

Wikidata-Flashback - Characterising the Temporal Evolution of Wikidata^{*}

Diego Rincon-Yanez^{1,*,\dagger}, Elmar Kiesling^{1,\dagger} and Nicolas Ferranti^{1,\dagger}

¹WU – Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

Abstract

Knowledge graphs have become a fundamental component of modern neurosymbolic AI systems, providing structured, linked, and semantically rich representations of multi-domain knowledge. While general-purpose knowledge graphs underpin much of the Web’s contextual data infrastructure, they are typically maintained as monolithic updates or complex commit histories, disregarding point-in-time states. Consequently, there remains a critical gap in accessible tooling to extract and reproduce time-resolved versions of large-scale knowledge graphs. In this paper, we introduce *Wikidata-Flashback*, a lightweight extraction framework that generates arbitrary time snapshots from Wikidata’s update history. To validate the tool’s functionality, we produce yearly snapshots of the KG from the project’s start in 2013 through the end of 2024 and conduct a longitudinal graph analysis. By enabling precise temporal reconstruction and seamless integration with existing Semantic Web stacks, Wikidata-Flashback establishes a reproducible foundation for temporal KG research and time-aware AI systems.

Keywords

Wikidata, Knowledge Graph, Temporal Knowledge Graph, Knowledge Evolution

1. Introduction

Large Scale Knowledge Graphs (LSKGs) have played a decisive role in how the Web evolves and in how it provides services [1]. With NeuroSymbolic AIs (NeSyAIs) established as a reliable paradigm for delivering contextualized and hybrid AI solutions, LSKGs such as Wikidata[2] or DBpedia[3] serve as foundational sources of information, hosting over a billion triples (each) derived from collaboratively curated data.

These collaborative Knowledge Graphs (KGs) are not static; they evolve over time [4]. In Wikidata, for instance, every contributor edit is stored with a timestamp, resulting in a continuous stream of updates that change entities, introduce new properties, and correct inaccurate information [5]. Figure ref fig: knowledgebase shows an example of the complete Wikidata data model structure [6].

In 2025, DBpediaTKG [7] was released as an extension of the official DBpedia extraction process to generate temporal versions of DBpedia. The framework reconstructs temporal dynamics from Wikipedia history dumps, yielding over a billion triples and more than 200 million lifespans, enabling consistent time snapshots of DBpedia [7]. To the best of our knowledge, however, no analogous tool exists for Wikidata; existing approaches either rely on periodic full dumps that lack edit granularity or require ad-hoc reconstruction pipelines that are difficult to replicate.

To address this, we present *Wikidata-Flashback*, an open-source extraction framework for recreating snapshots of the Wikidata knowledge base at any given point in time. The framework transforms Wikidata’s continuous changes and community edits into a fully reproducible KG, along with their respective difference (delta) files for added and removed triples. To this end, the pipeline first ingests the edit stream into an intermediate event table composed of $(\mathcal{S}, \mathcal{P}, \mathcal{O})$ and the creation (t_c, u_c) and deletion (t_d, u_d) user and timestamp identifiers; the event database is then swept in an incremental process to produce the desired: (i) snapshot, (ii) addition and (iii) deletion deltas.

Wikidata’26: Wikidata Workshop at ISWC 2026

*Corresponding author.

^{\dagger}These authors contributed equally.

✉ diego.rincon@wu.ac.at (D. Rincon-Yanez); elmar.kiesling@wu.ac.at (E. Kiesling); nicolas.ferranti@wu.ac.at (N. Ferranti)

id 0000-0002-8982-1678 (D. Rincon-Yanez); 0000-0002-7856-2113 (E. Kiesling); 0000-0002-5574-1987 (N. Ferranti)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

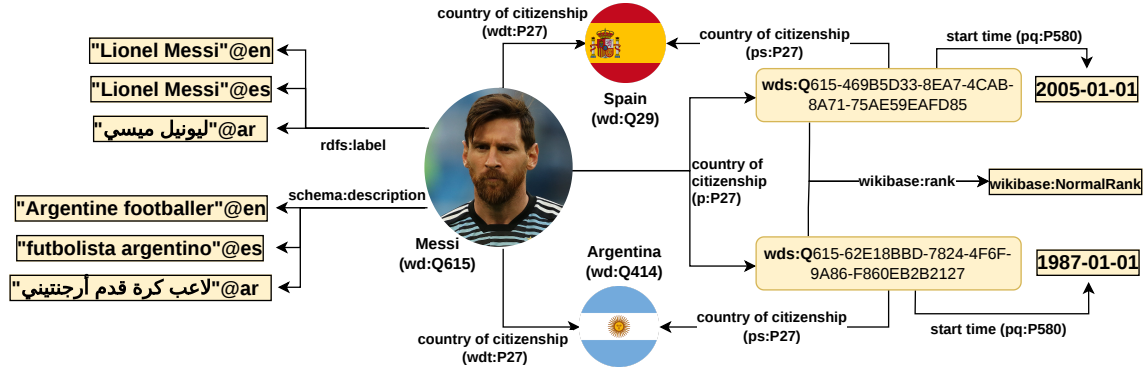


Figure 1: Wikidata structure, showing the knowledge base and the community editing logs

We evaluate Wikidata-Flashback by extracting annual snapshots of the entire history of Wikidata from its inception in 2013 through the end of 2024. Based on that, we analyse the Wikidata evolution and contrast it with that of DBpedia along multiple dimensions: First, we profile the KGs content, including the prevalence of entities and relations over time; Second, we study Wikidata’s evolution from a graph-analytic perspective along several key metrics. Third, we investigate growth patterns across Wikidata and contrast them with those of DBpedia. These results underscore the richness of Wikidata’s temporal dynamics and validate the utility of the presented contribution. Finally, the code to create the snapshots of Wikidata is available on GitHub¹

The resulting temporal graph aims to provide the Wikidata community and researchers with reliable, versioned snapshots of the Wikidata KG and its deltas, facilitating tasks such as temporal reasoning and incremental query optimisation, among others. The remainder of this paper is organised as follows: Section 2 reviews related work on KG snapshotting and relevant Wikidata studies; Section 3 details the architecture of Wikidata-Flashback; With yearly snapshots taken from Wikidata, Section 4.1 presents an analysis of the KGs, including graph characteristics comparable to a study with DBpedia. Section 4.3 depicts a longitudinal study of Wikidata’s evolution using the produced snapshots, exploring different growth metrics; finally, Section 6 discusses limitations and outlines future challenges.

2. Related Work

Extensive empirical work has analysed general aspects of Wikidata such as its quality [8], its interlinking within the Linked Data Ecosystem [9], and the evolution of community-driven schemas [10], but has typically largely disregarded the temporal dimension. Furthermore, there is also substantial work that studies the evolution of specific domains within Wikidata, such as the medical domain [11] or scientific publications [12].

Studies on the evolution of Wikidata as a whole are much more limited. Closest to our work, Wikidated [13] has similar objectives in that it aims to construct a dataset of Wikidata’s full revision history as sets of deletions and additions of RDF triples. However, the analyses in the paper are more limited, and the authors were not able to replicate the pipeline based on the published code in the repository.²

This work is also motivated by temporal datasets available for other KGs, most notably DBpedia, for which DBpedia-TKG [7] provides a framework for creating a sequence of temporal snapshots. Given the substantial differences like DBpedia and Wikidata, however, DBpedia-TKG works quite differently and re-extracts snapshots by re-running the whole DBpedia extraction pipeline on historic Wikipedia dumps.

¹GitHub Repository: <https://github.com/RyFoR/Wikidata-Flashback>

²The readme published in 2022 states that the authors are in the process of refactoring the code and preparing the dataset for initial public release, but no further information has been made available since then.

3. Temporal Snapshot Extraction for Wikidata

This section introduces the *Wikidata-Flashback* pipeline and explains how the raw edit history is transformed into temporal snapshots of truthy statements with explicit addition/deletion deltas. The framework was designed with specific key objectives: (i) *reproducibility*: for a given list of timestamps $\mathcal{T} = t_1 < t_2 < \dots < t_k$, multiple invocations of the pipeline should produce identical snapshots at respective time t_i , (ii) *scalability*: given its size, processing the Wikidata edit history in memory is generally not feasible on most hardware; the pipeline should therefore enable scalable processing (in a streaming manner with disk-backed event sweeps).

The pipeline proceeds in two stages:

1. **Event database construction** builds a statement-level lifecycle table from the raw page-revision dumps, stored in a DuckDB database
2. **Snapshot reconstruction**: performs an incremental sweep over this event table to yield full snapshots for arbitrary dates together with explicit addition/deletion artefacts.

These two steps deliver a pipeline that converts continuous Wikidata events into a sequence of discrete graph states.

Limitations: Note that we restrict snapshot extraction to truthy statements (see Section 3.1); the triples from the remaining auxiliary structure, such as discussion, edits, comments and unranked triples, are deliberately omitted from the temporal reconstruction.

3.1. Create the event database

The event database represents Wikidata’s evolution through *statement lifecycles*, that is, the statement-level histories of factual and schema-related statements. Each lifecycle records a qualified statement, the interval during which it was active, and the users responsible for its creation and deletion. Following the MAPL representation [14, 15], a statement can be expressed as $property(entity, value)@Q$, where the context set Q contains qualifiers, references, and rank information.

Drawing on Annotated RDF [16, 17], the Wikidata edit history is represented as a sequence of statement lifecycles of the form $property(entity, value)@Q[t_c, t_d, u_c, u_d]$. Here, t_c and t_d denote the creation and deletion times, while u_c and u_d identify the corresponding users. When a statement remains active at the end of the processed history, its deletion time and deleting user are left open. This event-based representation makes Wikidata’s temporal dimension explicit at the statement level and enables the active intervals and provenance of factual, schema, and constraint statements to be compared.

The event database is constructed from the Wikidata page meta-history dump of May 2025³. This dump that we use as input is organised as a sequence of page revisions rather than explicit statement histories. It is constructed by comparing consecutive revisions of each entity. RDF-level differences between revisions are used to identify statement creation and deletion events, which are then associated with the timestamp and editor of the corresponding revision. Processing the revisions in their recorded order provides a reproducible transformation from the selected historical dump to a statement-level event database.

The reconstruction accounts for Wikidata’s truthy rank semantics. Page revisions contain full statements and their ranks, whereas direct claims must be derived by selecting preferred-rank statements when they exist or normal-rank statements otherwise. The addition, deletion, or re-ranking of a full statement may therefore begin or end a truthy statement lifecycle, even when the underlying statement value is unchanged. The resulting records follow the schema:

$$(\mathcal{S}, \mathcal{P}, \mathcal{O}, t_c, u_c, t_d, u_d) \tag{1}$$

³<https://dumps.wikimedia.org/wikidatawiki/>

This provides the basis for reconstructing snapshots at arbitrary points in time and for deriving truthy-statement artefacts. An entity-to-file index supports targeted retrieval without requiring all lifecycle files to be scanned.⁴

3.2. Temporal Sweep and Snapshot Generation

The temporal sweep stage translates the event index built in the previous subsection into a set of temporally explicit graph states and corresponding change sets. At its core, the driver reads the event DB table and processes it against an ordered list of supplied timestamps. For each consecutive pair $(t_{\text{prev}}, t_{\text{curr}})$, the sweep algorithm applies all events whose validity intervals intersect the interval. The state after applying these transitions is materialised in three distinct CSV-based resulting triple artefacts: *Snapshot file* and *Deltas* (added/removed). Algorithm 1 specifies the sweep algorithm in more detail.

Algorithm 1 Sweep Algorithm for Snapshot and Delta Generation

Require: Event table E sorted by (t_c, t_d) ▷ c creation, d deletion
Require: Target timestamps $T = [t_1 < t_2 < \dots < t_k]$

- 1: $S \leftarrow \emptyset$ ▷ Current snapshot as a set of triples
- 2: $i \leftarrow 1$ ▷ Index into T
- 3: **procedure** SWEEP(E, T)
- 4: **for all** event $e \in E$ **do** ▷ Apply all events that belong to the current interval $(t_{i-1}, t_i]$
- 5: **while** $i \leq k$ **and** $(c_{e_t} \leq T[i] \vee d_{e_t} \leq T[i])$ **do**
- 6: **if** $c_{e_t} \leq T[i]$ **then**
- 7: $S \leftarrow S \cup \{(e_s, e_p, e_o)\}$
- 8: Record e_t in Δ_i^+
- 9: **end if**
- 10: **if** $d_{e_t} \neq \emptyset$ **and** $d_{e_t} \leq T[i]$ **then**
- 11: $S \leftarrow S \setminus \{(e_s, e_p, e_o)\}$
- 12: Record e_t in Δ_i^-
- 13: **end if**
- 14: **end while** ▷ Emit snapshot when we cross a target timestamp
- 15: **if** $i = k$ **or** $c_{e_t} > T[i]$ **then**
- 16: WRITE_SNAPSHOT($S, T[i]$) ▷ Compressed N-Triples
- 17: WRITE_DELTA($\Delta_i^+, \Delta_i^-, T[i-1], T[i]$)
- 18: $i \leftarrow i + 1$
- 19: **end if**
- 20: **end for**
- 21: **end procedure**

When invoked, the framework skips constructing the events table and instead performs the transition sweeps directly, thereby reducing I/O overhead on subsequent runs. After completing all intervals. In summary, the event sweep stage systematically materialises both full graph states and their associated change sets from an efficient event representation.

4. Characterising the Evolution of Wikidata

Based on the yearly snapshots extracted from Wikidata as described in Section 3, this section analyses temporal changes within Wikidata over the course of its history from inception until the end of 2024. We first track its general volume, growth, and addition/deletion dynamics in Section 4.1, then characterise

⁴https://anonymous.4open.science/r/wikidata_statements_lifecycle_machine-77B0/

its evolution from a network structural perspective in Section 4.2, and finally focus on *what* drove the growth in Section 4.3 by investigating entity-role composition and predicate usage patterns.

4.1. Growth

Yearly Growth Figure 2 plots the yearly growth in terms of number of triples, unique entities, unique types (*p31*), and unique relations on a logarithmic scale, showing that Wikidata mainly scaled by adding facts about more items under a relatively stable vocabulary. Whereas the number of triples grew from 26 million to 1.43 billion, i.e., a factor of approx. 56 from 2013 to 2024, the number of predicates (unique relations) grew by a factor of 15 (from 780 to 12,000), i.e., the schema remained fairly compact as the instance layer of the graph exploded. Also evident in the graph is a phase of rapid expansion from 2016 to 2017 (from 121M to 359M triples), which marks the start of massive scholarly imports that also shifted the graph composition considerably, as we will discuss in Section 4.3.

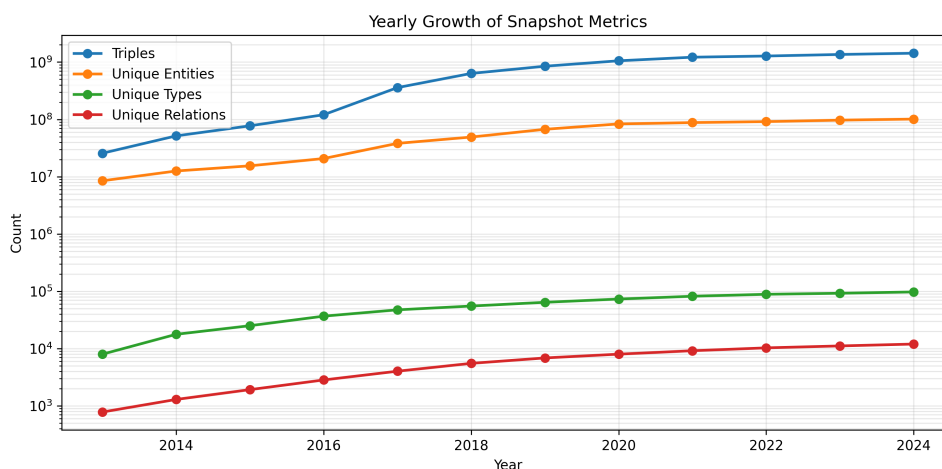


Figure 2: Yearly growth over the extracted snapshots

Triple Changes Figure 3 focuses on triple additions and deletions, showing high net growth every year, but with marked variations in three phases: (i) an early stage of continuous organic growth until 2016, followed by (ii) a phase of rapid expansion from 2017-2021, and (iii) a mature phase with relatively constant additions (80-96M p.a.) and rising removals (14-21M). Relative to additions, the share of deletions is rising, from around 2% in 2017 to approx. 23% in 2024 as the graph matures. Next, we will look at triple lifespans in more detail.

Triple Lifespan Figure 4 illustrates the distribution of triple lifespans, obtained by pairing create and delete events on the revision-level. Approx. 10% of the added triples – presumably reverts, bot errors, etc. – are removed within an hour; we see a near-year peak (59%) of triples removed in the 11-12 months interval. Because only triples that were eventually deleted are included in the analysis (i.e., statements that are still valid were not counted), we can conclude from the low number of triples deleted after 1 year that a large share that is not deleted after approx. one year is highly permanent.

4.2. Network structure

This section covers the temporal evolution from a network-analytic perspective – specifically discussing connectivity metrics and component analysis.

Out-Degree: Figure 5 shows that the rapid growth was not only due to an expansion of the covered entities, but also because these entities become more richly described over time. The out-degree

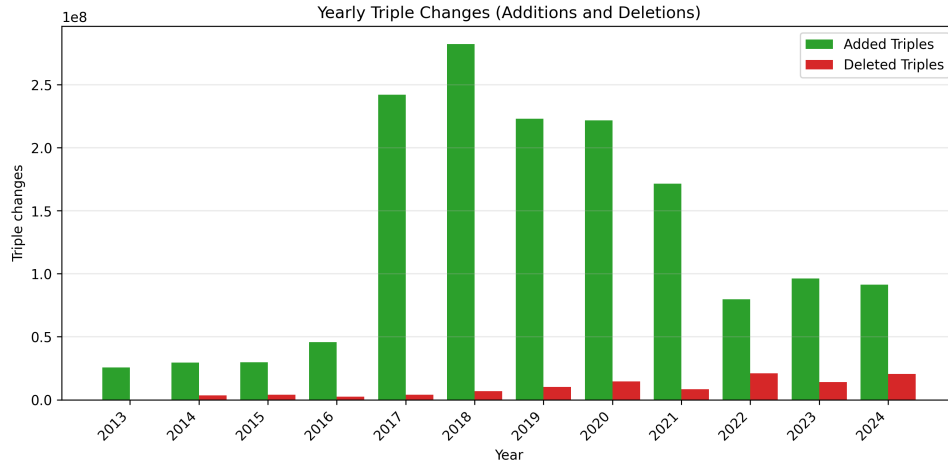


Figure 3: Triple Changes over Time

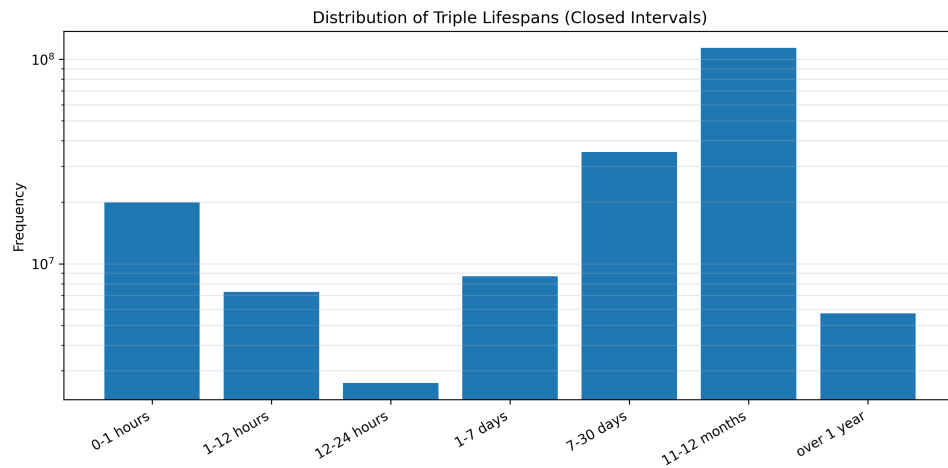


Figure 4: Triple Lifespan Distribution

distribution over all nodes plotted here includes literals. It shows that in the early stage of the evolution (2013), the share of items with an out-degree of approximately 1 was large, with maximum degrees on the scale of hundreds. In the period from 2017-2019, i.e., the scholarly import boom already identified in Section 4.1, the mass shifts right with a shoulder at approx. 5-20 as items gain a standard property set. In more recent years (2020-2024), the fraction of nodes with degree 1 is much lower, and we more commonly see hubs with approx. 10^3 statements, with some even exceeding 10^4 statements. Overall, the shape remains heavy-tailed over the evolution and exhibits scale-free properties. The maximum out-degree (cf. Fig. 6a) – excluding literals and focusing only on the pure entity-relation graph – has also grown strongly, but plateaued near approx. 8.300 entity relations since 2018 for the richest subjects.

In the following network-centric analyses, we will continue to focus exclusively on the pure relational knowledge graph, i.e., the entity-relation graph without literals.

In-Degree: Heavy-tailed hub scale is also evident in the maximum in-degree in Figure 6b, which is generally very large, exhibiting jumps especially after 2016 (from 3.2M in 2016 to 12M in 2017 to 40M in 2024). This is consistent with highly popular targets (classes, countries, journals, etc.) accumulating links.

Average degree and density: Figure 7a shows that the average degree rises from 4.4 to 14.3, i.e., the URI graph becomes considerably denser per node, which matches the out-degree shift we saw above.

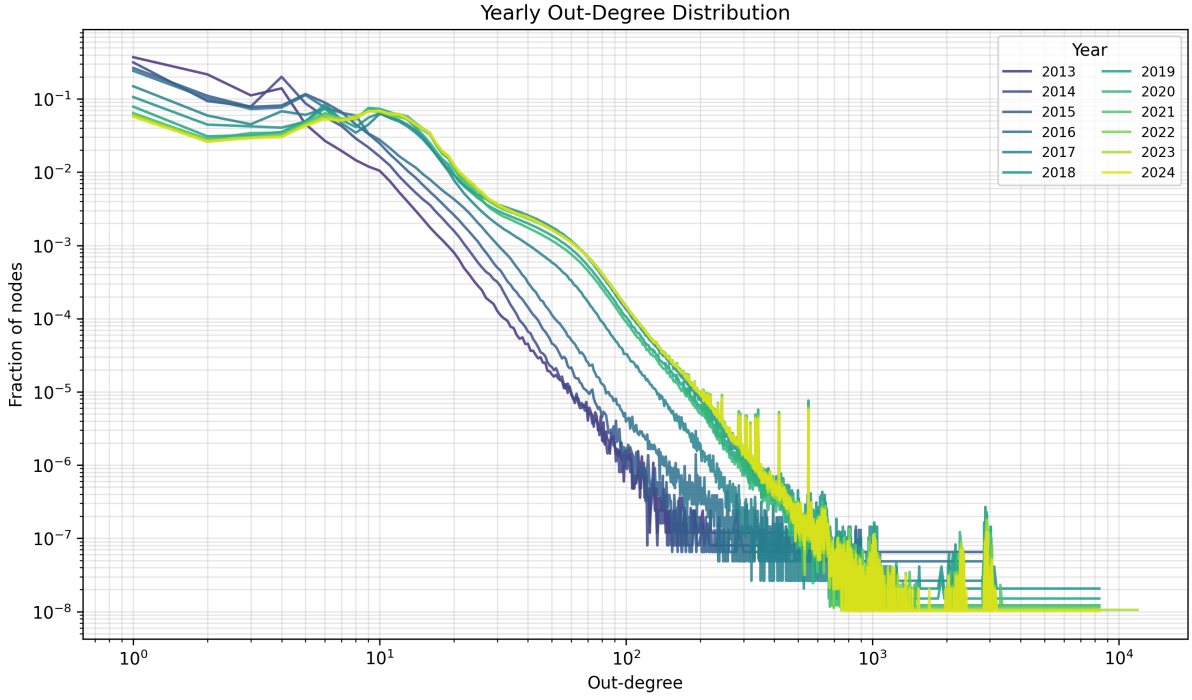


Figure 5: Wikidata out degree distribution

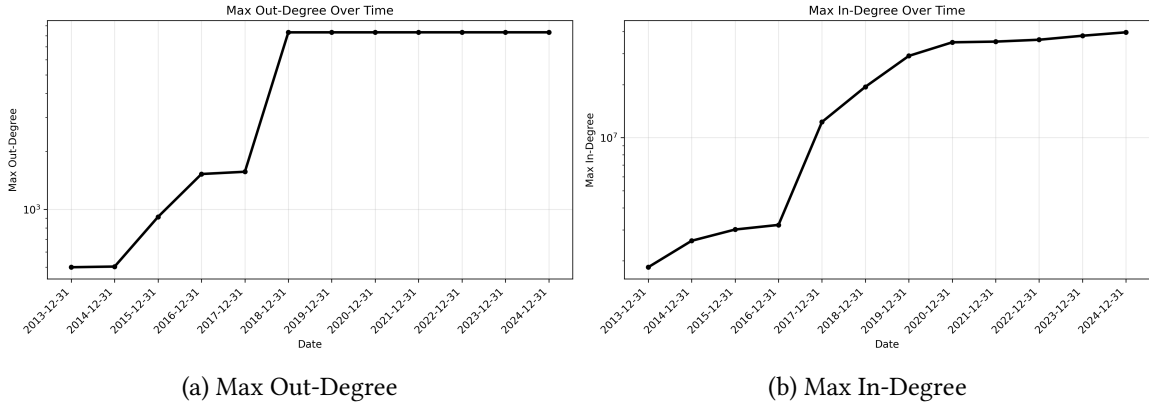


Figure 6: Max In and Out-Degrees over Time

One noteworthy exception occurred in the interval 2019-2020, where avg. degree temporarily dropped significantly as the addition of nodes outpaced the addition of new edges; from 2021 on, we can see that enrichment has caught up and average degree is stable since then. Global density – illustrated in Figure 7b – has fallen sharply after inception and has been relatively stable since 2021. This is typical for sparse KGs that get larger, as possible pairs grow as $O(n^2)$ while edges grow more slowly.

Components: Compared to the overall size of the graph, Figure 8a shows that Wikidata has essentially always been, and continues to be, a single giant weakly connected component (> 99.9% of nodes every year). Compared to the overall size of the graph, the number of other weakly connected components (cf. Fig. 8b) is tiny and has fallen.

4.3. Profiling the Wikidata Evolution

After investigating the global growth and structural evolution in a network-analytic sense, this section shifts focus to how Wikidata evolved in terms of content and composition.

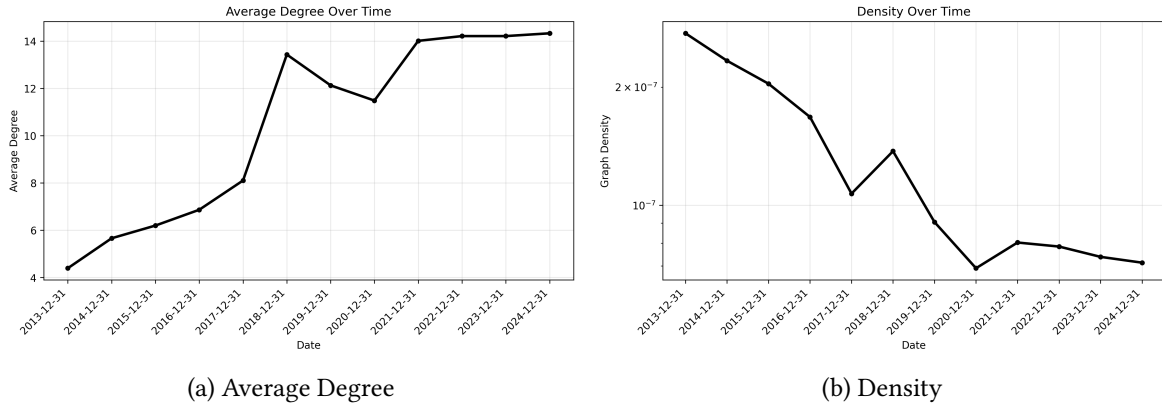


Figure 7: Average Degree and Density over Time

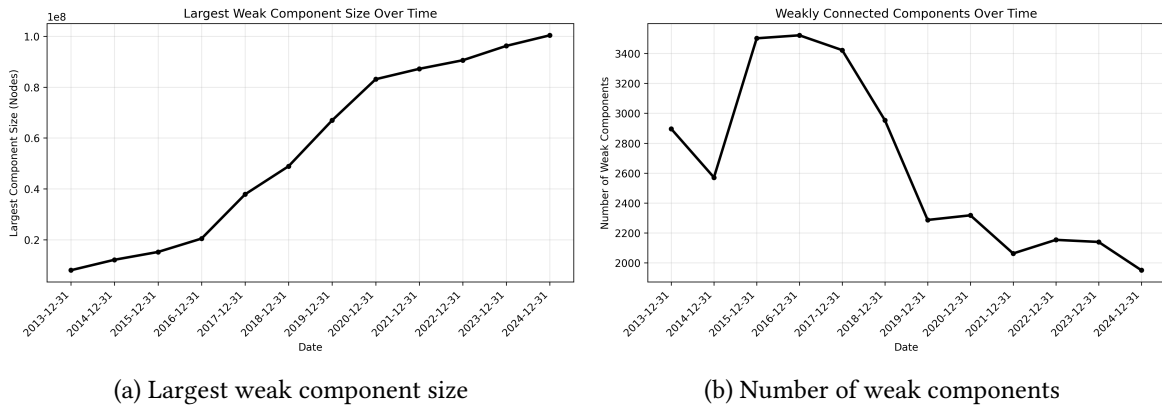


Figure 8: Weak Components over Time

Subjects vs object URI vocabulary To get a more nuanced understanding of how items appear in the graph, we can start by looking at the frequency of URIs appearing in subjects vs in object position. Figure 9 illustrates this over time, showing that subjects always dominate, i.e., that most items are described more than they are referenced, which seems natural. There are two notable sharp increases in object position URIs from 2017-2018 and 2020-2021 – likely as citation graphs and publication entities that are frequently referenced are added, which can also be seen in the sharp increase in the usage of P2860 (cites work) below.

Predicate prevalence: Until 2016, the most widely used predicates were typing and place-related – i.e., *instance-of* (P31), *country* (P17), *admin location* (P131), *coordinate location* (P625), etc. – as can be expected in a general knowledge base. From 2017 on, however, we can see a pronounced shift towards “scholarly” properties, with *cites work* (P2860) and *author name* (P2093) quickly becoming the most widely used. Fig. 10b highlights this by summarising the fraction of all triples using the respective property, showing that the citation network created by *cites work* (P2860) alone makes up a large part of the graph, peaking at > 25% in 2019 and > 20% in 2024. Taken together, “scholarly” properties in the top 15 made up approx. 48% of all triples in 2024. Fig. 10b

5. Wikidata vs DBpedia: Comparison

In the following, we compare Wikidata and DBpedia-TKG [7] in terms of growth, schema evolution, addition and deletion dynamics, and triple lifespan.

Note that DBpedia-TKG [7] is a temporal version of DBpedia that has been extracted ex post from

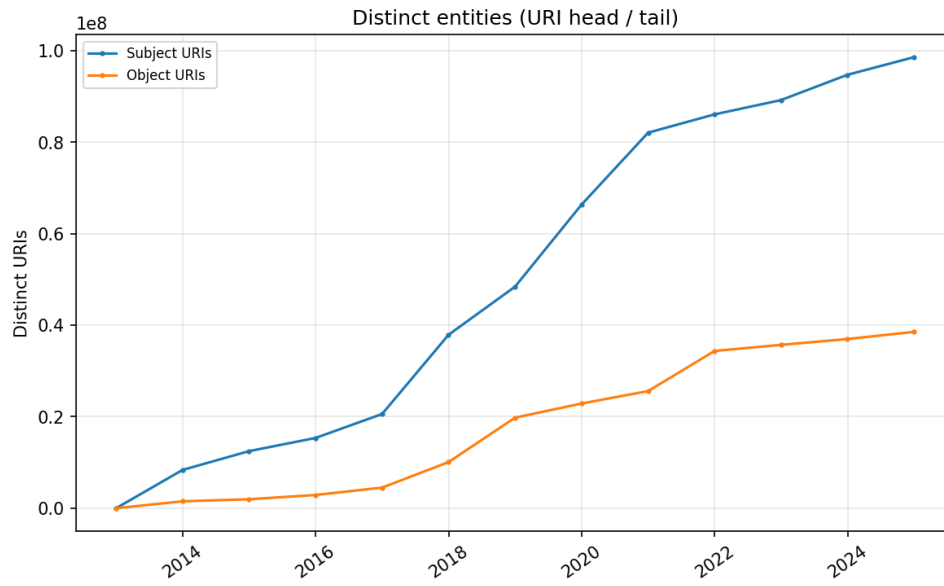
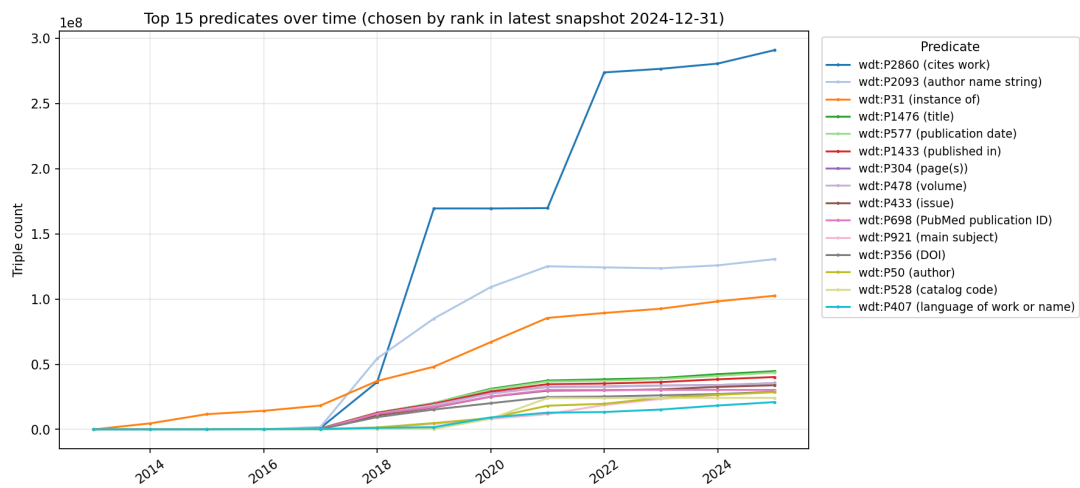
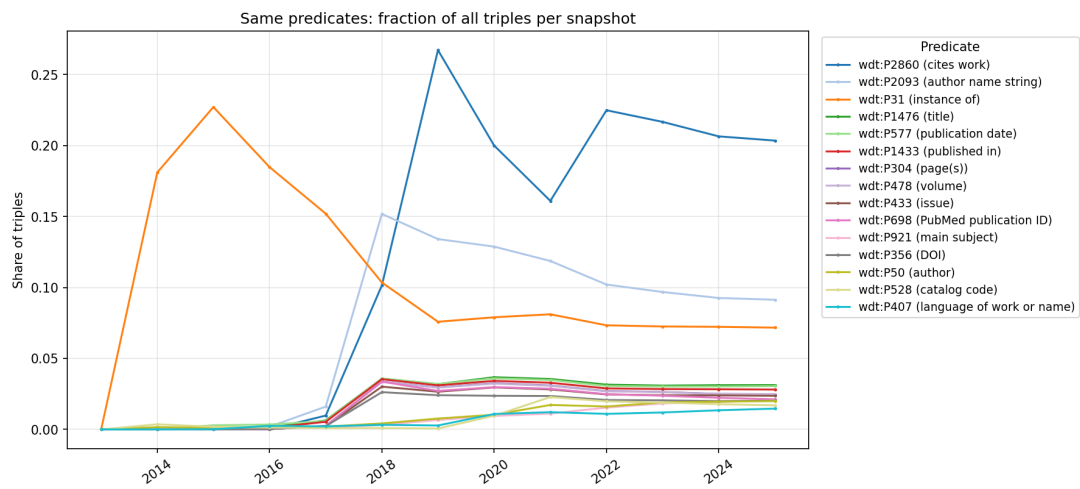


Figure 9: Distinct Subject and Object Position Entities over Time



(a) Top 15 predicates (2024) frequency



(b) Top 15 predicates (2024) (fraction of all triples)

Figure 10: Predicate prevalence over time

historic Wikipedia snapshots with the DBpedia extractors and schema as of 2024. It does therefore not match the historic DBpedia snapshots published, but reflects the evolution of the Wikipedia-extracted knowledge from a 2024 perspective.

Growth DBpedia was initiated in 2007; the DBpedia-TKG shows strong growth in the early years, which had already levelled off somewhat by 2012 when the Wikidata project started. Wikidata shows massive growth from its inception, especially in the 2017-2020 period, when it crossed DBpedia in terms of total number of triples (2018), reaching 1.43B vs 595M in 2024. In terms of unique entities covered, the gap is even wider (approx. 101M vs 24M). We can also see that in recent years, DBpedia’s entity growth is nearly linear, whereas DBpedia keeps adding more items at a faster pace.

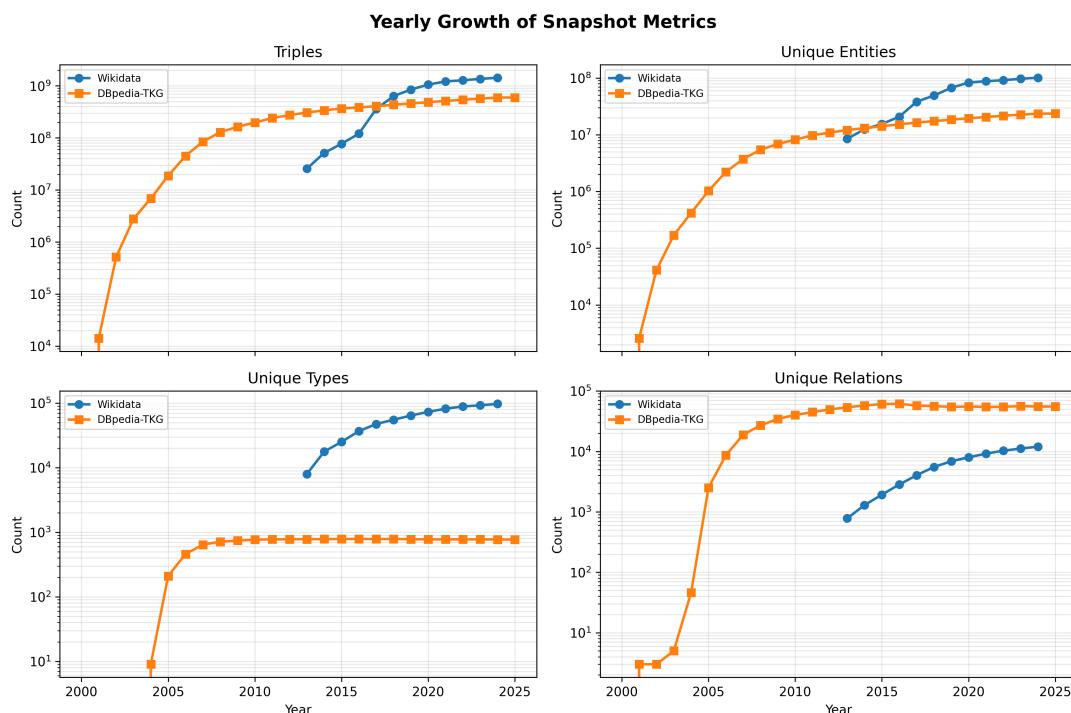


Figure 11: Yearly growth over the extracted snapshots: Wikidata vs DBpedia-TKG

Yearly additions and deletions Compared to Wikidata’s somewhat volatile but overall much higher add rate that peaks strongly in 2018 (282M p.a.) and has since settled at around 90M annual additions, DBpedia’s add rate is much flatter at approx. 44-48M p.a. since 2013.

In terms of removals, DBpedia has long exhibited high absolute deletion activity (14–28M/year, often approx. 30–60% of that year’s adds), while Wikidata deletions start near zero and only later approach DBpedia’s absolute level (approx. 20M) — still a smaller fraction of Wikidata’s peak adds.

Overall, we can see that DBpedia grows steadily with substantial deletions as articles are edited. In contrast, Wikidata’s addition and deletion dynamics are more import-driven, with large jumps and a relatively lower rate of deletion.

6. Conclusions and Future Work

In this work, we present Wikidata-Flashback, an open-source tool that reconstructs arbitrary point-in-time snapshots of the Wikidata knowledge graph and materialises them together with precise addition and deletion deltas, pruning the community aspect and concentrating on truthy statements only. These resulting yearly artefacts can be versioned, shared, and reused without reprocessing raw

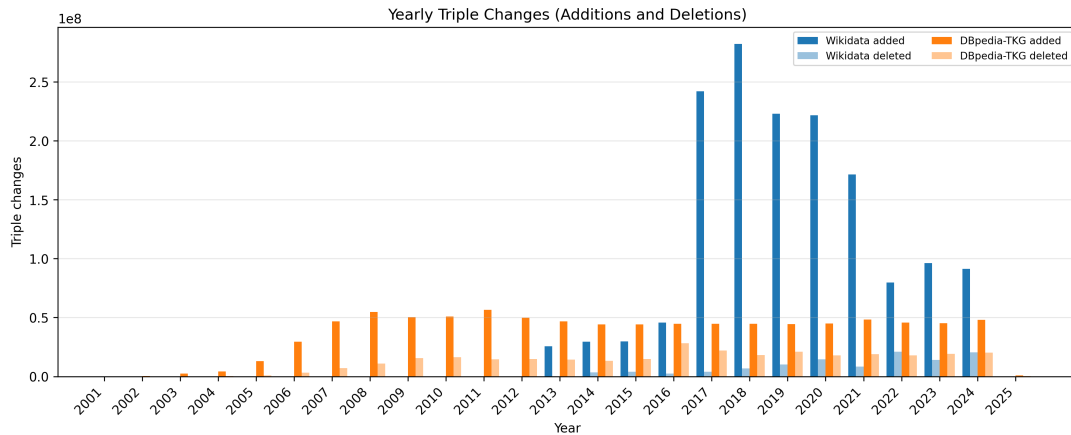


Figure 12: Yearly triple additions and deletions over time

dumps, thereby providing researchers with clean, reproducible data for a wide range of temporal KG studies.

Furthermore, a deep longitudinal analysis of these snapshots reveals clear growth trends in triples, entities, and property usage, offering new insights into the evolving structure of this community-built knowledge base and challenging the community at large to explore how knowledge evolution affects modern AI systems.

This work opens several promising research avenues. In particular, Wikidata-Flashback enables the construction of temporal-KG benchmarks that capture the full evolution of facts and facilitate systematic studies of knowledge dynamics. On the other hand, the remaining event-level comments, discussions and other annotated data could be exploited as a resource for mining how community contributions shape a large-scale graph, revealing motivations behind edits, patterns of change over time and the socio-technical forces driving the continual growth of Wikidata.

Acknowledgments

The work on this paper has been funded by the Austrian Science Fund (FWF) 10.55776/COE12.

Declaration on Generative AI

During the preparation of this work, the author(s) used Grammarly to check grammar and spelling. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

References

- [1] A. Polleres, F. Bauer, D. Dobriy, T. Käfer, L. Kubelka, A. Harth, T. Wehr, On the Historic Roots of Agentic AI in Semantic Web Services, in: *WWW Companion 2026 - Companion Proceedings of the ACM Web Conference 2026*, ACM, New York, NY, USA, 2026, pp. 10–14. doi:10.1145/3774905.3795076.
- [2] D. Vrandečić, M. Krötzsch, Wikidata: A free collaborative knowledgebase, *Communications of the ACM* 57 (2014) 78–85. doi:10.1145/2629489.
- [3] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, DBpedia: A nucleus for a Web of open data, in: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 4825 LNCS, Springer, Berlin, Heidelberg, 2007, pp. 722–735. doi:10.1007/978-3-540-76298-0_52.

- [4] A. Polleres, R. Pernisch, A. Bonifati, D. Dell’Aglia, D. Dobriy, S. Dumbrava, L. Etcheverry, N. Ferranti, K. Hose, E. Jiménez-Ruiz, M. Lissandrini, A. Scherp, R. Tommasini, J. Wachs, How Does Knowledge Evolve in Open Knowledge Graphs?, *Transactions on Graph Data and Knowledge* 1 (2023) 11:1–11:59. doi:10.4230/TGDK.1.1.11.
- [5] N. Ferranti, D. Guimarães, J. F. de Souza, A. Polleres, Formalizing repairs for wikidata constraint violations: A taxonomy and empirical analysis, in: *The Semantic Web – ISWC 2025*, Springer Nature Switzerland, Cham, 2026, pp. 369–387.
- [6] N. Ferranti, J. F. De Souza, S. Ahmetaj, A. Polleres, Formalizing and validating wikidata’s property constraints using shacl and sparql, *Semantic Web* 15 (2024) 2333–2380.
- [7] M. Hofer, M. M. Töpfer, C. Rost, E. Rahm, Dbpedia-tkg: Capturing wikipedia’s evolution as temporal knowledge graphs, in: *The Semantic Web - 22nd European Semantic Web Conference, ESWC 2025, Portoroz, Slovenia, June 1-5, 2025, Proceedings, Part II*, volume 15719 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 262–279. doi:10.1007/978-3-031-94578-6_15.
- [8] K. Shenyoy, F. Ilievski, D. Garijo, D. Schwabe, P. Szekely, A study of the quality of wikidata, *Journal of Web Semantics* 72 (2022) 100679. URL: <https://www.sciencedirect.com/science/article/pii/S1570826821000536>. doi:<https://doi.org/10.1016/j.websem.2021.100679>.
- [9] A. Haller, A. Polleres, D. Dobriy, N. Ferranti, S. J. Rodríguez Méndez, An analysis of links in wikidata, in: P. Groth, M.-E. Vidal, F. Suchanek, P. Szekley, P. Kapanipathi, C. Pesquita, H. Skaf-Molli, M. Tamper (Eds.), *The Semantic Web*, Springer International Publishing, Cham, 2022, pp. 21–38.
- [10] S. Baroncini, M. Scrocca, M. Martorana, Z. Śmiech, A. Polleres, et al., Analysing the evolution of community-driven (sub-) schemas within wikidata, in: *Proceedings of the 3rd Wikidata Workshop 2022 co-located with the 21st International Semantic Web Conference (ISWC2022)*, volume 3262, *CEUR Workshop Proceedings*, 2022, pp. 1–8.
- [11] T. Shafee, G. Masukume, L. Kipersztok, D. Das, M. Häggström, J. Heilman, Evolution of wikipedia’s medical content: past, present and future, *J Epidemiol Community Health* 71 (2017) 1122–1129.
- [12] Z. Saletti, C. Consonni, P. Frau-Amar, E. Gómez, Divinwd: Exploring the diversity of scientific publications in wikidata, in: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 20, 2026, pp. 2895–2909.
- [13] L. Schmelzeisen, C. Dima, S. Staab, Wikidated 1.0: An evolving knowledge graph dataset of wikidata’s revision history, *arXiv preprint arXiv:2112.05003* (2021).
- [14] D. L. Martin, P. F. Patel-Schneider, Wikidata constraints on MARS, in: *Proceedings of the 1st Wikidata Workshop (Wikidata 2020) co-located with 19th International Semantic Web Conference (OPub 2020)*, Virtual Conference, November 2-6, 2020, volume 2773 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: <https://ceur-ws.org/Vol-2773/paper-12.pdf>.
- [15] M. Marx, M. Krötzsch, V. Thost, Logic on MARS: ontologies for generalised property graphs, in: C. Sierra (Ed.), *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, ijcai.org, 2017, pp. 1188–1194. doi:10.24963/IJCAI.2017/165.
- [16] O. Udrea, D. R. Recupero, V. S. Subrahmanian, Annotated rdf, *ACM Trans. Comput. Logic* 11 (2010). doi:10.1145/1656242.1656245.
- [17] N. Lopes, A. Polleres, U. Straccia, A. Zimmermann, AnQL: SPARQLing up annotated RDFS, in: *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, volume 6496 of *LNCS*, Springer, Shanghai, China, 2010.