Table 1: Statistics of Entity Alignment datasets. ZH denotes Chinese, JA denotes Japanese, FR denotes french and EN denotes english.

Datasets	Language	Entities	Rel.	Attr.	Rel. triples	Attr. Triples	Align.
$DBP_{ZH\text{-}EN}$	ZH	19,388	1,701	7,780	70,414	379,684	15,000
	EN	19,572	1,323	6,933	95,142	567,755	
$DBP_{JA\text{-}EN}$	JA	19,814	1,299	5,681	77,214	354,619	15,000
	EN	19,780	1,153	5,850	93,484	497,230	
$DBP_{FR\text{-}EN}$	FR	19,661	903	4,431	105,998	528,665	15,000
	EN	19,993	1,208	6,161	115,722	576,543	
D_W_15K_V1	DBpedia	15,000	248	342	38,265	68,258	15,000
	Wikidata	15,000	169	649	42,746	138,246	
D_W_15K_V2	DBpedia	15,000	167	175	73,983	66,813	15,000
	Wikidata	15,000	121	457	83,365	175,686	
D_W_100K_V1	DBpedia	100,000	413	493	293,990	451,011	100,000
	Wikidata	100,000	261	874	251,708	687,860	
D_W_100K_V2	DBpedia	100,000	287	379	294,188	523,062	100,000
	Wikidata	100,000	32	38	400,518	749,787	

Table 2: Statistics of Knowledge Graph Alignment datasets. Mem-ST denotes memoryalpha-stexpanded, Star-SWT denotes starwars-swtor.

Datasets	Source	Inst.	Prop.	Class.	Triples.	Align.(Inst./class./prop.)
Star-SWT	Star Wars Wiki	145,033	700	269	8,246,033	1725/41/13
	The Old Republic Wiki	4,180	368	101	146,148	
Mem-ST	Memory Alpha	45,828	325	181	2,526,928	9295/53/14
	Star Trek Expanded Universe	13,426	202	283	567,386	

Statistics of Datasets

Table 1 and Table 2 summarize the statistics of the datasets used in our experiments. Following AttrGNN [2], we use the DBP15K version enriched with attribute values from the DBpedia dump (2016-10). We exclude OpenEA D–Y [5] due to its name bias, as almost all aligned entities in DBpedia and YAGO share the same names [3]. We also discard the SRPRS [1] and DWY100K [4] datasets, as they achieve perfect results using simple name-based heuristics [6].

References

[1] Guo, L., Sun, Z., and Hu, W. Learning to exploit long-term relational dependencies in knowledge graphs. In *International conference on machine learning* (2019), PMLR, pp. 2505–2514.

[2] Liu, Z., Cao, Y., Pan, L., Li, J., and Chua, T.-S. Exploring and evaluating attributes, values, and structures for entity alignment. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2020), pp. 6355–6364.

[3] Qi, Z., Zhang, Z., Chen, J., Chen, X., Xiang, Y., Zhang, N., and Zheng, Y. Unsupervised knowledge graph alignment by probabilistic reasoning and semantic embedding. In *International Joint Conference on Artificial Intelligence* (2021).

[4] Sun, Z., Hu, W., Zhang, Q., and Qu, Y. Bootstrapping entity alignment with knowledge graph embedding. In *IJCAI* (2018), vol. 18.

[5] Sun, Z., Zhang, Q., Hu, W., Wang, C., Chen, M., Akrami, F., and Li, C. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proceedings of the VLDB Endowment 13*, 11 (2020).

[6] Zhao, X., Zeng, W., Tang, J., Wang, W., and Suchanek, F. M. An experimental study of state-of-the-art entity alignment approaches. *IEEE Transactions on Knowledge and Data Engineering 34*, 6 (2020), 2610–2625.