

MultiSummGRAG: A Locally-Hosted GraphRAG Pipeline for the MultiSumm Task at MediaEval 2026.

Kimia Rabishokr^{1,*}, Andreas Weiler¹

¹IIS-ZHAW Team - Zurich University of Applied Sciences ZHAW, Switzerland

Abstract

This paper describes MultiSummGRAG, our approach for the MediaEval 2026 MultiSumm task on multimodal summarization of Food Sharing Initiatives across five European cities. Our approach is a six-phase, fully reproducible pipeline that generates structured HTML/PDF multimodal reports from a community-structured knowledge graph (GraphRAG) built over each city’s corpus, with explicit handling of non-English sources in the subtask. Paired with a strong synthesis model, GraphRAG produces grounded, geographically specific summaries that name real initiatives and active districts, including across the non-English subtask cities. Our analysis further reveals three main drivers of output quality: synthesis-model choice dominates over source language, GraphRAG requires a minimum corpus density to avoid hallucinated entities, and the multilingual bottleneck lies in upstream English-centric keyword components rather than in LLM language understanding.

1. Introduction

The MultiSumm 2026 task requires participants to generate structured, multimodal summaries of Food Sharing Initiatives (FSIs) from the ShareCity200 dataset for English-speaking main-task cities (Dublin, Brighton & Hove) and non-English subtask cities (Barcelona, Milan, London). Our system is a six-phase automated pipeline that processes each city’s FSI URL list into a rendered HTML/PDF report using only locally hosted open-weight models served via Ollama. The key architectural choice is a community-structured knowledge graph (GraphRAG) built over the full city corpus, rather than querying individual documents.

2. Related Work

At the first edition of MediaEval MultiSumm [1], Dhannya et al. [2] chunked scraped pages and summarized each segment with BART, merging per-chunk outputs into a city summary with optional translation for non-English content. Potyagalova and Jones [3] expanded the verified URL list via Google Search and produced summaries by answering a predefined question set over the extracted text with an LLM. Our pipeline builds on Wu et al.’s LLM-based retrieval and classification of FSIs [4], extending per-FSI labelling to corpus-wide synthesis driven by six predefined queries via the GraphRAG paradigm of Edge et al. [5], implemented in nano-graphrag [6]. This intermediate structured representation of the city corpus targets the global sensemaking queries MultiSumm requires.


MediaEval’26: Multimedia Evaluation Workshop, June 15–16, 2026, Amsterdam, Netherlands and Online

*Corresponding author.

✉ rabi@zhaw.ch (K. Rabishokr); wele@zhaw.ch (A. Weiler)



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

 CEUR Workshop Proceedings (CEUR-WS.org)

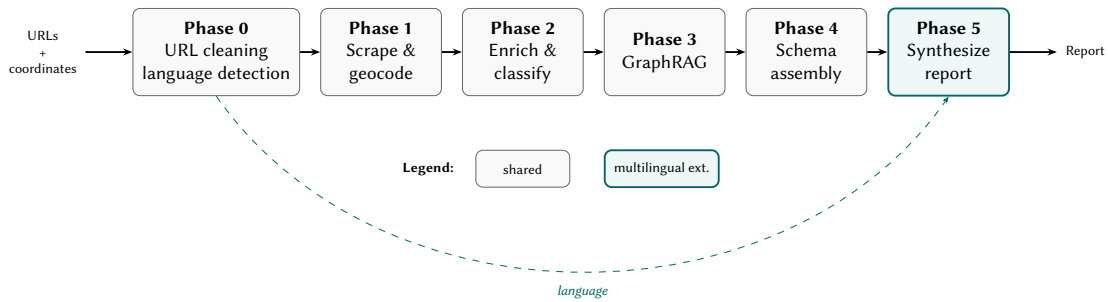


Figure 1: MultiSummGRAG pipeline (Phases 0–5). Gray boxes denote stages shared by both workflows; the teal box marks the multilingual extension. Phase 0 detects the source language of each URL using a character n-gram classifier (`langdetect`). Phases 1–4 are identical across both variants. Phase 5 is where the two branches diverge: in the multilingual variant the detected language is passed to the report synthesizer, which adjusts the LLM prompt to acknowledge non-English source material.

3. Approach

Our pipeline processes FSI web content through six sequential phases (Figure 1): URL preprocessing, data collection, geographic enrichment, knowledge graph construction, schema assembly, and report generation.

3.1. Data Collection and Preprocessing

Per-city input CSVs are cleaned before scraping by removing exact duplicates, login-walled social-media URLs, and dead links, and by geocoding coordinate-free entries via Nominatim with a flagged fallback to city-centre coordinates.

Because FSI websites are frequently JavaScript-heavy or WordPress-hosted, we scrape with headless Chromium via Playwright and `playwright-stealth` (concurrency 3), merging up to five same-domain subpages per seed URL; WordPress sites are additionally queried via their REST API as a higher-fidelity fallback. Text is extracted with Trafilatura; images are filtered to $\geq 200 \times 200$ px and deduplicated city-wide via perceptual hashing (pHash).

3.2. Geographic Enrichment and FSI Classification

Each FSI is assigned to a city district via point-in-polygon intersection against official administrative boundary GeoJSONs; FSIs falling outside every polygon are snapped to the nearest district centroid. Classification of FSI type and operational level is performed by an LLM at `temperature=0`; model choices and supported languages are listed in Table 1. For subtask cities, the system prompt includes multilingual term mappings so that classification is meaning-based regardless of source language. Popularity is then estimated via a weighted English keyword scoring scheme and ranked into city-wide percentile tiers (*high/medium/low*); keyword sparsity for non-English cities is the primary driver of the bimodal popularity distributions seen in Table 2.

3.3. Knowledge Synthesis and Report Generation

We build a community-structured knowledge graph with nano-graphrag, enabling both corpus-wide (global) and entity-specific (local) retrieval. Entity and relationship extraction uses `llama3.2-graphrag` across all cities; community summarization and query answering use

Table 1

Per-city configuration. Boundary GeoJSONs are in WGS84 (EPSG:4326); The pipeline normalizes CRS at load time. ML Prompt indicates a multilingual term-mapping system prompt for FSI classification.

Task	City	# Districts	Boundary	Classifier	ML Prompt
Main	Dublin	35	[7]	11ama3.2	-
	Brighton & Hove	23	[8]	11ama3.2	-
Subtask	London	33	[9]	qwen3:14b	EN
	Barcelona	10	[10]	qwen3:14b	EN/ES/CA
	Milan	9	[11]	qwen3:14b	EN/IT

qwen3:14b for sub-task cities and 11ama3.2-graphrag for main task cities (following the same task split as Table 1). Semantic embeddings use `nomic-embed-text` with 1800-token chunks (overlap 100 for sub-task, 50 for main task). Six predefined queries cover geographic distribution, initiative types, operational models, reach and activity, per-district summaries, and notable initiatives.

A schema assembly step aggregates enriched records by district, computing FSI counts and type breakdowns, and assigning green/yellow/red density tiers by city-wide percentile. Keyword-based fact extraction across ≈ 20 dimensions (sustainability, education, target audience, collaboration, digital and physical presence) produces a verified numeric facts dictionary. Each of the six report sections is synthesized at `temperature=0` via a two-part prompt that separates verified `DATA POINTS` from graph-derived `CONTEXT`; a post-hoc number check flags any figure absent from the facts dictionary. Three visualizations (district density map, horizontal bar chart, FSI-type donut chart) and four notable-initiative images are embedded as base64 assets; the report is assembled with Jinja2 and exported to PDF via WeasyPrint.

4. Results and Analysis

Table 2 summarizes outputs across all five cities, covering 635 FSI records in total. The sub-task cities achieve full or near-full district coverage, while Brighton & Hove is a striking outlier with 76 FSIs concentrated in just 3 of 23 council wards. English-speaking Cities show clean label distributions dominated by `food_sharing` and near-zero unknown rates, whereas multilingual cities carry markedly higher unknown operational-level shares — a direct consequence of keyword-based classification degrading on non-English text. Milan is the only city where NGO-led initiatives predominate; all others are strongly community-led. The popularity columns reveal a further cross-lingual gap: English cities spread across all three tiers, while Barcelona and Milan collapse to a near-binary high/low split due to the English-only popularity keyword set.

GraphRAG output quality varied substantially with corpus size, language, and synthesis model. London, Barcelona, and Milan (using qwen3:14b) produced rich, geographically specific answers naming real initiatives and active districts. Dublin and Brighton & Hove performed considerably weaker, with both using 11ama3.2-graphrag for synthesis and Dublin’s small corpus (69 FSIs) producing sparse inter-entity edges that led to hallucinated organization names. Additional failure modes include Barcelona’s 39% unknown operational level, reflecting a genuine taxonomic mismatch between the UK/Irish-oriented schema and Catalan solidarity-economy terminology, and a partial chain-of-thought leak in Milan’s district summaries, where qwen3:14b’s internal `<think>` reasoning blocks were not consistently stripped before being inserted into the report.

Table 2

Pipeline output statistics per city. Dominant type and operation are the modal label with share in parentheses. Unk. = unknown label share. Popularity tiers: H / M / L (%). Abbreviations: *fs* food_sharing, *cg* community_garden, *fb* food_bank; *cl* community_led, *nl* ngo_led.

City	FSIs	Dist.	Dominant		Unknown (%)		Popularity (%)		
			Type	Op	Type	Op	H	M	L
Dublin	69	19/35	<i>fs</i> (49%)	<i>cl</i> (67%)	–	–	32	29	39
Brighton & Hove	76	3/23	<i>fs</i> (67%)	<i>cl</i> (83%)	–	–	26	36	38
London	106	26/33	<i>cg</i> (17%)	<i>cl</i> (41%)	21	33	25	43	31
Barcelona	237	10/10	<i>cg</i> (17%)	<i>cl</i> (31%)	–	39	53	0	47
Milan	147	9/9	<i>fb</i> (18%)	<i>nl</i> (43%)	–	27	71	0	29

5. Discussion and Outlook

Our approach proved capable of producing the corpus-level synthesis the task requires; output quality, however, was strongly determined by two factors treated as secondary during development: synthesis model choice and per-city corpus size.

GraphRAG enables grounded corpus-level synthesis. When paired with qwen3 : 14b on a sufficiently dense corpus, the community-graph approach produced district-specific, initiative-named answers across both English and non-English cities. The community summarization step in particular delivered coherent district-level paragraphs without per-document prompting, validating the corpus-level framing of the task.

The synthesis model dominates output quality. Across cities, qwen3 : 14b produced grounded, initiative-specific answers while llama3 . 2-graphrag hallucinated organizations, and this gap tracked the model rather than the source language.

GraphRAG requires a minimum corpus density. Dublin’s 69 documents produced a graph too sparsely connected for meaningful synthesis. Below roughly 100 FSI records, direct vector similarity search over enriched records is likely more reliable than GraphRAG; future work should apply GraphRAG selectively by corpus size.

Keyword-based components fail silently on non-English content. The bimodal popularity distributions for Barcelona and Milan illustrate a general failure mode: percentile normalization compensates for uniformly sparse signals but not when sparsity is systematically correlated with source language. An LLM-based activity assessment would be language-agnostic by construction.

FSI taxonomy mismatch for cooperative economies. Barcelona’s high unknown rates reflect that Catalan and Spanish food cooperatives and agroecological networks do not map onto a schema developed from UK/Irish food aid contexts. Multilingual FSI research should develop culturally-grounded taxonomies before applying a uniform classification pipeline.

Outlook. The three most impactful near-term changes are: qwen3 : 14b for synthesis across all cities; an LLM-based popularity scorer; and corpus-size-gated GraphRAG. More broadly, the bottleneck in multilingual FSI extraction is rarely language understanding since modern LLMs handle Spanish, Italian, and Catalan content well, but the upstream keyword-based components were designed with English in mind. The pipeline’s reliance only on locally hosted open-weight models also suggests applicability to the wider ShareCity200 collection and to other domains where centralized API access is undesirable for cost, privacy, or reproducibility reasons.

Acknowledgements

This work was carried out at the Zurich University of Applied Sciences (ZHAW) for the MediaEval2026 MultiSumm task. The authors thank the MultiSumm 2026 task organizers for providing the ShareCity200 dataset, and for their support throughout the challenge.

Declaration on Generative AI

The pipeline implementation, experimental design, scientific claims and writing draft for the paper are the work of the authors. The authors have used Claude (Anthropic) to assist with debugging of the pipeline, help with formatting LaTeX tables and LaTeX figures, help with text shortening, and proofreading. The authors have also used Grammarly to correct grammatical mistakes in the draft of the paper. The authors take full responsibility for the publication's content.

References

- [1] A. Potyagalova, G. J. F. Jones, MultiSumm – multimodal summarisation task at MediaEval 2025, in: Working Notes Proceedings of the MediaEval 2025 Workshop, CEUR Workshop Proceedings, Dublin, Ireland, 2025. URL: <https://2025.multimediaeval.com/paper53.pdf>.
- [2] S. M. Dhannya, S. Lakshmi Priya, M. Shalini, A. Vijayakumar, S. Ranjan, MultiSumm: An AI-powered system for automated multimodal news summarization, in: Working Notes Proceedings of the MediaEval 2025 Workshop, CEUR Workshop Proceedings, Dublin, Ireland, 2025. URL: <https://2025.multimediaeval.com/paper32.pdf>.
- [3] A. Potyagalova, G. J. F. Jones, LLM in multi-document summarisation: MultiSumm task at MediaEval 2025, in: Working Notes Proceedings of the MediaEval 2025 Workshop, CEUR Workshop Proceedings, Dublin, Ireland, 2025. URL: <https://2025.multimediaeval.com/paper51.pdf>.
- [4] H. Wu, H. Cho, A. R. Davies, G. J. F. Jones, LLM-based automated web retrieval and text classification of food sharing initiatives, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), ACM, 2024, pp. 4983–4990. doi:10.1145/3627673.3680090.
- [5] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitansky, R. O. Ness, J. Larson, From local to global: A graph RAG approach to query-focused summarization, arXiv preprint arXiv:2404.16130 (2024). URL: <https://arxiv.org/abs/2404.16130>.
- [6] G. Ye, nano-graphrag contributors, nano-graphrag: A simple, easy-to-hack GraphRAG implementation, <https://github.com/gusye1234/nano-graphrag>, 2024.
- [7] Dublin Postal Districts – GeoJSON Boundaries, <https://zenodo.org/records/4589220>, 2021. doi:10.5281/zenodo.4589220.
- [8] missinglink, UK Postcode Polygons: BN GeoJSON, <https://geoportal.statistics.gov.uk/>, 2016.
- [9] R. Radoi, London Borough Boundaries GeoJSON, <https://github.com/radoi90/housequest-data>, 2015.
- [10] M. González, BCN Geodata: Districtes de Barcelona, <https://github.com/martgnz/bcn-geodata>, 2017.
- [11] Comune di Milano, Municipi: Superficie e Confini – Open Data, <https://dati.comune.milano.it/dataset/ds379-infogeo-municipi-superficie>, 2023.