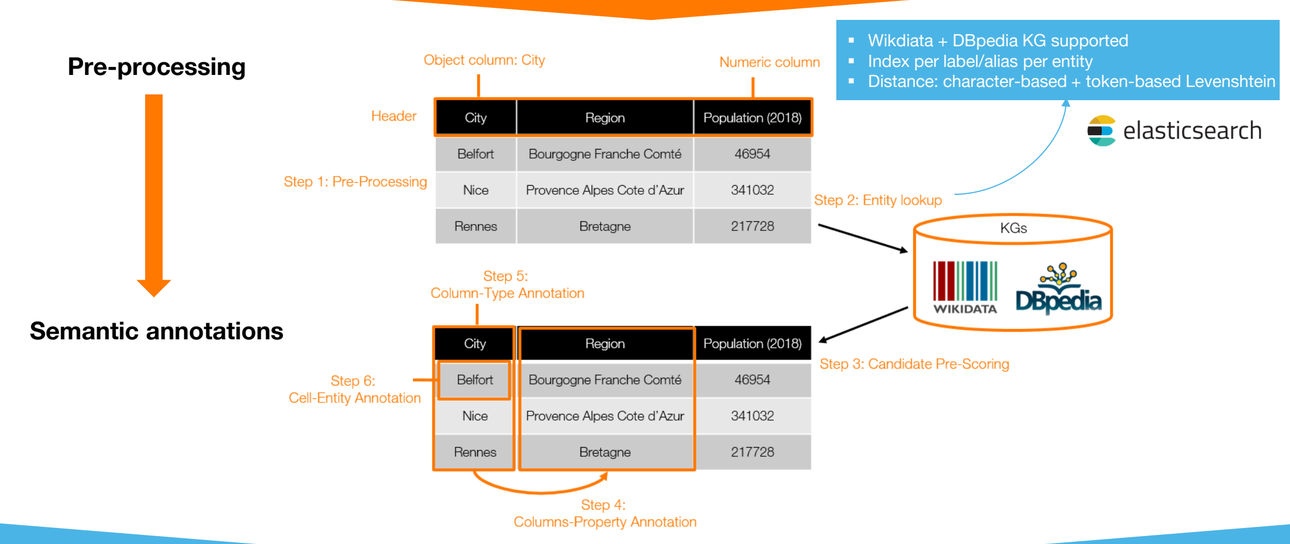
Viet-Phi Huynh, Jixiong Liu, Yoan Chabot, Frédéric Deuzé,
Thomas Labbé, Pierre Monnin, Raphaël Troncy, **Guillaume Dive**
yoan.chabot@orange.com raphael.troncy@eurecom.fr

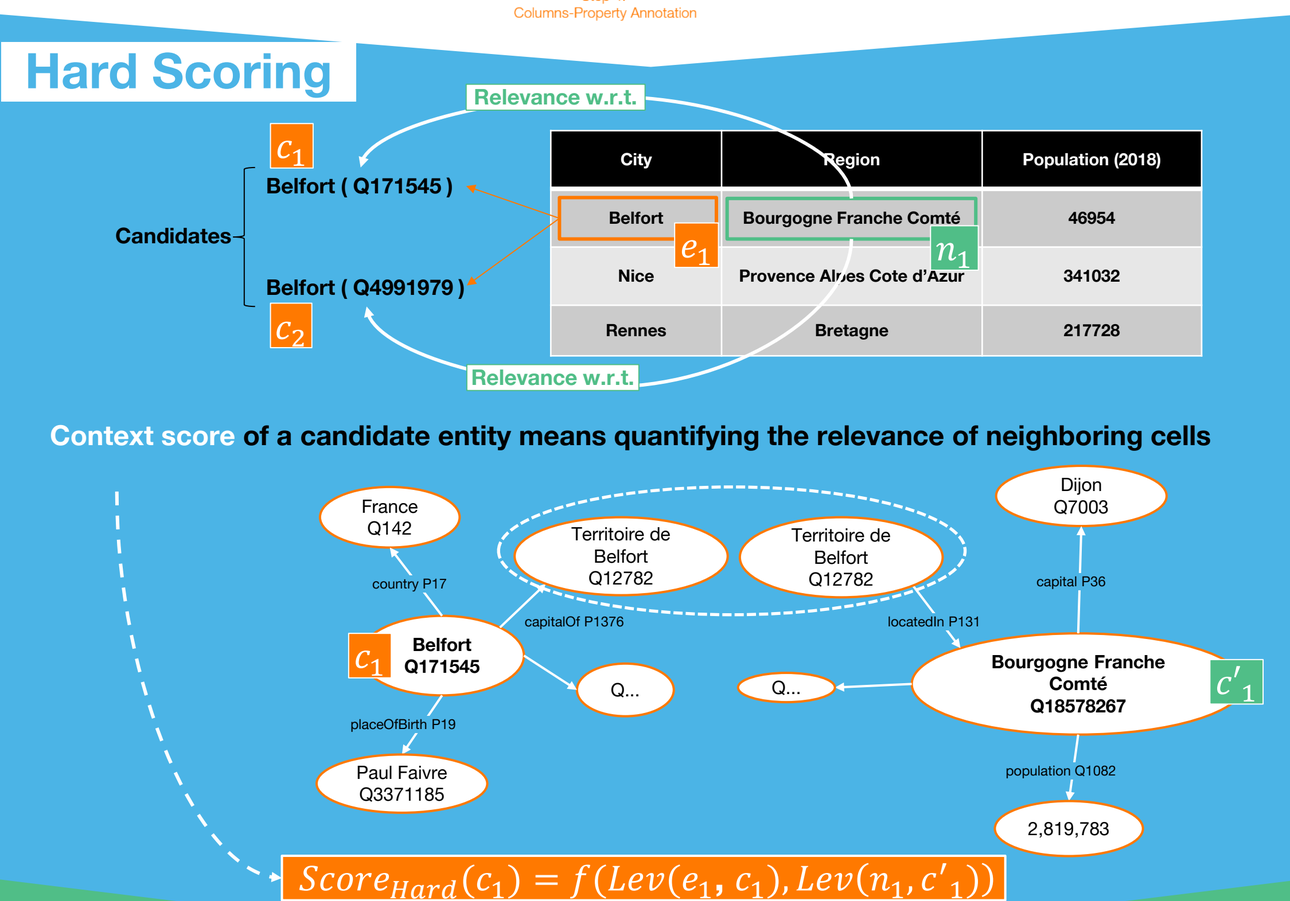
ISWC 2021

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

Annotation Workflow

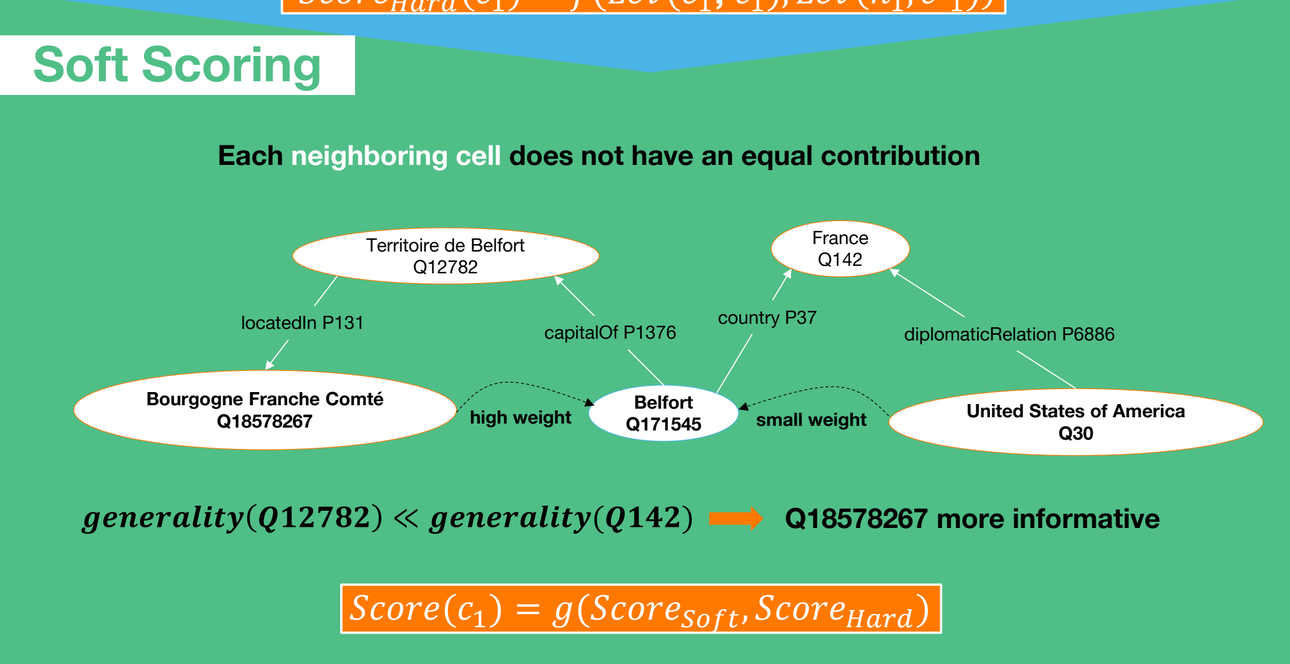


Hard Scoring



$$Score_{Hard}(c_1) = f(Lev(e_1, c_1), Lev(n_1, c'_1))$$

Soft Scoring

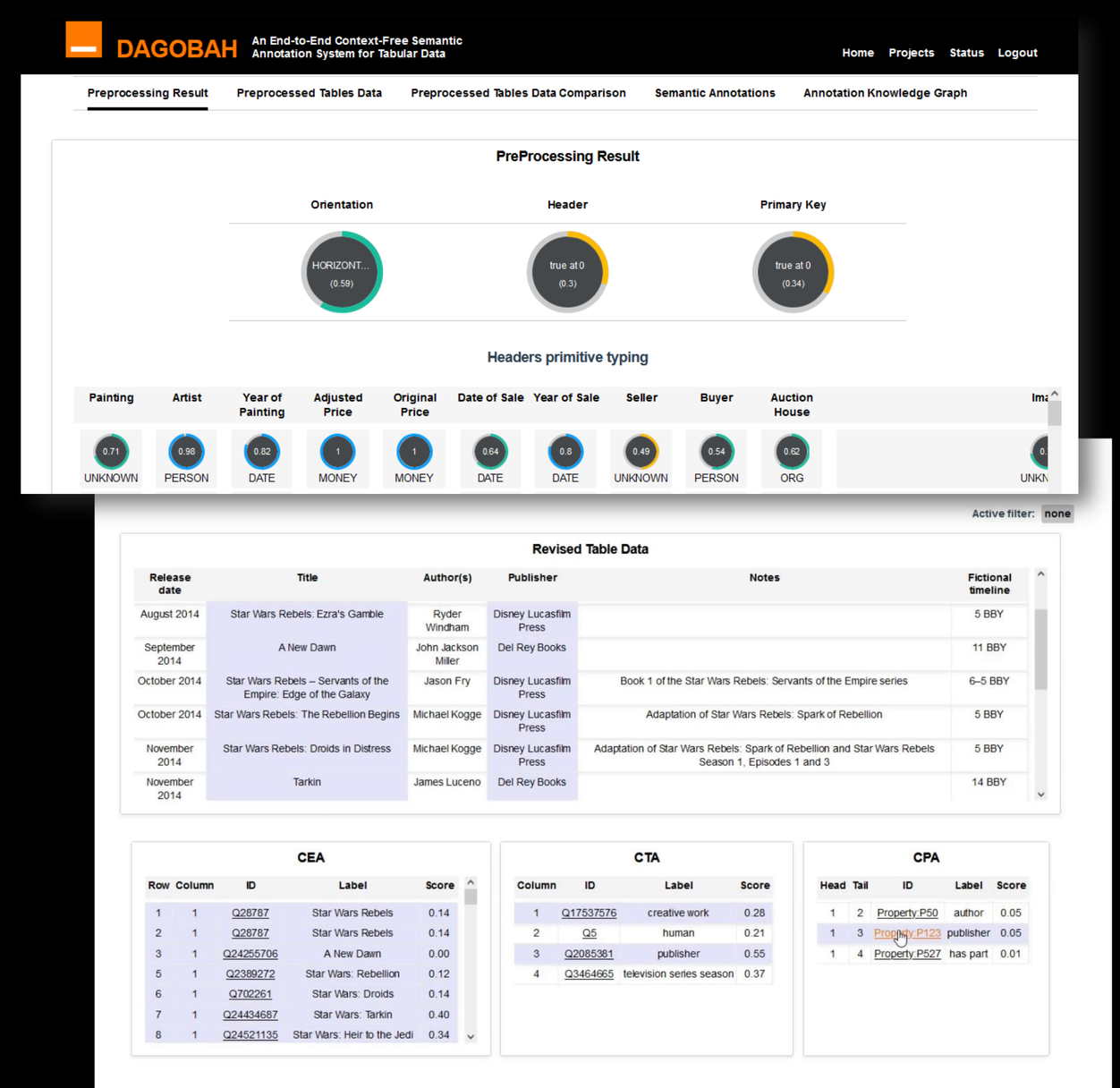


$generality(Q12782) \ll generality(Q142)$ → Q18578267 more informative

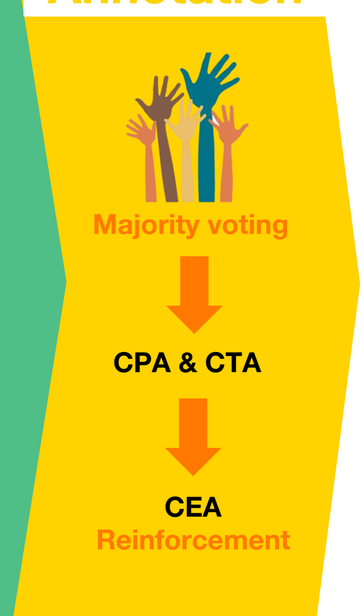
$$Score(c_1) = g(Score_{Soft}, Score_{Hard})$$

Usability

- ✓ API for developers
- ✓ UI for everyone



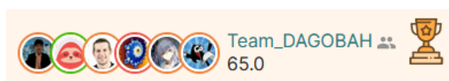
Annotation



SemTab @ ISWC 2021 Results

Task	Setting	CTA		CEA		CPA	
Criteria		F1	Precision	F1	Precision	F1	Precision
Round 1-WDTables	2*	0.832	0.832	0.923	0.923	x	x
Round 1-DBPTables	2*	0.422	0.424	0.945	0.946	x	x
Round 2-BioTables	3	0.916	0.916	0.970	0.970	0.899	0.899
Round 2-HardTables	3	0.976	0.976	0.975	0.976	0.996	0.996

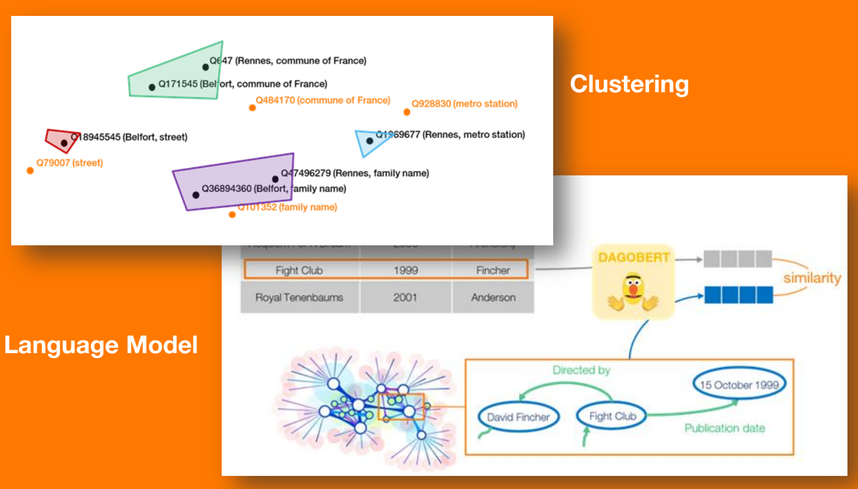
Hard Scoring + Soft Scoring → Best results



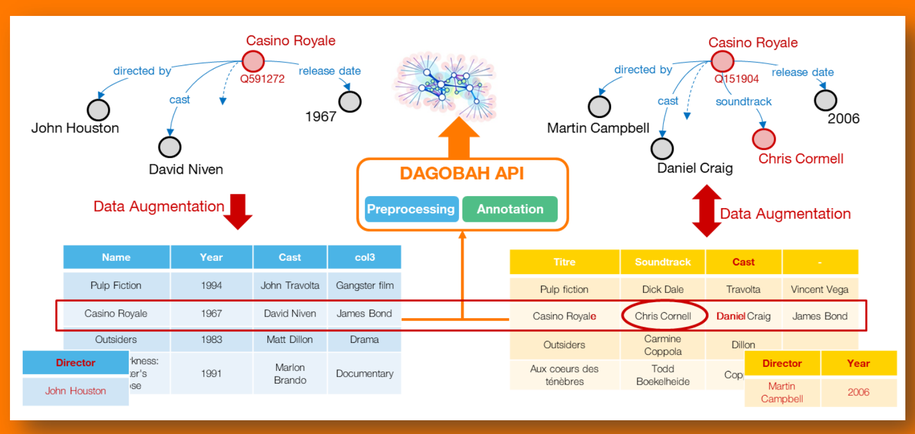
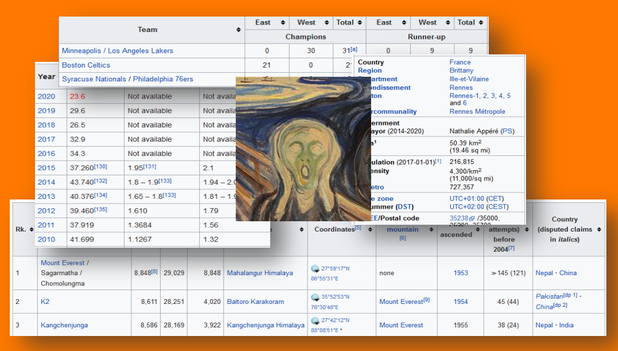
Top-ranked system (Rounds 1 + 2)
(with few submissions)

Future Works

- ✓ Annotation enhancement (dictionaries, embeddings)



- ✓ Data augmentation



- ✓ Hardcore corpus