

LLM in Multi-Document Summarisation : MultiSumm Task at MediaEval 2025

Anastasiia Potyagalova, Gareth J. F. Jones

ADAPT Centre, School of Computing, Dublin City University, Ireland

Abstract

We describe our participation in the MediaEval 2025 MultiSumm task on multimodal summarization of multiple websites. Our system focuses on the automated summarization of Food Sharing Initiatives (FSIs) in urban environments. The approach combines web data crawling, classification, and summarization pipelines developed within the H2020 CULTIVATE project. We generate concise summaries by integrating multimodal (text + image) features and LLM-based evaluation metrics. The system achieves strong coverage and coherence on multiple test cities (Dublin, Brighton, London, Milan, Barcelona).

1. Introduction

Our approach for the MultiSumm 2025 task is based on an automated Google Search pipeline that detects, extracts, and analyzes relevant web pages describing Food Sharing Initiatives (FSIs). The system gathers publicly available online content — such as official websites, social media pages, and organization directories — and processes this material to generate multimodal summaries. Each retrieved page is cleaned, parsed, and analyzed to extract textual information and representative visual content (e.g., photographs, logos, and event images).

The extracted text and associated images are then used as input to the summarization module, which produces structured city-level reports describing the characteristics and activities of FSIs. These summaries aim to provide an integrated view of the local food sharing ecosystem, combining descriptive and visual evidence from multiple independent sources.

The primary objective of this approach is to automate the generation of coherent and verifiable summaries that reflect the diversity of food sharing practices across different cities. The pipeline ensures reproducibility and scalability, allowing large-scale analysis of community-driven initiatives by applying a unified search and summarization workflow.


2. Approach


Our system pipeline follows the task definition presented in the MultiSumm 2025 overview paper [?], which focuses on multimodal summarization of multiple websites representing Food Sharing Initiatives (FSIs). The proposed approach consists of five main stages: data acquisition, preprocessing, classification and labeling, multimodal summarization, and evaluation.


2.1. Data Acquisition

The input to the system comprises lists of manually verified URLs corresponding to FSIs in each target city. These lists were provided for the training city (Cork) and evaluation cities (Dublin,

MediaEval'25: Multimedia Evaluation Workshop, October 25–26, 2025, Dublin, Ireland and Online

 anastasia.potyagalova2@mail.dcu.ie (A. Potyagalova); Gareth.Jones@dcu.ie (G. J. F. Jones)

 © 2025 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)

Brighton, London, Milan, and Barcelona). The pipeline uses an automated Google Search component to retrieve additional web pages that are likely to describe related initiatives [1], [2]. For each link, the system downloads the web content, filtering out duplicates, advertisements, and irrelevant domains. Crawling is performed using a custom Python-based scraper with rate limiting and HTML structure validation.

2.2. Data Preprocessing

All collected web pages are cleaned and normalized before analysis. Text content is extracted using BeautifulSoup, tokenized, and stored in a structured JSON format together with metadata (URL, title, and text from "about-me" page). Visual materials such as embedded photographs, logos, or event banners are downloaded and linked to the corresponding text segments. Each data item therefore contains both textual and visual information for later multimodal summarization.

2.3. Multimodal Summarization

To guide the generation of the textual summaries, we prepared a structured set of potential questions for the language model [3], [4]. These questions were designed to capture the main dimensions of the FSI landscape within each city, including organizational type, funding model, target audiences, sustainability practices, and community engagement activities. For every city in the dataset, the LLM received this predefined question set and produced concise answers based on the information available in the previously extracted text segments [5], [6], [7]. If a particular piece of information was not present in the source material, the system omitted the corresponding answer to avoid speculative or unsupported content. This procedure ensured that the generated textual reports remained factual and directly grounded in the crawled web data. The visual component of the report was constructed in a parallel manner. For each FSI, the system automatically retrieved images from representative web pages—most notably the *About Us* sections, media galleries, or event announcements. These images were then analyzed by the LLM, which evaluated their semantic relevance and selected the most appropriate visual examples to accompany the textual summary. This multimodal selection process ensured that the final reports combined informative written descriptions with contextually meaningful imagery, thereby providing a more complete and credible representation of the food sharing landscape in each city.

2.4. Evaluation

Evaluation combines both automatic and manual components. LLM-as-judge assessment is employed to measure factual accuracy, coherence, and multimodal alignment between text and image content [8], [9], [10]. Human evaluators further review a subset of the generated summaries for completeness and clarity. Quantitative measures include coverage of available FSIs and ratio of successfully summarized pages per city. The system outputs are finally compiled into structured JSON reports.

Overall, the proposed pipeline provides an end-to-end workflow for large-scale, multimodal summarization of food sharing web data, aligning with the objectives of the MultiSumm 2025 challenge.

3. Results

Table 1 presents the performance of **Team 3 (ADAPT DCU)** for all cities included in the MultiSumm 2025 evaluation. The testing data includes Dublin and Brighton as the main-task test set and London, Milan, and Barcelona were used as the sub-task evaluation set.

Table 1
MultiSumm 2025 results for Team 3

City	Task Type	Coverage (%)	Coherence Score
Dublin	Main	96.4	0.84
Brighton	Main	94.7	0.81
London	Subtask	92.1	0.79
Milan	Subtask	95.0	0.83
Barcelona	Subtask	90.6	0.78
Average		94.8	0.82

The system successfully generated summaries for almost all available Food Sharing Initiative (FSI) URLs across the evaluation cities, achieving an overall coverage of approximately 95%. The summarization outputs demonstrated consistent coherence and factual completeness, particularly for cities with higher-quality web content (Dublin and Milan).

Qualitative inspection revealed that image–text integration improved interpretability in roughly two-thirds of the summaries. The most common errors were linked to duplicated or incomplete sentences when the source websites contained fragmented HTML structures. Nevertheless, the pipeline produced balanced and informative multimodal summaries suitable for downstream analysis of food-sharing ecosystems.

Future improvements will focus on refining the LLM prompting for more concise outputs, improving robustness to mixed-language content, and developing automatic metrics for measuring multimodal alignment.

4. Discussion and Insights

The evaluation results demonstrate that the proposed pipeline achieved strong overall performance across both the main and subtask cities.

A notable trend was the dependency between web content heterogeneity and summarization quality. Cities where FSIs maintain active and content-rich websites produced more coherent summaries, while limited-text or social-media-only pages occasionally resulted in less structured outputs. This highlights the importance of robust preprocessing and domain-specific filtering when dealing with heterogeneous web data.

The manual evaluation further indicated that multimodal enrichment, specifically the inclusion of images, logos, or event photographs, enhanced the quality of the generated reports. At the same time, the system occasionally generated redundant descriptions and provided duplicated or very similar images when duplicate pages were crawled from multiple sources. Future iterations of the pipeline will incorporate improved deduplication and clustering techniques, as well as multilingual adaptation for cities where initiatives operate in non-English contexts.

Overall, the presented approach demonstrates that large-scale, automated summarization of community-driven initiatives is feasible and can support the documentation and comparative analysis of food sharing ecosystems across European cities.

5. Conclusions and Future Directions

This paper presented the participation of Team 3 (ADAPT DCU) in the MediaEval 2025 MultiSumm task on multimodal summarization of multiple websites. Our approach, built around a Google Search-driven crawling and summarization pipeline, successfully automated the extraction, classification, and synthesis of textual and visual information about FSIs across several European cities.

By combining structured data preprocessing, LLM-based summarization, and image-text alignment, the method produced concise and verifiable reports suitable for both quantitative analysis and public communication. The outputs could be valuable for documenting local sustainability activities and facilitating cross-city comparisons within the CULTIVATE project.

Future work will focus on three main directions:

- We plan to integrate multilingual summarization capabilities to support FSIs operating in non-English environments.
- Aim to improve multimodal alignment by incorporating automatic visual-semantic consistency metrics.
- Explore adaptive prompting strategies to enhance factual precision and reduce redundancy in large-scale summarization outputs.

These steps are expected to increase the quality, interpretability, and scalability of automated FSI documentation for future editions of the MultiSumm challenge.

6. Acknowledgement

This research received support from the SFI ADAPT II, and European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101083377.

References

- [1] 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, p. 4983–4990. URL: <https://doi.org/10.1145/3627673.3680090>. doi:10.1145/3627673.3680090.
- [2] S. Lanka, A. Srivallop, C. A. Pinto, A specialized framework of web application for efficient data retrieval on social media tools, in: 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2021, pp. 161–166.
- [3] OpenAI, Gpt-4 technical report, 2024. [arXiv:2303.08774](https://arxiv.org/abs/2303.08774).
- [4] Anonymous, Llms-as-judges: A comprehensive survey on llm-based evaluation methods, arXiv preprint [arXiv:2412.05579](https://arxiv.org/abs/2412.05579) (2024). URL: <https://arxiv.org/html/2412.05579v2>.
- [5] Y. Chae, T. Davidson, Large language models for text classification: From zero-shot learning to fine-tuning, Open Science Foundation (2023).
- [6] Y. Zhang, M. Wang, C. Ren, Q. Li, P. Tiwari, B. Wang, J. Qin, Pushing the limit of llm capacity for text classification, arXiv preprint [arXiv:2402.07470](https://arxiv.org/abs/2402.07470) (2024).
- [7] X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, G. Wang, Text classification via large language models, arXiv preprint [arXiv:2305.08377](https://arxiv.org/abs/2305.08377) (2023).
- [8] Y. Lu, X. Yang, X. Li, et al., Llmscore: Unveiling the power of large language models in text-to-image synthesis evaluation, arXiv preprint [arXiv:2305.11116](https://arxiv.org/abs/2305.11116) (2023). URL: <https://arxiv.org/abs/2305.11116>.
- [9] Z. Sheng, K. Yang, et al., Multi-document summarization via deep learning techniques, ACM Transactions on Information Systems (2021). doi:10.1145/3529754.

- [10] H. Wei, S. He, T. Xia, F. Liu, A. Wong, J. Lin, M. Han, Systematic evaluation of llm-as-a-judge in llm alignment tasks: Explainable metrics and diverse prompt templates, 2025. URL: <https://arxiv.org/abs/2408.13006>. `arXiv:2408.13006`.