Conference (ISWC 2022), volume 3320 of *CEUR Workshop Proceedings*, CEUR-WS.org, Virtual conference, October 23-27, 2022, 2022.

[10] V. Efthymiou, E. Jiménez-Ruiz, J. Chen, V. Cutrona, O. Hassanzadeh, J. Sequeda, K. Srinivas, N. Abdelmageed, M. Hulsebos, A. Khatiwada, K. Korini, B. Kruit (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2023, co-located with the 22nd International Semantic Web Conference (ISWC 2023), volume 3557 of *CEUR Workshop Proceedings*, CEUR-WS.org, Athens, Greece, November 6-10, 2023, 2023.

[11] O. Hassanzadeh, N. Abdelmageed, M. Cremaschi, V. Cutrona, F. D'Adda, V. Efthymiou, B. Kruit, E. Lobo, N. Mihindukulasooriya, N. H. Pham (Eds.), Proceedings of the Semantic Web Challenge on Tabular Data to Knowledge Graph Matching, SemTab 2024, co-located with the 23rd International Semantic Web Conference (ISWC 2024), volume 3889 of *CEUR Workshop Proceedings*, CEUR-WS.org, Baltimore, USA, November 11-15, 2024, 2024.

[12] M. Cremaschi, F. Belotti, J. D'Souza, M. Palmonari, Mammotab 25: A large-scale dataset for semantic table interpretation - training, testing, and detecting weaknesses, in: The Semantic Web - ISWC 2025 - 24th International Semantic Web Conference, Nara, Japan, November 2-6, 2025, Proceedings, Part II, volume 16141 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 131–148. URL: https://doi.org/10.1007/978-3-032-09530-5_8. doi:10.1007/978-3-032-09530-5_8.

[13] A. Jiomekong, J. Bikim, P. Negoue, J. Chin, Secu-table: a comprehensive security table dataset

for evaluating semantic table interpretation systems, 2025. URL: https://arxiv.org/abs/2511.06301. arXiv:2511.06301.

[14] B. Foko, A. Jiomekong, H. Tapamo, J. Buisson, S. Tiwari, Exploring Naive Bayes Classifiers for Tabular Data to Knowledge Graph Matching, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.

[15] J. P. Bikim, C. Atezong, A. Jiomekong, A. Oelen, G. Rabby, J. D'Souza, S. Auer, Leveraging GPT Models For Semantic Table Annotation, in: SemTab'24: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2024, co-located with the 23rd International Semantic Web Conference (ISWC), 2024.

[16] S. Walter, H. Bast, Knowledge graph entity linking via interactive reasoning and exploration with grasp, in: Proceedings of the 20th International Workshop on Ontology Matching (OM 2025) co-located with the 24th International Semantic Web Conference (ISWC 2025), Nara, Japan, 2025.

[17] S. Walter, H. Bast, Grasp: Generic reasoning and sparql generation across knowledge graphs, in: The Semantic Web – ISWC 2025: 24th International Semantic Web Conference, Nara, Japan, November 2–6, 2025, Proceedings, Part I, Springer-Verlag, Berlin, Heidelberg, 2025, p. 271–289. URL: https://doi.org/10.1007/978-3-032-09527-5_15. doi:10.1007/978-3-032-09527-5_15.

[18] K. Bar, T. Sagi, Llm-driven retrieval, debate, and verification for robust table-to-knowledge-graph matching, in: Proceedings of the 20th International Workshop on Ontology Matching (OM 2025) co-located with the 24th International Semantic Web Conference (ISWC 2025), Nara, Japan, 2025.

[19] Y. Tachioka, Y. Terao, Cell entity annotation for semtab 2025 mammotab via iterative refinement with transposed contexts and unsupervised scoring, in: Proceedings of the 20th International Workshop on Ontology Matching (OM 2025) co-located with the 24th International Semantic Web Conference (ISWC 2025), Nara, Japan, 2025.

[20] W. Baazouzi, M. Kachroudi, S. Faiz, Kepler-aSI at SemTab 2023, in: SemTab'23: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2023, co-located with the 22nd International Semantic Web Conference (ISWC), 2023.

[21] W. Baazouzi, M. Kachroudi, S. Faiz, Kepler-aSI : Semantic Annotation for Tabular Data, in: SemTab'24: Semantic Web Challenge on Tabular Data to Knowledge Graph Matching 2024, co-located with the 23rd International Semantic Web Conference (ISWC), 2024.