

boosted by wiki

bbw:

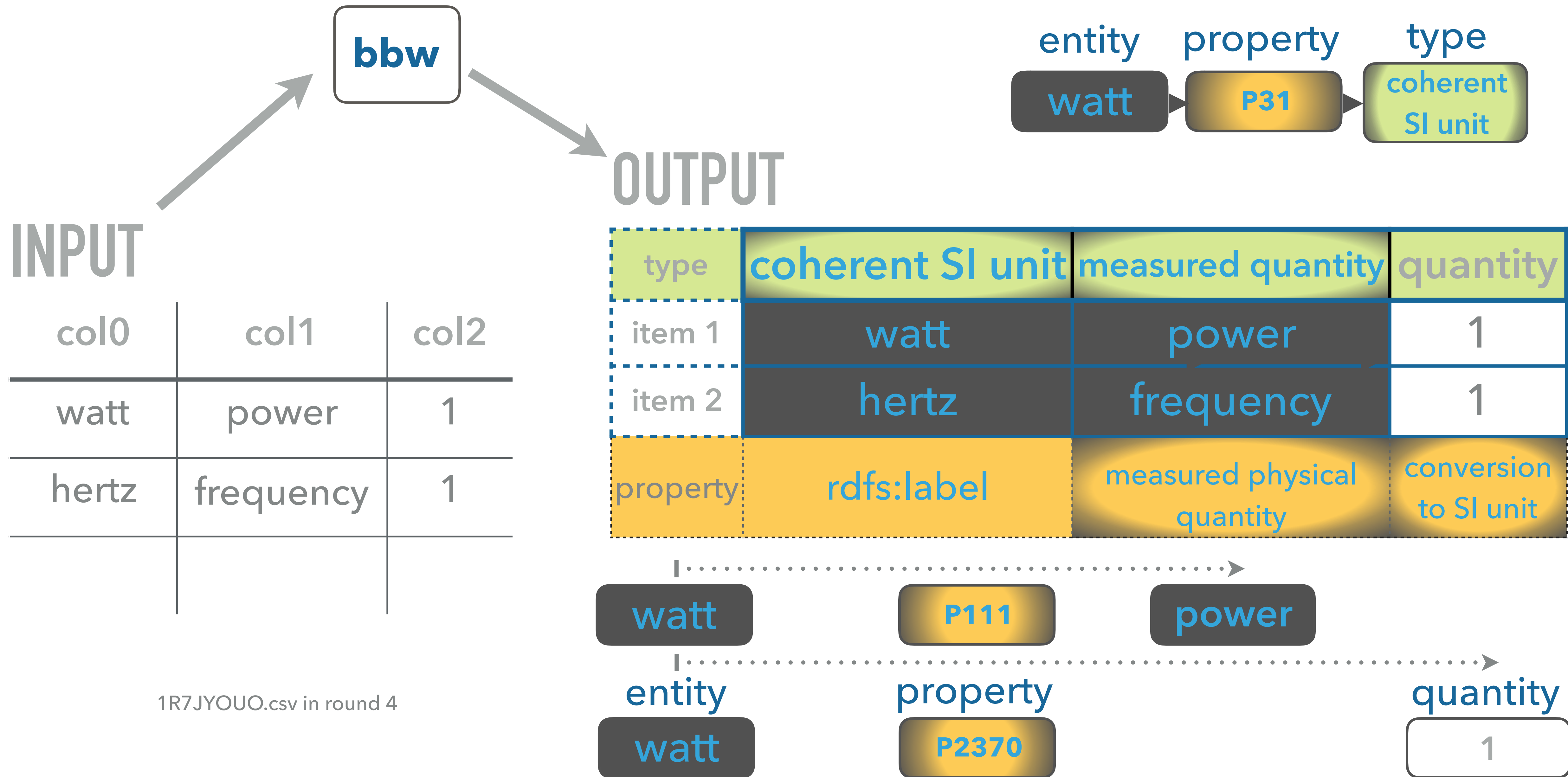
Matching CSV to Wikidata via Meta-lookup

Renat Shigapov, Philipp Zumstein, Jan Kamlah, Lars Oberländer, Jörg Mechnich and Irene Schumm

Mannheim University Library, University of Mannheim

Business and Economics Research Data Center Baden-Württemberg

SEMTAB2020: ENTITY, PROPERTY AND TYPE ANNOTATIONS



Meta-lookup

lookup over many search and metasearch engines



DuckDuckGo.



Contextual matching

matching an Item using the at least two features

hertz + frequency = hertz

herzt + coherent SI unit = hertz

herzt + P111 frequency = hertz
P2370 1

7-STEPS WORKFLOW

Step 1:
Preprocessing

Steps 2-6:
Lookup-Matching

Step 7:
Postprocessing

`ftfy.clean_text()`

predict 4 datatypes

Meta-lookup

Matching

a few Entities,
Properties and Types

most frequent Entity

most frequent Property

common Type via
`gas:service`

RESULTS, DISCUSSION AND CONCLUSIONS

Round	CPA			CTA			CEA		
	F1	P	Po	AF1	AP	Po	F1	P	Po
1	-	-	-	-	-	-	-	-	-
2 (<1 st deadline)	-	-	-	-	-	-	-	-	-
2 (>1 st deadline)	0.991	0.992	4	0.914	0.929	7	0.892	0.960	10
2 (>2 nd deadline)	0.992	0.994	3	0.964	0.978	6	0.944	0.981	7
3 (<deadline)	0.949	0.957	4	0.960	0.966	4	0.954	0.974	5
3 (>deadline)	0.989	0.994	3	0.971	0.974	4	0.976	0.985	4
4	-	-	-	-	-	-	-	-	-

F1 is F1 score
P is Precision
AF1 is approximate F1 score
AP is approximate Precision
Po is position



- * We used multiple APIs (without the Wikidata dump files)
- * Meta-lookup resolves even difficult mistakes in a word
- * Contextual matching gives F1-score above 0.97 in all tasks

