LSAT Focus: EAGLE's Embedded Entities Highlighting Technique for NTCIR-18 Lifelog-6

Introduction

This paper focusing on automatic searching methods for finding distinct life moments. Our experiments explore and compare different retrieval strategies, including keyword matching-based search combined with embedding extraction, vector embedding-based semantic search using a multimodal model, and hybrid methods that take advantage of both approaches. Our proposed method improved retrieval accuracy by directing the model's attention to key query terms while prioritizing semantic relevance and the presence of requested entities in the retrieved moments. Experimental results demonstrated that the best-performing method relies on embeddings incorporating extended descriptions and highlighted keywords. Conversely, the hybrid methods in our experiments have less effective results, likely due to limitations in the keyword-matching search algorithm. This work's findings underscore the richer descriptive entities within queries to enhance the retrieval of life moments, ensuring a focus on core semantic and visual elements.

Methodology

Keyword-based retrieval matches query terms with metadata, excelling at named entities and explicit concepts. However, it struggles with abstract or vague queries due to vocabulary limitations.

Semantic retrieval (e.g., using CLIP) embeds queries and images into a shared latent space, enabling broader concept matching beyond exact words. Yet, it may miss specific constraints like time or named entities.

Hybrid Approaches

To leverage the strengths of both methods, we explore two hybrid strategies:

Precision-Oriented (Consensus): Retrieve only items found by both keyword and semantic methods

Recall-Oriented (Filter-Rerank): Use keyword search to shortlist candidates, then rerank by semantic similarity

Enhancing Semantic Search with Entity Emphasis

We improve query embeddings by appending key entities to the query before encoding. This simple tweak makes the model focus more on critical concepts, improving alignment with user intent during vector search.

Although the semantic search pipeline remains unchanged, this simple modification encourages the embedding model to assign greater weight to salient terms. By explicitly repeating key concepts, the model is guided to attend more strongly to the most informative parts of the query, resulting in a vector that better reflects the user's search intent. After this enhancement step, the nearest-neighbour search proceeds as usual, using the modified query vector.


```
{
    '_index': 'ntcir-es',
    '_id': '20190109_150833_000',
    '_score': 20.574127,
    '_source': {
        'image_name': '20190109_150833_000.jpg',
        'timestamp': '20190109',
        'tags': 'text,indoor,wall,ceiling,screen',
        'categories': 'Office',
        'semantic_time:__new_timezone': 'Europe/Athens',
        'semantic_time:__weekday': 'Wednesday',
        'semantic_time:__year': '2019',
        'semantic_time:__month': 'January',
        'semantic_time:__semantic_time': 'early evening',
        'music:__album': None,
        'music:__artist': None,
        'music:__song': None,
        'time:__utc_time': '2019-01-09 15:08:01',
        'time:__local_time': '2019-01-09 17:08:01',
        'time:__minute_id': '20190109_1508',
        'activity:__movement': 'Inside',
        'activity:__movement_prob': 0.9933500886,
        'activity:__stop': True,
        'location:__lat': 40.6359851,
        'location:__lng': 22.9355687,
        'location:__semantic_name': 'Zeus Conference Rooms',
        'location:__original_name': 'Zeus Conference Rooms',
        'location:__parent': 'Mediterranean Palace Hotel',
        'location:__city': 'Thessaloniki Regional Unit, Thessaloniki Municipal Unit, Macedonia and Thrace, Greece',
        'location:__country': 'Greece',
        'visual_concepts:__OCR': None,
        'visual_concepts:__Caption': 'a few people in a conference room',
        'visual_concepts:__CaptionScore': 0.4219984114
    }
}
```


Experiments and Results

Incorporating both descriptions and narrative fields further improves embedding performance by enriching contextual information, while sometimes introducing noise in keyword-based methods. Enhancing query representations through entity highlighting consistently boosts retrieval effectiveness, with the best configuration achieving mAP 0.2075 and Recall@100 of 0.5002.

	mAP	P@5	P@10	P@100	R@5	R@10	R@100
Text Matching	0.0121	0.0462	0.05	0.0435	0.0052	0.0058	0.0274
Text Matching Narrative	0.0117	0.0462	0.0538	0.0412	0.0052	0.0067	0.0263
Embedding	0.1608	0.2538	0.2385	0.1323	0.1522	0.2167	0.4779
Embedding Narrative	0.1964	**0.3077**	**0.2692**	**0.155**	0.1566	0.2018	0.4806
Filter and Rerank	0.0758	0.1538	0.1308	0.0877	0.0692	0.0916	0.1769
Filter and Rerank Narrative	0.0988	0.1692	0.1385	0.0938	0.0759	0.09	0.2204
Combined Intersection	0.0507	0.152	0.116	0.0548	0.0752	0.0858	0.1139
Combined Intersection Narrative	0.0975	**0.1913**	**0.1565**	**0.0626**	0.0913	0.0963	0.2129
Embedding-based with Entity Highlight	0.2031	0.2462	0.2462	0.1496	0.1309	0.2136	0.491
Embedding-based with Entity Highlight Narrative	**0.2075**	0.2538	0.2423	0.135	**0.1686**	**0.2337**	**0.5002**

In contrast, hybrid methods that combine keyword and vector approaches underperform due to the weak keyword component filtering out relevant results, highlighting the importance of strong individual components in hybrid strategies.