

Small Description Text for Zenodo

WebNLG⁺⁺ is a harmonized and up-to-date version of the WebNLG⁺ test set, aligned with the September 2025 release of DBpedia.

It contains **1,779 texts** and **5,639 triples**, each manually reviewed and validated with LLM assistance to ensure consistency between ontology and entity URIs.

The dataset addresses changes in DBpedia entities and predicates, providing a benchmark that reflects the evolving knowledge base used by modern LLMs.

Three JSON files correspond to different use cases — updated triples, predicate-only updates, and full alignment with WebNLG⁺⁺ — plus one statistics file.

Files included

- webnlg_plus_mod_ground_truth.json
- webnlg_plus_mod_pred_ground_truth.json
- webnlg_plus_plus_ground_truth.json
- webnlg_plus_plus_statistics.xlsx
- README.md

Applications

- Evaluation of LLM-based triple extraction systems
- Study of ontology and URI evolution in DBpedia
- Temporal fact consistency and reasoning analysis

License: CC BY-SA 4.0 (???)

Keywords: WebNLG, DBpedia, triple extraction, LLMs, knowledge graph, benchmark