

## Semantic Web Challenge on Tabular Data to Knowledge Graph Matching

Tabular data in the form of CSV files is the common input format in a data analytics pipeline. However, a lack of understanding of the semantic structure and meaning of the content may hinder the data analysis process. Thus gaining this semantic understanding will be very valuable for data integration, data cleaning, data mining, machine learning and knowledge discovery tasks. For example, understanding what the data is can help assess what sorts of transformation are appropriate on the data.

Tables on the Web may also be the source of highly valuable data. The addition of semantic information to Web tables may enhance a wide range of applications, such as web search, question answering, and knowledge base (KB) construction.

Tabular data to Knowledge Graph (KG) matching is the process of assigning semantic tags from Knowledge Graphs (e.g., Wikidata or DBpedia) to the elements of the table. This task however is often difficult in practice due to metadata (e.g., column names) being missing, incomplete or ambiguous.

This challenge aims at benchmarking systems dealing with the tabular data to KG matching problem, so as to facilitate their comparison on the same basis and the reproducibility of the results.

The 2019 edition of this challenge will be colocated with the 18th International Semantic Web Conference and the 14th International Workshop on Ontology Matching.

### Results and Challenge Prizes

Results of all four rounds available [here](#). Summary of SemTab 2019 results.

Prizes sponsored by SIRIUS and IBM Research:

- 1st Prize (CTA, CEA and CPA): MTab Team.
- 2nd Prize (CTA, CEA and CPA): IDLab Team.
- 3rd Prize (CTA, CEA and CPA): Tabularis Team.
- 4th Prize (CEA): ADGOS Team.
- Outstanding Improvement (CEA): Team STI.

### Datasets and Evaluator

The challenge datasets and ground truths are now open: <https://doi.org/10.5281/zenodo.3518539>

You can cite the dataset as:  
*Cihan Hassanzadeh, Vasilis Eftimios, Jieyan Chen, Ernesto Jimenez-Ruiz, and Kavitha Srinivas. (2019). SemTab 2019 Semantic Web Challenge on Tabular Data to Knowledge Graph Matching - 2019 Data Sets (Version 2019) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3518539>*

The codes of the AIcrowd evaluator are also available [here](#).

### ISWC Challenge Presentations

The results of the challenge will be presented on October 30 (11:40-12:40). See full ISWC program [here](#). Four participating teams will also present their systems.

- 11:40-12:00: Challenge overview & announcement of awards. ([slides](#)) ([photos](#))
- 12:00-12:10: MTAB Matching Tabular Data to Knowledge Graph using Probability Models by [Phuc Nguyen, Nathawut Kerdkachorn, Ryoko Ichise and Hidaki Takeda](#). ([slides](#))
- 12:10-12:30: MantisTable: An Automatic Approach for the Semantic Table Interpretation by [Aviit Thawani, Minds Hu, Erdong Hu, Husain Zafar, Naren Teja Divvala, Amandeep Singh, Ehsan Qasemi, Pedro Szekely and Jay Pujara](#). ([slides](#))
- 12:30-12:40: DAGOBA: An End-to-End Context-Free Tabular Data Semantic Annotation System by [Yvan Chabot, Thomas Labbe, Jixiong Liu and Raphaël Troncy](#). ([slides](#))

Presentations during the Ontology Matching workshop on October 26:

- Challenge overview. ([slides](#))
- Presentation during OM workshop: ISWC challenge: transforming tabular data into semantic knowledge by [Gilles Vandewiele, Bram Steenwinckel, Filip De Turck, Femke Ongena](#). ([slides](#))

### System papers

Papers published in the [Vol-2533 of CEUR Workshop Proceedings](#).

- [Daniela Oliveira and Matheus d'Águila. ADGG - Annotating Data with Ontologies and Graphs](#)
- [Marco Cremaschi, Roberto Avogadro, and David Chiragere. MTab: Matching Tabular Data to Knowledge Graph using Probability Models](#)
- [Marco Cremaschi, Roberto Avogadro, and David Chiragere. MantisTable: An Automatic Approach for the Semantic Table Interpretation](#)
- [Aviit Thawani, Minds Hu, Erdong Hu, Husain Zafar, Naren Teja Divvala, Amandeep Singh, Ehsan Qasemi, Pedro Szekely and Jay Pujara. Entity Linking to Knowledge Graphs to Infer Column Types and Properties](#)
- [Gilles Vandewiele, Bram Steenwinckel, Filip De Turck, and Femke Ongena. CVS2KG: Transforming Tabular Data into Semantic Knowledge](#)
- [Yvan Chabot, Thomas Labbe, Jixiong Liu and Raphaël Troncy. DAGOBIA: An End-to-End Context-Free Tabular Data Semantic Annotation System](#)
- [Hiroaki Morikawa. Semantic Table Interpretation using LOD4ALL](#).

### Challenge Tasks

The challenge includes the following tasks organised into several evaluation rounds:

- Assigning a semantic type (e.g., a KG class) to a column: **CTA task**. Datasets: Round 1, Round 2 (targets), Round 3 (targets), Round 4 (targets)
- Matching a cell to a KG entity: **CEA task**. Datasets: Round 1, Round 2 (targets), Round 3 (targets), Round 4 (targets)
- Assigning a KG property to the relationship between two columns: **CPA task**. Datasets: Round 1, Round 2 (targets), Round 3 (targets), Round 4 (targets)

The challenge will be run with the support of the [AIcrowd](#) platform.

NEW: please register your system details [here](#).

NEW: we have created a [discussion group](#) for the challenge.

### Support for ontology alignment and link discovery

Ontology alignment and link discovery systems are welcome to participate. Please follow the instructions for the **CEA task**.

- Round 2 datasets in RDF (ttl format): [tables](#) (single and multiple files) and [dbpedia knowledge graph](#) (single and multiple fragments).

### Challenge Prizes

There will be prizes sponsored by [SIRIUS](#) and [IBM Research](#) for the best systems and the best student systems in the challenge.

The prize winners will be announced during the [ISWC conference](#) (on October 30, 2019).

We will take into account all evaluation rounds specially the one running till the conference dates, the covered tasks and the novelty of the applied techniques (we encourage the submission of a [system paper](#)).

### Important Dates

- Open: Please register your system details [here](#).
- April 15: Round 1 opens
- June 1: Round 1 closes
- July 1: Best systems in Rounds 1 and 2 are invited to present their results during [ISWC conference](#) and the [Ontology Matching workshop](#). Check [ISWC 2019 student travel grants](#).
- July 17: Round 2 opens
- September 22 (extended): Round 2 closes
- September 27 (extended): Round 3 closes
- September 27 (extended): System paper submissions (preliminary version, e.g., `system_name_prelim.pdf`). Please use [this form](#).
- October 15: Round 3 opens (tentative)
- October 15: Round 4 opens (tentative)
- October 20: Round 4 closes (tentative)
- October 26: [Ontology Matching workshop](#)
- October 30: [Challenge Presentation](#) and prize announcement
- November 10: System paper submissions (final version, e.g., `system_name_final.pdf`). Please use [this form](#).

### Guidelines for System Papers

We encourage participants to submit a system paper. The paper should be no more than 8 pages long and formatted using the [LaTeX template](#). System papers will be reviewed by 1-2 challenge organisers. Please use [this form](#) for the submission (requires a Google account and a valid email).

To ensure easy comparability among the participants we suggest the following outline:

1. Problem definition statement
  - 1. State purpose, general statement
  - 2. Specific techniques used
  - 3. Adaptations made for the evaluation
  - 4. Link to the system and parameters
2. Results
  - 2. (x) a comment for each task/dataset performed
  - 3. General comments (if relevant)
    - 1. Comments on the system (strength and weaknesses)
    - 2. Discussions on the way to improve the proposed system
    - 3. Comments on the challenge procedure
    - 4. Comments on the challenge test cases
    - 5. Comments on the challenge measures
    - 6. Proposal of new datasets, tasks or measures
4. Conclusions
5. References

### Organisation

#### Challenge chairs

This track is organised by [Kavitha Srinivas](#) (IBM Research), [Ernesto Jimenez-Ruiz](#) (City, University of London), [Alan Turing Institute](#), [University of Oslo](#), [Ola Hassanein](#) (IBM Research), [Jieyan Chen](#) (University of Oxford) and [Vasilis Eftimios](#) (IBM Research). If you have any problems working with the datasets or any suggestions related to this challenge, do not hesitate to contact us.

#### Challenge committee members

- Udayan Khurana (IBM Research)
- Erik Broberg Myklevoll (University of Oslo)
- Vasilis Eftimios (IBM Research)
- Monika Solanki (Agricenter)
- Ole Magnus Holtet (University of Oslo)
- Pedro Szekely (University of California, Berkeley)
- Udo Holler (University of Bielefeld, University of Oslo)
- Marco Cremaschi (University of Milano - Bicocca)
- Asan Aglibov (Medical University of Vienna)

### Acknowledgements

The challenge is currently supported by the [AIDA project](#), the [SIRIUS Centre for Research-driven Innovation](#), and [IBM Research](#).