DAGOBAH: Table and Graph Contexts For Efficient Semantic Annotation Of Tabular Data

Viet-Phi Huynh, Jixiong Liu, Yoan Chabot, Frédéric Deuzé, Thomas Labbé, Pierre Monnin, Raphael Troncy, Guillaume Dive
yoan.chabot@orange.com raphael.troncy@eurecom.fr

Annotation Workflow

Pre-processing

Semantic annotations

Hard Scoring

Context score of a candidate entity means quantifying the relevance of neighboring cells

ScoreHard(\(c_1\)) = \(f(Lev(e_1, c_1), Lev(n_1, c_1))\)

generality(\(Q12782\)) \(\ll\) generality(\(Q142\)) \(\Rightarrow\) \(Q18578267\) more informative

Score\(_{\text{Hard}}\)(\(c_1\)) = \(g(\text{Score}_{\text{Soft}}, \text{Score}_{\text{Hard}})\)

Soft Scoring

Each neighboring cell does not have an equal contribution

Usability

API for developers
UI for everyone

Future Works

- Annotation enhancement (dictionaries, embeddings)
- Data augmentation
- Hardcore corpus

SemTab @ ISWC 2021 Results

Best results