

Knowledge Graph Extension based on Generative Retrieval

1. Introduction

Knowledge Graph Extension: the basic idea is to extend a given *knowledge graph* (KG) using a *Large Language Model* (LLM).¹ The LLM is used to add information to the existing graph exploiting the *implicit knowledge* baked into the LLM.

Generative Retrieval: this method uses a LLM to perform a retrieval-like task. The LLM scans a large body of text (e.g. a book, a transcription, a database, etc.) and generates condensed, abstractive representations of the information most relevant to a query. This approach can be seen as a hybrid between information retrieval and abstractive summarization.

Knowledge Graph Extension based on Generative Retrieval: Instead of relying on the implicit knowledge baked into the LLM, through GR a large body of text data can be used as a knowledge base for the KG extension. This method has a few different advantages: it reduces the possibility of hallucination, it allows to deal with up-to-date information, it allows to deal with specific or niche bodies of knowledge and, most importantly, provides direct grounding to knowledge.

Procedure: provide the LLM with data and prompt it to extract the relevant information and give a response in natural language. The extracted information is then used as the content of the knowledge graph extension, that will be performed by another instance of the LLM.

Notice that I use two distinct instances of the LLM instead of one. This has a few advantages:

- Lighter use of the context window in each instance, permitting to use more data and longer prompts in both instances.
- More control over the prompts, leading to better testability and more control over the results.
- The possibility to use two different specialized models, since the two tasks are quite different.

2. Use case example

Extending a KG about a movie with information provided by the viewers in social media.

As a use case, I show the extension of a KG that represents basic information about a movie. The information for the extension is extracted from a collection of social media posts and comments through GR. For instance, since the comments are written by users who watched the movie, we can explore the following themes: immediate reactions from the viewers, perceived meaning of the movie and perceived cultural or political references. An instance of the LLM is prompted to produce a short phrase in natural language that describes one of these aspects. Then, another instance of the LLM is given the Base KG and the short phrase and is prompted to extend the Base KG with the content of the short phrase.

¹ De Giorgis, S., Gangemi, A., Russo, A. Neurosymbolic graph enrichment for Grounded World Models. Information Processing & Management, 62, 4, 2025

Procedure:

Input data:

- **Base KG:** a simple KG containing basic information about a movie. The graph represents an entry on IMDB.
- **Text Data:** a large amount of text from social media posts and comments regarding the movie. In this case, the data will be extracted from Reddit through its API.

Processing:

- A LLM is prompted to extract information from the Text Data regarding a specific topic and generate a response in natural language.
- A new instance of a LLM is prompted to produce an extended KG, enriching the Base KG with the information provided by the response at the previous step.

Result:

- **Extended KG:** a KG extended with grounded information.

In this example, GR is useful for the following reasons:

- The possibility to extract knowledge about specific movies that are not known by the LLM (especially movies that came out after its knowledge cut-off).
- The possibility to represent knowledge that is grounded in the viewers' experience.
- The fact that we can track back knowledge to its original source.

3. Application

Base KG:²

The base graph represents the IMDB entry for the movie *Lilo & Stitch*.

```
@prefix imdb: <http://example.org/imdb/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

imdb:tt11655566 a dbo:Film ;
  dct:title "Lilo & Stitch" ;
  dbo:releaseYear "2025"^^xsd:gYear ;
  dbo:genre imdb:Genre_Action ,
            imdb:Genre_Drama ,
            imdb:Genre_Comedy ;
  owl:sameAs <https://www.imdb.com/title/tt11655566> .

imdb:Genre_Action a dbo:Genre ;
  rdfs:label "Action" ;
  owl:sameAs <http://dbpedia.org/resource/Action_film> .

imdb:Genre_Drama a dbo:Genre ;
  rdfs:label "Drama" ;
  owl:sameAs <http://dbpedia.org/resource/Drama_film> .
```

² This graph was created with the help of a LLM, starting from an entry of the IMDB. This is not essential for the aim of this work. On the contrary, this graph may contain errors and it would be better to use a well curated graph.

```
imdb:Genre_Comedy a dbo:Genre ;
  rdfs:label "Comedy" ;
  owl:sameAs <http://dbpedia.org/resource/Comedy_film> .
```

Text Data:

The data consists of social media posts and comments about the movie *Lilo & Stitch* from the following Reddit posts:

```
I watched Lilo & Stitch today... Here are my thoughts
https://www.reddit.com/r/moviecritic/comments/lkus4lf/i_watched_lilo_stitch_today_here_are_my_thoughts/

"Lilo and Stitch" sets a record in its debut and becomes the highest-grossing film in the United States
https://www.reddit.com/r/moviecritic/comments/lkzml2/lilo_and_stitch_sets_a_record_in_its_debut_and/

What are your thoughts on Lilo & Stitch live action remake?
https://www.reddit.com/r/moviecritic/comments/lj9usjo/what_are_your_thoughts_on_lilo_stitch_live_action/

Lilo & Stitch (casting)
https://www.reddit.com/r/moviecritic/comments/lkvz3l/lilo_stitch_casting/
```

GR:

- **Model:** OpenAI GPT 5 Mini, accessed through API
- **Prompt:**

```
input=f"You are an assistant specialized in the analysis of comments about a movie from Reddit users. Read all the comments and provide a very brief response, in which you describe the immediate reaction of the viewers. Use a simple syntax, later your response will be transformed in a knowledge graph by another LLM. Limit the response to a single phrase starting like 'the viewers reacted...'. These are the comments from the viewers: ###{comments}###",
reasoning={ "effort": "low" },
text={ "verbosity": "low" }
```

- **Response:**

the viewers reacted with polarized nostalgia—many families and kids loved it and praised Stitch and performances, while longtime fans angrily criticized character/ending changes (notably Nani's custody, Jumba/Gantu and casting), sparking cultural debate despite strong box-office success.

KG Extension:

- **Model:** OpenAI o4 Mini, accessed through API
- **Prompt:**³

```
{"role": "system", "content": f"Extend a knowledge graph (KG) with the information from the text provided by the user, using a frame semantics approach. The user will provide the knowledge graph to be extended and the text containing the information for the extension. You must use the following resources: OWL, dbpedia, Wordnet, framester, friend of a friend and DOLCE. You will produce the graph only, without any comment or explanation. The extended graph must include all the content of the starting graph. Don't include parts of the text as a string description. Remember to always specify prefixes in front of every element. The KG should be expressed according to: - the semantics of OWL2 logic language - frame semantics principles as formalised in the Framester schema (https://w3id.org/framester/schema/) - RDF language serialized in Turtle syntax. The alignments in KG are to PropBank, WordNet, VerbAtlas, Wikidata, DBpedia, DOLCE-Zero, etc."},
{"role": "user", "content": f"This is the graph: ---{graph}--- This is the text: ---{text}---"}
```

- **Response:**

```
@prefix imdb: <http://example.org/imdb/> .
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix pb: <http://propbank.org/pb/> .
@prefix wn: <http://wordnet-rdf.princeton.edu/wn30/> .
@prefix vap: <https://w3id.org/framester/verbAtlas/> .
@prefix fr: <https://w3id.org/framester/schema/> .
```

³ Part of the prompt is derived from De Giorgis, S., Gangemi, A., Russo, A. Neurosymbolic graph enrichment for Grounded World Models. Information Processing & Management, 62, 4, 2025

```

@prefix fne: <https://w3id.org/framester/fn/> .
@prefix dbp: <http://dbpedia.org/resource/> .
@prefix dzero: <http://www.ontologydesignpatterns.org/ont/dolce/zero/0.1#> .

imdb:tt11655566 a dbo:Film ;
    dct:title "Lilo & Stitch" ;
    dbo:releaseYear "2025"^^xsd:gYear ;
    dbo:genre imdb:Genre_Action ,
              imdb:Genre_Drama ,
              imdb:Genre_Comedy ;
    owl:sameAs <https://www.imdb.com/title/tt11655566> .

imdb:Genre_Action a dbo:Genre ;
    rdfs:label "Action" ;
    owl:sameAs <http://dbpedia.org/resource/Action_film> .

imdb:Genre_Drama a dbo:Genre ;
    rdfs:label "Drama" ;
    owl:sameAs <http://dbpedia.org/resource/Drama_film> .

imdb:Genre_Comedy a dbo:Genre ;
    rdfs:label "Comedy" ;
    owl:sameAs <http://dbpedia.org/resource/Comedy_film> .

imdb:ViewerGroup a foaf:Group, dzero:SocialAgent ;
    rdfs:label "Viewers" .

imdb:Group_FamiliesKids a foaf:Group, dzero:SocialAgent ;
    rdfs:label "Families and kids" .

imdb:Group_LongtimeFans a foaf:Group, dzero:SocialAgent ;
    rdfs:label "Longtime fans" .

imdb:Stitch a dbo:Character ;
    rdfs:label "Stitch" ;
    owl:sameAs dbp:Stitch_(Disney) .

imdb:Performance a dbo:Performance ;
    rdfs:label "Performances" ;
    owl:sameAs dbp:Performance_(arts) .

# Events and changes
:reaction1 a fr:ReactionEvent ;
    fr:evokedBy pb:react-01, wn:react%2:38:00 ;
    fr:Experiencer imdb:ViewerGroup ;
    fr:Stimulus imdb:tt11655566 ;
    fr:CauseEmotion dbp:Nostalgia ;
    rdfs:comment "polarized nostalgia" .

:lovel a fr:PositiveEvaluationEvent ;
    fr:evokedBy pb:love-01, wn:love%2:37:01, vap:Love ;
    fr:Experiencer imdb:Group_FamiliesKids ;
    fr:Stimulus imdb:tt11655566 ;
    rdfs:label "loved it" .

:praisel a fr:PraiseEvent ;
    fr:evokedBy pb:praise-01, wn:praise%2:31:00 ;
    fr:Experiencer imdb:Group_FamiliesKids ;
    fr:Stimulus imdb:Stitch, imdb:Performance ;
    rdfs:label "praised Stitch and performances" .

:critiquel a fr:CriticismEvent ;
    fr:evokedBy pb:criticize-01, wn:criticize%2:41:00 ;
    fr:Experiencer imdb:Group_LongtimeFans ;
    fr:Stimulus :characterEndingChanges ;
    fr:Manner dbp:Anger ;
    rdfs:label "angrily criticized changes" .

:characterEndingChanges a fr:Change ;
    rdfs:label "character/ending changes" ;
    fr:hasPart :naniCustodyChange, :jumbaGantuCastingChange ;
    owl:sameAs dbp:Change_(word) .

:naniCustodyChange a fr:Change ;
    rdfs:label "Nani's custody change" .

:jumbaGantuCastingChange a fr:Change ;
    rdfs:label "Jumba/Gantu and casting change" .

```

```
:debatel a fr:DebateEvent ;
    fr:evokedBy pb:debate-01, wn:debate%2:32:00 ;
    fr:Theme dbp:Cultural_debate ;
    fr:Cause :critique1 ;
    rdfs:label "sparked cultural debate" .

:boxofficel a fr:SuccessEvent ;
    fr:Experiencer imdb:tt11655566 ;
    fr:Attribute dbp:Box_office ;
    fr:Polarity "positive" ;
    rdfs:label "strong box-office success" .

# Frame alignments
fr:ReactionEvent          owl:equivalentClass dbp:Reaction .
fr:PositiveEvaluationEvent owl:equivalentClass dbp:Positive_evaluation .
fr:PraiseEvent            owl:equivalentClass dbp:Praise .
fr:CriticismEvent         owl:equivalentClass dbp:Criticism .
fr:DebateEvent            owl:equivalentClass dbp:Debate .
fr:SuccessEvent           owl:equivalentClass dbp:Success .
```