Sophocles' Antigone as a Knowledge Graph through a Hybrid Collaborative Workflow with Ontology-Guided LLM Extraction

Apostolos Baniotis
Department of Computer Science, University of Crete & FORTH-ICS
Heraklion, Greece
baniotis@ics.forth.gr

Marsel Senka
Department of Computer Science, University of Crete
Heraklion, Greece
senkamarselo@gmail.com

Entisa Tzeortziana Komoritsan
Department of Computer Science, University of Crete
Heraklion, Greece
csdp1463@csd.uoc.gr

Alexandros Manos
Department of Computer Science, University of Crete
Heraklion, Greece
alexandros.manos00@gmail.com

Christina Kanetou
Department of Computer Science, University of Crete & FORTH-ICS
Heraklion, Greece
csdp1465@csd.uoc.gr

Konstantinos Drenoviadis
Department of Computer Science, University of Crete & FORTH-ICS
Heraklion, Greece
kdren@csd.uoc.gr

Evangelia Milolidaki
Department of Computer Science, University of Crete & FORTH-ICS
Heraklion, Greece
csdp1451@csd.uoc.gr

Yannis Tzitzikas
Department of Computer Science, University of Crete & FORTH-ICS
Heraklion, Greece
tzitzik@ics.forth.gr

Abstract

Building a knowledge graph (KG) from a Greek tragedy poses challenges absent from tabular or encyclopedic sources: the play has a fixed dramatic hierarchy (scene, speech, line), scholarly readings (theme, emotion, motivation, myth) must be modeled explicitly, and multiple language editions must share a single identity graph rather than fork the structure per language. We target a KG for Sophocles' *Antigone* that supports two coupled uses: structured retrieval, through integrity and competency questions expressed in SPARQL over dramatic structure and interpretive annotations; and interactive exploration, through a lightweight read client that navigates lines across languages, shows scene context, and reports corpus statistics. To build it, we introduce a hybrid construction workflow in which ontology-constrained Large Language Models (LLMs) propose RDF triples and human reviewers validate them at producer and aggregator stages. A custom ontology, aligned with the well-established FRBRoo and CIDOC-CRM standards, fixes the only permitted vocabulary. Each Ancient Greek line acts as the single reference line for its verse, and translations, here Modern Greek, English, and Mandarin, attach to it as `TranslationVariant` nodes, keeping identity in one place and cross-lingual queries simple. Construction proceeds in three LLM passes (structure, semantics, translations), governed by machine-readable rules, dramatic-boundary chunking, and peer review. The published corpus comprises 185 modular Turtle files, 1259 reference lines, 291 speeches, 3500+ translation variants, and 25,481 triples. We report the methodology, validation outcomes, documented merge challenges, and lessons for humanities knowledge graph construction at verse scale.

CCS Concepts

• **Computing methodologies → Ontology engineering**; **Knowledge representation and reasoning**; • **Information systems →** *Information extraction.*

Keywords

Knowledge Graph Construction, Ontology-based Extraction, Large Language Models, Literary Text, Collaborative Workflow, RDF, SPARQL

ACM Reference Format:
Apostolos Baniotis, Marsel Senka, Entisa Tzeortziana Komoritsan, Alexandros Manos, Christina Kanetou, Konstantinos Drenoviadis, Evangelia Milolidaki, and Yannis Tzitzikas. 2026. Sophocles' Antigone as a Knowledge Graph through a Hybrid Collaborative Workflow with Ontology-Guided LLM Extraction. In *Proceedings of The KGCSW-SETN 2026 Workshop on Knowledge Graph Construction (KGCSW-SETN 2026).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Motivation. Composed almost two and a half millennia ago, Sophocles' *Antigone* remains one of the most influential texts in human

KGCSW-SETN 2026, Chania, Greece

ACM ISBN 978-1-4503-XXXX-X/26/06
https://doi.org/XXXXXXX.XXXXXXX

literary history and is still staged every year across many countries, continuing to frame debates on divine versus human law, conscience, and civic obedience. A single performance text combines a dramatic spine (play → scene → speech → line), performative roles (characters, chorus), and interpretive layers (theme, emotion, motivation, myth, ethical conflict) that classicists attach to utterances, not to isolated words, and thus resist shallow text mining. Yet no existing knowledge graph unites these aspects for a complete tragedy at verse scale: prior resources emphasize text encoding (TEI; [1]), corpus access without play-specific interpretive graphs (Perseus; [2]), or bottom-up expert annotation later reorganized into an ontology (Euporia; [16]); knowledge graphs of fragmentary tragedians [4] do not combine a full play with multilingual line alignment and multi-faceted interpretive predicates under one ontology.

Objective. We set out to produce a knowledge graph (KG) of *Antigone* that serves two coupled purposes. First, structured querying: every edition line and speech should be typed, linkable, and retrievable with SPARQL, e.g., who speaks at line *n*, which speeches develop a theme, how Ancient Greek and translation variants align. Second, interactive exploration: the same graph should power a browser client where readers navigate scenes and lines, inspect speaker and semantic context, and view corpus-level statistics, enabling non-linear reading of the play with value for both analysis and education. Achieving both requires explicit modeling of dramatic structure and scholarly readings, plus a multilingual alignment pattern in which Ancient Greek, Modern Greek, English, and Mandarin share one identity graph.

Novelty. To the best of our knowledge, this is the first published knowledge graph of a Greek tragedy that unifies (i) dramatic structure, (ii) interpretive semantics, and (iii) multilingual textual alignment within a single queryable representation. This enables answering complex questions like (i) *"Retrieve the English and Mandarin translation variants of all individual verses spoken by Creon in scenes where he experiences the emotion of anger while asserting civic law, alongside any associated mythological references used to justify his decisions"*, or (ii) *"Identify and compare the English and Mandarin translations of all verses where Antigone references the Oedipus myth while experiencing grief, but only within speeches that develop the theme of divine law versus civic obedience".* As regards the KG construction its novelty lies in three contributions: (a) an ontology-guided, multi-pass prompting method that treats the ontology not merely as a target schema but as an active constraint throughout LLM-based extraction, limiting vocabulary drift, unsupported relations, and structural inconsistencies, (b) a shared-reference modeling pattern that maintains a unified structural and semantic hierarchy for the work, aligning multiple translations to the same verse entities to support cross-lingual querying without reproducing the graph for each language, and (c) a human-in-the-loop construction workflow in which expert knowledge guides LLM generation through the ontology and demonstration examples, while human annotation, refinement, and integration progressively transform independently generated outputs into a coherent and reliable knowledge graph.

Results and evaluation. The workflow, illustrated in Figure 1,

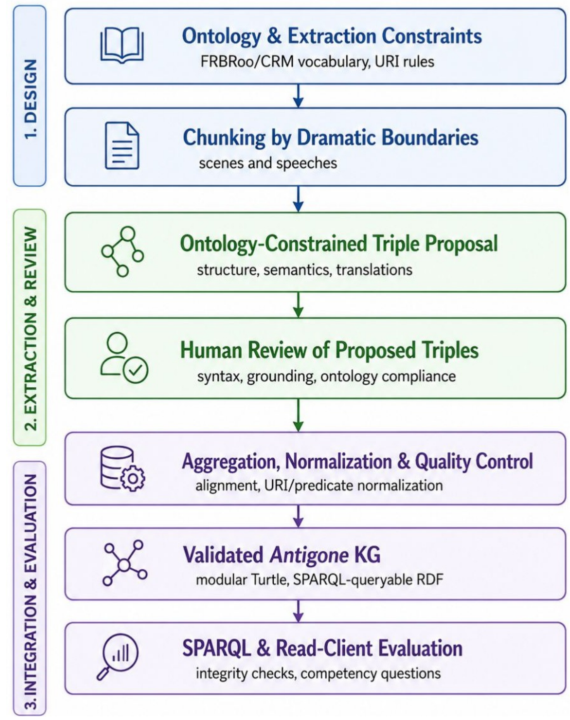


Figure 1: Collaborative KG construction workflow for *Antigone*.

yielded a substantial published corpus of 25,481 triples, including all 1259 available canonical reference lines populated across the play's 1353 edition line span, produced over approximately six months of collaborative effort. Quality was assessed through peer review at extraction and merge, integrity SPARQL after each load, competency-question patterns, and deployment of an interactive read client that stress-tests the endpoint. A live demonstration of the knowledge graph and read client is publicly available[1]. The complete dataset, ontology, extraction scripts, and read client source code are open-source and available on GitHub[2]. Ultimately, while the LLM proposal accelerated throughput, it required human correction for interpretive literary data, especially speech boundaries and theme labeling, confirming the necessity of a hybrid methodology. The rest of this paper is organized as follows: Section 2 situates the work among literary encoding, tragedy annotation, and LLM-to-KG methods. Section 3 describes ontology design, extraction, aggregation, and publication. Section 4 reports validation and corpus statistics. Section 5 describes the interactive access services. Section 6 concludes the paper.

2 Background and Related Work

Semantic technologies and knowledge graphs increasingly represent historical and cultural data [16, 19]. TEI [22] remains the dominant encoding for classical texts [1]; Perseus [2, 6] provides

[1] https://demos.isl.ics.forth.gr/antigone/
[2] https://github.com/marselosenka/Antigone

access but not a play-specific interpretive graph . Antonopoulos et al. [4] structure fragmentary tragedians.

Mugelli et al. [16] annotate ritual and religious phenomena in Greek tragedy through the Euporia system: a bottom-up domain-specific language lets experts tag passages with interpretive hashtags, register competing scholarly readings of the same lines, and later reorganize tags in an OWL ontology queried with SPARQL. Their workflow, annotate, then search, then structure, inspired our emphasis on human judgment over interpretive content. We differ in scope and method: we target one play at speech and line granularity, with a fixed ontology declared before extraction (to constrain LLMs), four aligned translations on a Greek reference line, and dramatic structure (FRBRoo) alongside ethics and affect (CRM), not only ritual tags. *Antigone* is a natural test case for that tradition: the plot turns on burial rites, civic edict, pollution (*miasma*), and supplication, themes our corpus encodes as `Theme`, `Motivation`, and `Law` individuals with `description` literals.

LLM-Assisted Knowledge Graph Construction. In recent years, the paradigm of Knowledge Graph construction has shifted from purely rule-based or supervised NLP pipelines to approaches leveraging machine learning [24] and Large Language Models (LLMs) [5, 23]. LLMs possess remarkable zero-shot and few-shot capabilities for named entity recognition and relation extraction. However, applying them directly to historical or literary texts introduces severe challenges. Unconstrained LLMs suffer from structural hallucination and inventing non-existent ontology properties, nesting blank nodes uncontrollably, or failing to maintain consistent URIs across different prompts. Furthermore, when tasked with deep semantic interpretation (e.g., identifying the "motivation" of a character in a complex verse), LLMs often default to anachronistic or superficial summaries rather than scholarly readings. Recent methodologies advocate for "hybrid" or "human-in-the-loop" workflows [21], where the LLM acts as an initial high-recall triple proposer bounded by strict prompts, while human experts perform the final validation. Our work adapts this hybrid paradigm specifically for the granular, highly structured, and multilingual domain of classical tragedy.

3 Antigone KG Construction Process

At first we describe the methodology (in §3.1), then the general conceptually modelling approach (in §3.2) and then we describe how we tackled multilingualism (in §3.3). Subsequently we describe how we used the ontology in the LLM extraction protocol (in §3.4), and discuss the obverved behaviour of LLMs (in §3.5), and finally we describe the aggregation process (in §3.6).

3.1 Methodology

As stated earlier, we followed the workflow illustrated in Figure 1. A distinctive characteristic of our methodology is that we followed a *collaborative* approach in which student teams (in total 14 students), assumed defined *roles* under faculty supervision over *approximately six months*. The distinct roles are listed in Table 1. In brief, we had 3 main roles: KG producers (R_{KG}), KG aggregators (R_{AGG}) and GUI creators (R_{GUI}). For getting a clear picture of the effort spent on each step, we followed a process similar to [21], i.e. each contributor logged hours spent per task. Table 2 summarizes reported person-hours by team.

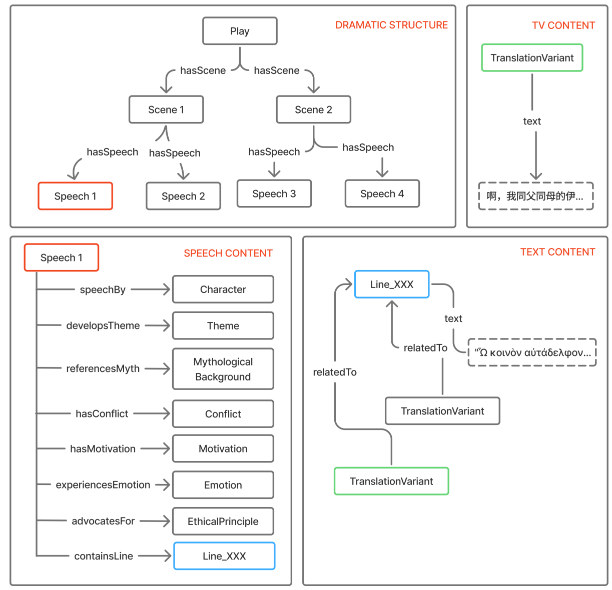


Figure 2: Conceptual Modeling Overview

Table 1: Team roles in the construction workflow.

Team	**Role**
R_{KG}	Produce RDF/Turtle from verse passages via ontology-guided LLM extraction; peer-review each chunk
R_{AGG}	Issue demo rules; plan chunks; merge heterogeneous TTL; cross-lingual alignment; publish to Fuseki
R_{GUI}	Deploy SPARQL access services; exercise the graph through layered read queries
Faculty	Ontology engineering, extraction contract, project management

Table 2: Reported construction effort by team (six-month project).

Team	**Students**	**Hours**	**Primary activities**
R_{KG} (RDF producers)	7	125	ontology-guided LLM extraction, peer review
R_{AGG} (aggregators)	3	85	Chunk planning, merge, alignment, Fuseki publication
R_{GUI} (access services)	4	140	SPARQL client, endpoint testing, UI integration
Total	**14**	**350**	–

3.2 Conceptual Modeling

An overview of the conceptual modeling approach is shown in Figure 2. To ensure interoperability with existing cultural heritage data, the Antigone ontology was explicitly aligned with the CIDOC

CRM (ISO 21127) [8][3] and its extension for bibliographic and performative works, FRBRoo[4].

FRBRoo is particularly suited for our domain because it elegantly models the distinction between a conceptual work (the Play), its performative expressions (the Scenes and Speeches), and its material manifestations (the translated texts). For instance, an utterance by Antigone is modeled not merely as a string of text, but as a discrete dramatic event (Speech) that occurs within the temporal bounds of a Scene.

CIDOC CRM is utilized to model the actors (`E21_Person` mapped to `Character`) and the conceptual interpretive layers (`E28_Conceptual_Object` mapped to `Theme`, `Emotion`, and `MythologicalBackground`). By grounding our custom classes in these established standards, the resulting Knowledge Graph remains highly expressive for classicists while remaining technically interoperable with broader Linked Open Data cultural initiatives.

Table 3: Ontology layers instantiated in the corpus.

Layer	Representative classes	Alignment
Dramatic structure	`Play`, `Scene`, `Speech`, `Line`	FRBRoo F23
Persons	`Character`, `Role`	CRM E21
Interpretive	`Theme`, `Emotion`, `Motivation`, `Conflict`	CRM E28
Intertextual	`MythologicalBackground`	CRM E28
Linguistic	`TranslationVariant`	—

Construction-critical properties include `hasSpeech`, `containsLine`, `speechBy`, `developsTheme`, `experiencesEmotion`, `hasMotivation`, `hasConflict`, `referencesMyth`, and `relatedTo` (for `TranslationVariant` → `Line` linking). `Theme_*` and `Emotion_*` are reified individuals with `description` literals. Most interpretive triples attach to `Speech`; lines inherit context through `containsLine`. An example is shown next.

Example (interpretive layer).

```
:Speech_Antigone_1_10
a :Speech ;
:speechBy :Antigone_Character_Antigone ;
:containsLine :Line_001, :Line_002, :Line_003, :Line_004,
              :Line_005, :Line_006, :Line_007, :Line_008,
              :Line_009, :Line_010 ;
:referencesMyth :Myth_Zeus, :Myth_OedipusTroubles ;
:developsTheme :Theme_FateOfOedipusLine, :Theme_FamilySuffering,
                :Theme_PoliticalDecreeThreat, :Theme_EnemiesHarm .
:Theme_FateOfOedipusLine
a :Theme ;
:description "The continuing evils inherited from Oedipus, presented
↪  as still being fulfilled upon his children while they
↪  live."^^xsd:string .
```

Each `description` records the *grounded reading* adopted for that chunk, analogous in spirit to registering one interpretive variant in expert-driven tragedy annotation [16], but typed under a shared ontology rather than free hashtags. These triples will be leveraged by the interactive access services that will be described in Section 5.

[3]https://cidoc-crm.org/
[4]https://cidoc-crm.org/frbroo

```
# 01_prologue/translations/english/1_to_19.ttl
:TV_Line_001_en a :TranslationVariant ;
:text "Ismene, sweet Ismene, my dear sister!"^^xsd:string ;
:relatedTo :Line_001 .

# 01_prologue/translations/greek/1_to_19.ttl
:TV_Line_001_gr a :TranslationVariant ;
   :text "Ω Ισμήνη, αδελφή μου από τον ίδιο πατέρα και την ίδια
μητέρα,"^^xsd:string ;
   :relatedTo :Line_001 .

# 01_prologue/translations/chinese/1_to_19.ttl
:TV_Line_001_zh a :TranslationVariant ;
   :text "啊，我同父同母的伊斯墨涅。"^^xsd:string ;
   :relatedTo :Line_001 .
```

Figure 3: Multilingual alignment: `TranslationVariant` entities anchoring to a shared canonical verse.

3.3 Reference Line Multilingual Linking

Each edition line maps to exactly one `Line_NNN` URI carrying the canonical Ancient Greek `text`. As shown in Table 4, *translations* use the prefix *TV_* (*T*ranslation *V*ariant): `TV_Line_NNN_{GR|EN|ZH}` individuals link to the reference line via the `relatedTo` property. Crucially, `Scene`, `Speech`, and `Character` URIs are shared across all languages and are never duplicated.

Table 4: URI naming pattern for verse alignments (line 1).

URI	Meaning
`:Line_001`	Canonical line 1 (Ancient Greek reference)
`:TV_Line_001_GR`	Modern Greek translation variant
`:TV_Line_001_EN`	English translation variant
`:TV_Line_001_ZH`	Mandarin translation variant

To avoid structural duplication, global extraction rules strictly restrict translation files to this pattern (as shown in Figure 3). Interpretive semantics are deliberately excluded from these language files and remain attached to the `Speech` level to ensure a clean separation of concerns.

Why this is novel. This is novel for two reasons: (a) query decoupling, and (b) property isolation.

Query Decoupling: Classicists can run complex, interpretive SPARQL queries (e.g., tracking the emotional trajectory of Creon) over the language-invariant core (:Speech and :Character), and then join the linguistic layer (TranslationVariant) at query runtime. This prevents O(N) graph duplication for N translations.

Property Isolation: Semantic annotations (themes, emotions, mythological references) attach only to the canonical Speech or Line level, preserving the purity of the translations which only hold raw string literals.

3.4 Ontology as Extraction Contract and LLM Extraction Protocol

A fundamental premise of our methodology is that the ontology must be engineered *a priori* and passed to the LLM as a strict, non-negotiable contract, rather than allowing the model to induce a schema bottom-up. In particular, the R_{KG} team transformed verse passages into ontology-compliant Turtle. The guiding principle: output must behave like a *controlled semantic representation*, ontology-locked vocabularies, no undefined properties, every resource a typed individual.

Specifically we followed a *Decoupled Sequential Extraction (DSE)* protocol that enforces a strict separation of concerns across the LLM generation space. Rather than executing a monolithic joint-extraction prompt, the pipeline is divided into three consecutive phases where the output of phase N serves as the deterministic structural anchor for phase N+1. This multi-pass decoupling mathematically bounds the branching factor of the generated graph in each turn, resulting in near-zero syntactic syntax errors (Turtle malformation) and significantly reducing semantic hallucinations. Note that in LLM operations, asking a model to extract structure, interpretive semantics, and translations simultaneously induces multi-task interference and causes the context window to degrade.

In particular, the extraction has three passes: Each R_{KG} contributor received a dramatic line range (e.g., prologue lines 1–160) and produced up to three Turtle files:

(1) *Pass A (structure):* `canonical/structure.ttl` — `Scene`, `Speech`, `Line_NNN` with Greek `text`.
(2) *Pass B (semantics):* `canonical/semantics.ttl` — `Theme`, `Emotion`, `Motivation`, etc.
(3) *Pass C (translations):* `translations/{greek|english|chinese}/X_to_Y.ttl` — `TV_Line_*` only.

3.4.1 Schema-Constrained In-Context ABox Generation.

Most LLM-based KGC approaches fall into two categories: (a) OpenIE, where the LLM extracts entities and relations without being constrained by a predefined ontology, often resulting in property drift and inconsistent relation representations [25]; or (b) Post-hoc Mapping, where the LLM outputs free-form triples that are later mapped to an ontology [13]. We have followed a different approach, that could be named *Schema-Constrained In-Context ABox Generation.* Instead of treating prompt composition in an ad hoc manner, we model it as a formal mapping that restricts the search space of the Large Language Model (LLM) to a closed-world representation defined by the ontology.

Let Σ^* denote the set of Unicode strings. Let $O = (\mathcal{T}, \mathcal{A}_{\text{init}})$ be the *Antigone* ontology, where $\mathcal{T}$ represents the terminological box (TBox) defining classes C and properties $\mathcal{P}$, and $\mathcal{A}_{\text{init}}$ represents the initial assertion box (ABox) containing predefined entities (e.g., characters, scenes). Let $o \in \Sigma^*$ be the serialization of O (specifically, `1.Antigone-Ontology.ttl`), $r \in \Sigma^*$ denote the global syntax rules (e.g., forbidden blank nodes, unique URI conventions), $t \in \Sigma^*$ be a pass-specific template (such as TPL_{sem}), and $x \in \Sigma^*$ be the assigned input verse range.

We define the prompt constructor Π as a mapping:

$$\Pi : \text{Pass} \times \Sigma^* \times \Sigma^* \times \Sigma^* \times \Sigma^* \to \Sigma^* \tag{1}$$

which expands to the XML-tagged composition:

$$\begin{aligned}\Pi(\text{pass}, o, r, t, x) = {} & \langle\text{ontology}\rangle \parallel o \parallel \langle/\text{ontology}\rangle \\ & \parallel \langle\text{constraints}\rangle \parallel r \parallel \langle/\text{constraints}\rangle \\ & \parallel \langle\text{task\_desc}\rangle \parallel t \parallel \langle/\text{task\_desc}\rangle \\ & \parallel \langle\text{source}\rangle\langle\text{TEXT}\rangle \parallel x \parallel \langle/\text{TEXT}\rangle\langle/\text{source}\rangle \\ & \parallel \text{“Output only valid RDF/Turtle triples.”}\end{aligned} \tag{2}$$

where $\parallel$ denotes string concatenation.

The LLM acts as an extraction function Φ that projects this prompt into a set of proposed RDF assertions $\mathcal{A}_{\text{proposed}}$ that must parse as syntactically valid Turtle:

$$\Phi : \Sigma^* \to \mathcal{A}_{\text{proposed}} \tag{3}$$

To ensure the semantic integrity and strict ontological alignment of the generated graph, the output of Φ is subject to a schema-contract enforced by the TBox $\mathcal{T}$. For every proposed triple $(s, p, o) \in \mathcal{A}_{\text{proposed}}$, the following constraints must hold: $p \in \mathcal{P}$ and
$((p, \text{rdfs:domain}, C_1) \in \mathcal{T} \implies (s, \text{rdf:type}, C_1) \in \mathcal{A})$
$\wedge\ ((p, \text{rdfs:range}, C_2) \in \mathcal{T} \implies (o, \text{rdf:type}, C_2) \in \mathcal{A})$ where $\mathcal{A} = \mathcal{A}_{\text{proposed}} \cup \mathcal{A}_{\text{init}}$ represents the complete, active ABox.

By compiling the TBox schema directly into the context window as a non-negotiable syntactic boundary, the LLM is transformed from an open-ended generator into a schema-conforming parser operating under a closed-world assumption. Any generated triple violating these domain and range constraints represents a semantic or syntactic fault, triggering immediate rejection and manual refinement in the peer consensus phase (Section 3.6).

Table 5 maps each slot to its repository artifact; Table 6 maps the six TPL_{sem} categories to ontology classes and properties.

Table 5: Ontology-constrained prompt assembly (all passes).

Slot	Symbol / artifact	Role in function Π
S1	o, `1.Antigone-Ontology.ttl`	Permitted vocabulary
S2	r, rules in `demo/demo_grc.ttl`	Global structural constraints
S3	t, TPL_{sem}	Category-specific instructions
S4	x, assigned verse range	Grounding text in `<TEXT>. . . </TEXT>`

Table 6: Semantic annotation categories in the Pass B prompt (ontology-guided).

Reading (prompt)	Class	Property
Intentions / objectives	`Motivation`, `MoralDecision`	`hasMotivation`, `makesMoralDecision`
Moral justifications	`EthicalPrinciple`, `MoralDecision`	`advocatesFor`
Emotional states	`Emotion`	`experiencesEmotion`
Semantic contradictions	`Conflict`	`hasConflict`, `conflictBetween`
Causal relationships	`Motivation`, `MoralDecision`, `owl:Thing`	`relatedTo`
Thematic signals	`Theme`	`developsTheme`, `hasTheme`

3.5 Observed LLM Behavior and Prompt Engineering.

During the initialization phase, we observed that LLMs naturally resist outputting strict, schema-compliant RDF without extensive conditioning. Common failure modes included malformed Turtle, untyped entities, invented predicates (e.g., `:hasFeeling` instead of `:experiencesEmotion`), and speaker-label confusion where models parsed stage directions like "CHORUS:" as dialogue. Providing the full Turtle ontology file within the XML `<ontology>` tags, combined with negative constraints, significantly reduced these syntactic hallucinations. However, semantic hallucinations remained a persistent issue. For example, models struggled with the collective identity of the Chorus or misinterpreted sarcastic dialogue as genuine statements of intent. This empirical behavior strictly necessitated the human peer-review phase that follows extraction.

Language models Used. R_{KG} contributors invoked various LLMs through chat services, including ChatGPT-5, Gemini 3.0, Claude, and DeepSeek. Model choice was left to each contributor. Identifiers were recorded in R_{KG} methodology reports, but not version-pinned in the repository. Published Turtle uses corpus-stable identifiers (e.g., `Theme_FateOfOedipusLine`) rather than the unstandardized ad-hoc labels initially generated by the models (e.g., `AntigoneMotivation1`); R_{AGG} normalized naming during merge.

Initial Quality Review. After each verse-range output, R_{KG} performed manual evaluation to verify: (i) Turtle syntax and mandatory typing; (ii) semantic grounding in the source passage; and (iii) the documentation of skipped or ambiguous lines. Contributors used inline `# REPORT` comments (e.g., *# REPORT: Skipped line 21 due to missing references*) to log verse-splitting problems or unresolved conflicts, which the R_{AGG} team later utilized to guide the merge process.

3.6 Aggregation and Publication

Here the aggregators (the members with role R_{AGG}) had to integrate TTL from multiple R_{KG} pipelines. Early tasks included creating `demo` reference files (expected naming, scene, speech, line flow) and assigning line ranges by *dramatic boundaries* rather than equal line counts, avoiding mid-speech cuts that fragment `Speech` individuals.

Table 7 documents challenges encountered during merge and the human mitigations applied.

Table 7: Common aggregation challenges and mitigations.

Challenge	Manifestation	Mitigation
Verse block layout	Textbook groups lines; wrong `lineNumber`	Manual checks; re-extract; `# REPORT`
Speaker labels	"CHORUS:" parsed as text	Prompt rules; speech audit
Inconsistent modeling	Emotions on `Line` vs. `Character`	R_{AGG} normalization
Edition skew	Merged/split Modern Greek verses	Ancient Greek ground truth; TV for text only
Semantics on variants	Themes on `TranslationVariant`	Consolidate on `Speech`/`Character`

Semantic consolidation policy: Ancient Greek structure carries interpretive authority. R_{AGG} moved interpretive links to canonical `Speech` and `Character`; `TranslationVariant` holds translated `text` only.

Example (R_{AGG} consolidation). Published files keep translation text on variants and interpretive links on speeches (from `01_prologue/`):

```
# translations/english/1_to_19.ttl
:TV_Line_001_en a :TranslationVariant ;
:text "Ismene, sweet Ismene, my dear sister!"^^xsd:string ;
:relatedTo :Line_001 .

# canonical/semantics.ttl
:Speech_Antigone_1_10
a :Speech ;
:developsTheme :Theme_FateOfOedipusLine, :Theme_FamilySuffering,
               :Theme_PoliticalDecreeThreat, :Theme_EnemiesHarm .
```

Following transcription and merge, R_{AGG} published modular Turtle to *Apache Jena Fuseki 6.x* (TDB2). Corpus contains 13 section folders. Each section follows the same layout:

```
Prologue/
  canonical/structure.ttl    # Pass A: Scene, Speech, Line (Greek)
  canonical/semantics.ttl    # Pass B: Theme, Emotion, Myth, ...
  translations/
    greek/1_to_19.ttl        # Pass C: TV_Line_*_gr
    english/1_to_19.ttl      # Pass C: TV_Line_*_en
    chinese/1_to_19.ttl      # Pass C: TV_Line_*_zh
```

Each contributor's files merge monotonically; blank nodes are forbidden by extraction rules. After upload, R_{AGG} ran validation to ensure syntactic correctness and ontology alignment before sharing the endpoint with R_{GUI}.

4 Evaluation

To ensure data quality, we implemented peer review of produced triples, ontology compliance checking, integrity SPARQL after each load, and predefined queries derived from competency questions.

4.1 Data Validation and Integrity Checks

For data validation and integrity check, we adopted a *Producer-Aggregator Consensus Model:* The R_{KG} (producers) and R_{AGG} (aggregators) act as a dual-gate consensus engine:

- **Gate 1: Syntactic & Local Semantic Grounding (R_{KG})**: Ensures that the extracted RDF graph is locally consistent with the specific dramatic passage (e.g., correct speaker attribution for line X).
- **Gate 2: Global Schema Integration & Normalization (R_{AGG})**: Resolves semantic discrepancies across adjacent verse blocks. For instance, if one student team labeled an emotion as :Sadness and another as :Grief, R_{AGG} executes a normalization pass mapped to the pre-coordinated subclasses of CIDOC-CRM E28_Conceptual_Object to maintain global graph cohesion.

SPARQL as Active Compiler/Linter: We use SPARQL validation as Continuous Integration (CI) Linters for KG assembly. Every pull request (merge of a Turtle file) must pass the SPARQL test suite to guarantee there are no orphan translation variants or speakerless speeches. In particular, after each Fuseki load, R_{AGG} executed integrity queries (Appendix A) to catch structural errors during the merge. For example, checking for orphan translation variants (which should return **zero** rows):

```
PREFIX : <http://example.org/antigone#>
SELECT ?tv WHERE {
  ?tv a :TranslationVariant ; :relatedTo ?line .
  FILTER NOT EXISTS { ?line a :Line }
}
```

Other integrity checks included: duplicate `lineNumber`; speeches missing a `speechBy` property; and overall property usage histograms.

4.2 Evaluation through Competency Questions

Team members contributed natural-language *competency questions*, manually translated into SPARQL by R_{AGG}. This serves the same functional role as structured retrieval in tragedy annotation projects [16], but here as *hand-authored graph patterns* over RDF rather than an automated natural-language interface. Below are three representative patterns (indicative list in Appendix A).

CQ: Who speaks at line 1?

```
PREFIX : <http://example.org/antigone#>
SELECT ?speech ?character WHERE {
  ?speech a :Speech ;
          :speechBy ?character ;
          :containsLine :Line_001 .
}
```

CQ: What are the Ancient Greek text and its corresponding Modern Greek and English translations for every line?

```
PREFIX : <http://example.org/antigone#>
SELECT ?line ?n ?ancient
       (SAMPLE(?modernRaw)  AS ?modern)
       (SAMPLE(?englishRaw) AS ?english)
WHERE {
  ?line a :Line ; :lineNumber ?n ; :text ?ancient .
  OPTIONAL {
    ?tvModern a :TranslationVariant ; :relatedTo ?line ;
              :text ?modernRaw .
    FILTER(CONTAINS(STR(?tvModern), "_GR"))
  }
  OPTIONAL {
    ?tvEnglish a :TranslationVariant ; :relatedTo ?line ;
               :text ?englishRaw .
    FILTER(CONTAINS(STR(?tvEnglish), "_EN"))
  }
}
GROUP BY ?line ?n ?ancient
ORDER BY ?n
```

CQ: What themes and mythological references are developed in the speech containing line 1?

```
PREFIX : <http://example.org/antigone#>
SELECT DISTINCT ?speech ?theme ?myth WHERE {
  ?speech a :Speech ; :containsLine ?line .
  ?line :lineNumber ?n .
  OPTIONAL { ?speech :developsTheme  ?theme }
  OPTIONAL { ?speech :referencesMyth ?myth }
  FILTER(?n = 1)
}
```

The R_{GUI} web client embeds these patterns for scene reading and line navigation (Figure 4(a)).

4.3 Knowledge Graph Statistics and Reproducibility

Table 8 summarizes the construction snapshot and representative class instantiation for the published corpus.

Edition numbering reaches 1353; 1259 canonical `Line` individuals are populated in the RDF corpus (the line navigator indexes the edition range for navigation).

Table 8: Corpus statistics and representative class instantiation.

Metric / Class	Value	Description / Role
General Corpus Statistics		
RDF triples	25,481	Total graph statements
Turtle files	185	Modular extraction files
Dramatic sections	13	Structural subdivisions
Edition line span	1–1353	Canonical verse numbering
Languages	el, en, zh	Included text layers
Core Class Individuals		
`Line`	1259	Canonical verse reference
`Speech`	291	Utterance hub
`Trans.Variant`	3500+	Multilingual text
`Theme`	100+	Interpretive concept
`Emotion`	50+	Affective reading
`Character`	27	Dramatic person

Reproducibility. The ontology, corpus modules, extraction rules (`demo_grc.ttl`), Pass B prompt (`pass_b_prompt.tex`), Fuseki configuration, bulk-load script, and web client are available in the project repository. Integrity and competency SPARQL queries are listed in Appendix A.

4.4 Evaluation of LLMs

Here we report a few results that access how good triple extraction with LLM were in our case. The R_{KG} files contained 41,654 raw triple occurrences, corresponding to 25,229 unique triples after duplicate removal. Translation related triples were excluded because the multilingual editions required substantial manual alignment and restructuring. This left 15,873 triples.

The R_{AGG} compared these proposals with the final corpus and classified them as accepted, modified, or rejected. The review followed a predefined annotation contract rather than independent annotation. Line identities, speech and scene membership, and the placement of semantic assertions on canonical line entities followed fixed representation rules, while equivalent semantic concepts were merged where necessary.

As detailed in Table 9, while over half of the proposed triples (53.74%) were accepted as-is, nearly a third (32.94%) were rejected entirely, and 13.32% required modification by human reviewers. These high rejection and modification rates primarily stemmed from the models struggling with collective identities (e.g., the Chorus), misinterpreting sarcasm, and hallucinating dramatic boundaries. This empirical breakdown strongly validates the necessity of our hybrid, human-in-the-loop workflow.

The final published corpus contains 25,481 triples because it also includes translation-related statements, structural and ontology declarations, and triples added or reconstructed during aggregation. Therefore, its total does not correspond directly to the accepted and modified LLM proposals.

4.5 Ablation Study

To assess the effect of ontology-constrained prompting, we conducted a small ablation on lines 110–210, comparing the constrained R_{KG} output with a free-form RDF extraction for the same passage,

Table 9: Disposition of evaluated LLM-proposed triples.

Disposition	Triples	Percentage
Accepted	8,530	53.74%
Modified	2,114	13.32%
Rejected	5,229	32.94%

excluding `TranslationVariant` triples. Both outputs were syntactically valid Turtle, yielding a malformed-Turtle rate of 0.00% in both cases. However, the free-form output produced 110 triples, of which 51 used predicates outside the project ontology contract (an unsupported-predicate rate of 46.36%). Because these assertions relied on generic namespaces and custom URI conventions, they exhibited no direct schema overlap with the canonical R_{AGG} graph and would require complete manual re-linking to be integrated. By contrast, the ontology-constrained R_{KG} output for the same passage produced 1,484 unique triples with zero unsupported predicates, of which 39.96% were accepted as-is and 12.47% were retained after modification. The free-form output did capture some broad literary content, including major figures such as Chorus, Creon, Polyneices, Eteocles, Zeus, Ares, and Nike, but it used non-canonical identifiers such as `ex:Line110` instead of `:Line_110`, direct CRM/FRBRoo predicates rather than the *Antigone* ontology predicates, and recovered only 6 of the 97 canonical line nodes in the passage. This ablation indicates that ontology-constrained prompting mainly improved schema compatibility, line-level integration, and downstream corpus usability, while human review remained necessary for interpretive correction.

5 Access Services and Interactive Exploration

Several access services over the produced KG are possible. Apart from plain SPARQL, one could provide a *keyword search* service (as in [9, 12, 17]), a *faceted exploration* method (as in [10, 18]), a *Query by Example* method (as in [3, 7]), a CIDOC-CRM dedicated QA service (as in [11]), or a RAG-based QA service (as in [20]). In this project, we focused on a dedicated (for tragedies) *interactive system*, that allows users to visually explore the KG, through click-based interactions, facilitating a non-linear, multifaceted reading of the tragedy. From an educational and analytical standpoint, that means that the KG provides a novel, non-linear way to read a tragedy. Unlike a traditional printed edition or a flat digital text, the KG enables classicists and students to instantly surface macroscopic trends, such as visualizing which specific scenes contain the highest density of references to "Divine Law" versus "Civic Law," or tracking the emotional trajectory of Creon line-by-line. The reference-line multilingual pattern ensures that an English reader can navigate the interpretive depth of the original Greek structure without needing to read Ancient Greek, bridging a significant accessibility gap in classical studies.

To this end team R_{GUI} developed a lightweight JavaScript read client over the published SPARQL endpoint. Note that the client acts as an extra validation layer that exposes the graph's structural, multilingual, and interpretive dimensions. An overview is given in Figure 4.

- *Multilingual Line Navigation:* Retrieves the canonical Greek text alongside translation variants, speaker metadata, and dramatic context for any selected verse. The corresponding screen is shown in Figure 4(a).
- *Interactive Context:* Surfaces scene-level and speech-level semantic annotations, linking specific lines to their associated themes, emotions, and mythological references is shown in Figure 4(b).
- *Corpus-level Analytics:* Aggregates graph statistics (e.g., live triple counts, 25,481) and generates progressive knowledge networks connecting characters, themes, and events. The corresponding screens are shown in Figures 4(c) and 4(d).

The client successfully confirms that the competency-question patterns execute correctly against the merged corpus. For example, selecting line 1 dynamically executes a SPARQL query to return `:Speech_Antigone_1_10`, the speaker `:Antigone_Character_Antigone`, and interpretive concepts such as `Theme_FateOfOedipusLine`.

5.1 Discussion and Qualitative Insights

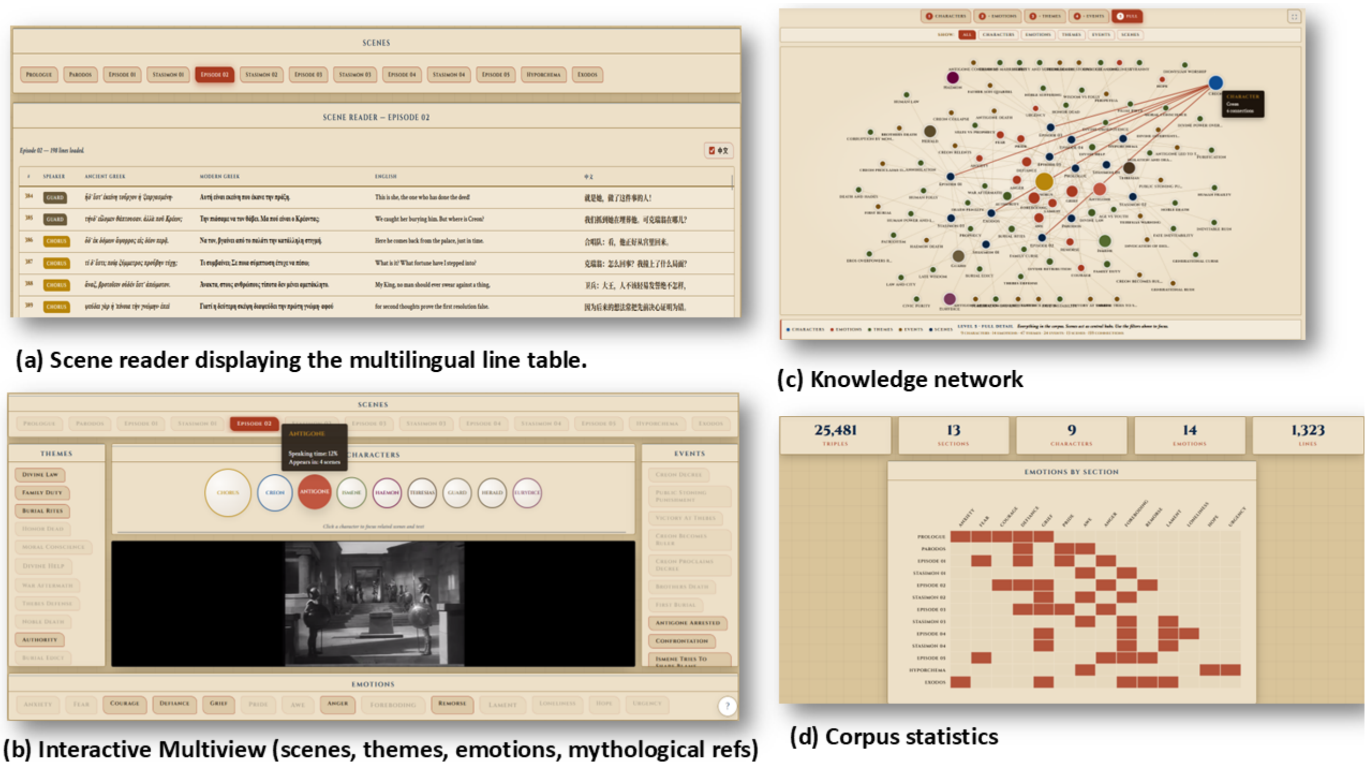
The Necessity of Human Oversight. The corpus was constructed over approximately six months, representing more than 350+ person-hours across the participating teams. The experience of the R_{KG} and R_{AGG} teams showed that the main effort lay not in the initial generation of RDF, which LLMs produced rapidly, but in the subsequent validation and correction of dramatic boundaries, speaker attribution, and interpretive annotations (as we described in §4.4).

Human Oversight. The main cost was the subsequent validation of dramatic structure and interpretive annotations. Recurring errors included merged speeches, incorrect speaker attribution, particularly in long dialogue passages and choral sections, and overly generic semantic labels. For example, models sometimes assigned broad concepts such as "Sadness" instead of ontology-specific themes grounded in the dramatic context. These observations indicate that ontology-constrained prompting improved output consistency but did not make the generated RDF suitable for direct publication. Human review remained necessary for speech segmentation, speaker attribution, multilingual alignment, and interpretive decisions.

6 Concluding Remarks

This paper presented a collaborative workflow for representing Sophocles' *Antigone* as a knowledge graph that combines dramatic structure, multilingual line text, and interpretive annotations over speeches. The workflow uses ontology-constrained LLM extraction to propose RDF triples, followed by producer review, aggregation, normalization, and SPARQL-based validation. The resulting corpus contains 25,481 RDF triples, 1259 canonical line individuals, 291 speeches, 3500+ translation variants, and 185 modular Turtle files, demonstrating the feasibility of semi-automated, schema-guided construction for literary drama at verse scale.

The main methodological finding is that LLMs are effective as ontology-constrained triple proposers, but not as autonomous literary annotators. Human review remained necessary for validating dramatic boundaries, speaker attribution, multilingual alignment, and interpretive claims. This supports a hybrid construction model in which the ontology constrains generation, while human reviewers preserve structural consistency and scholarly interpretation. In this setting, the graph does not claim to encode a single authoritative reading of *Antigone*; rather, it records explicit interpretive



(a) Scene reader displaying the multilingual line table.

(c) Knowledge network

(b) Interactive Multiview (scenes, themes, emotions, mythological refs)

(d) Corpus statistics

Figure 4: Overview of the interactive access services. The client provides (a) multilingual line navigation, (b) interactive context for scenes, themes, emotions and muythological refs, (c) the knowledge network of play entities, and (d) corpus-level analytics.

choices through typed entities and natural-language descriptions reviewed before publication.

From a structural perspective, our approach to modeling multilingual literary text proved highly effective. Assigning one stable Ancient Greek reference line per edition verse, and linking translations as `TranslationVariant` nodes, simplified aggregation and cross-lingual querying. This avoided duplicating the dramatic structure for each language and allowed interpretive annotations to remain anchored in the canonical Greek graph. Similarly, chunking the play by dramatic boundaries, rather than fixed line counts, helped preserve the context needed to annotate complete speeches.

The read client demonstrated that the resulting KG supports the intended access patterns: scene navigation, line-level lookup, translation joins, and retrieval of speakers, themes, emotions, and mythological references. In this sense, the interface served as an access service and practical validation layer, showing that the constructed graph can support both close reading and broader exploratory analysis of the tragedy. The current version of the prototype application is available at this link [5].

Applicability. We believe this methodology can extend to other Greek tragedies that share the same dramatic hierarchy and would benefit from multilingual presentation.

Future Research and Work. One direction for future research is to investigate the automation of the structural validation earlier in the merge process. Other possible extensions include:

(a) Aligning our thematic vocabulary with shared knowledge organization systems, so that interpretive concepts can be compared and linked across projects instead of remaining local labels,
(b) Expanding multilingual coverage. The current graph structure lowers the cost of adding further languages: because each edition line carries a single stable identity and translations attach to it only through `TranslationVariant` and `relatedTo`, a new language can be introduced as an additional translation layer without touching the canonical structure or re-annotating interpretive content,
(c) Publication as linked open data, as well as in CIDOC-CRM Portal[6] [14]. At present, no comparable line- and speech-level tragedy graph exists to link to, so this step depends on the emergence of similarly granular resources, of which fragmentary-tragedian graphs [4] are an early example, (d) Leverage the KG for offering advanced and verifiable QA services, by exploiting recent natural language question to SPARQL techniques, like [15].

Acknowledgments

Many thanks to all students who participated in this project, namely: Ioannis Titos Papadakis, Stavros Petrakis, Michail Iason Giannakopoulos, Dimitrios Marinos Grammenidis, Patrick Zygmuntowicz, Christina Ntoka, Konstantinos Anagnostakis and Zacharias Varouchakis.

[5] https://demos.isl.ics.forth.gr/antigone/

[6] https://demos.isl.ics.forth.gr/CIDOC-CRM_Portal/

References

[1] [n. d.]. TEI P5: Guidelines for Electronic Text Encoding and Interchange. https://tei-c.org/. Accessed 2026.

[2] Gregory R. Crane (Ed.). 2026. *Perseus Digital Library*. https://www.perseus.tufts.edu/

[3] Akritas Akritidis and Yannis Tzitzikas. 2024. Querying knowledge graphs through positive and negative examples and feedback. *Journal of Intelligent Information Systems* 62, 5 (2024), 1165–1186. doi:10.1007/s10844-024-00846-z

[4] Andreas P Antonopoulos, Stylianos Chronopoulos, Nikolaos Ntaliakouras, Panagiota Taktikou, Anastasia Psomiadou, and Iraklis Markelis. 2023. Developing a Database for the Greek Fragmentary Tragedians. *Digital Classics Online* (2023), 15–29.

[5] Marijke Beersmans, Alek Keersmaekers, Evelien de Graaf, Tim Van de Cruys, Mark Depauw, and Margherita Fantoli. 2024. "Gotta catch 'em all!": Retrieving people in Ancient Greek texts combining transformer models and domain knowledge. In *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024)*, John Pavlopoulos, Thea Sommerschield, Yannis Assael, Shai Gordin, Kyunghyun Cho, Marco Passarotti, Rachele Sprugnoli, Yudong Liu, Bin Li, and Adam Anderson (Eds.). Association for Computational Linguistics, Hybrid in Bangkok, Thailand and online, 152–164. doi:10.18653/v1/2024.ml4al-1.16

[6] Gregory Crane. 1996. Building a Digital Library: The Perseus Project as a Case Study in the Humanities. In *Proceedings of the First ACM International Conference on Digital Libraries*. 3–10.

[7] Gonzalo Diaz, Marcelo Arenas, and Michael Benedikt. 2016. SPARQLByE. *Proceedings of the VLDB Endowment* 9, 13 (2016), 1533–1536. doi:10.14778/3007263.3007302

[8] Martin Doerr. 2005. The CIDOC CRM, an ontological approach to schema heterogeneity. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

[9] Shady Elbassuoni and Roi Blanco. 2011. Keyword search over RDF graphs. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11)*. Association for Computing Machinery, New York, NY, USA, 237–246. doi:10.1145/2063576.2063615

[10] Sébastien Ferré. 2017. Sparklis: An expressive query builder for SPARQL endpoints with guidance in natural language. *Semantic Web* 8, 3 (2017), 405–418. doi:10.3233/SW-150208

[11] Nikos Gounakis, Michalis Mountantonakis, and Yannis Tzitzikas. 2023. Evaluating a radius-based pipeline for question answering over cultural (CIDOC-CRM based) knowledge graphs. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*. 1–10.

[12] Giorgos Kadilierakis, Pavlos Fafalios, Panagiotis Papadakos, and Yannis Tzitzikas. 2020. Keyword search over RDF using document-centric information retrieval systems. In *European Semantic Web Conference*. Springer, 121–137.

[13] Lars-Peter Meyer, Claus Stadler, Johannes Frey, Norman Radtke, Kurt Junghanns, Roy Meissner, Gordian Dziwis, Kirill Bulert, and Michael Martin. 2024. *LLM-assisted Knowledge Graph Engineering: Experiments with ChatGPT*. Springer Fachmedien Wiesbaden, 103–115. doi:10.1007/978-3-658-43705-3_8

[14] Michalis Mountantonakis. 2023. Why We Need Ontology-Specific Data Portals: A Case Study for CIDOC-CRM. In *Proceedings of the CIDOC-CRM Special Interest Group (SIG) Meeting (CEUR Workshop Proceedings, Vol. 3540)*. CEUR-WS.org, paper9. https://ceur-ws.org/Vol-3540/paper9.pdf

[15] Michalis Mountantonakis and Yannis Tzitzikas. 2025. Generating SPARQL Queries over CIDOC-CRM using a Two-Stage Ontology Path Patterns Method in LLM Prompts. *ACM Journal on Computing and Cultural Heritage* 18, 1 (2025), Article 12. doi:10.1145/3708326

[16] Gloria Mugelli, Federico Boschetti, Andrea Bellandi, Riccardo Del Gratta, Anas Fahad Khan, Andrea Taddei, et al. 2021. Annotating ritual in ancient Greek tragedy: a bottom-up approach in action. *Digital Humanities Quarterly* 15, 1 (2021).

[17] Christos Nikas, Giorgos Kadilierakis, Pavlos Fafalios, and Yannis Tzitzikas. 2020. Keyword search over RDF: is a single perspective enough? *Big Data and Cognitive Computing* 4, 3 (2020), 22.

[18] Maria-Evangelia Papadaki and Yannis Tzitzikas. 2024. Unifying faceted search and analytics over RDF knowledge graphs. *Knowledge and Information Systems* 66, 7 (2024), 3921–3958.

[19] Juan-Antonio Pastor-Sánchez, Efstratios Kontopoulos, Tomás Saorín, Thomas Bebis, and Sándor Darányi. 2021. Greek Mythology as a Knowledge Graph: From Chaos to Zeus and Beyond. https://api.semanticscholar.org/CorpusID:235657768

[20] Iordanis Sapidis, Valantis Zervos, and Michalis Mountantonakis. 2025. and QA over Documents Using LLMs, RAG and Knowledge Graph Verbalization. In *New Trends in Theory and Practice of Digital Libraries: TPDL 2025 Short Papers and Workshops, Tampere, Finland, September 23–26, 2025, Proceedings*. Springer Nature, 248.

[21] Sophia Sideri, Emmanouil Patronakis, Evangelos Nikiforos, Isidoros Chatzichrysos, Apostolos Baniotis, Valantis Zervos, Stavros Tzormpatzakis, Michalis Saridakis, Iosif Oikonomakis, Michalis Mountantonakis, Pavlos Fafalios, and Yannis Tzitzikas. 2026. To Automate or Not to Automate the Transcription of Ancient Earthquakes: Toward a Global Knowledge Graph About Ancient Earthquakes. In *Knowledge Graphs*. Springer Nature Singapore, Singapore, 390–405.

[22] Text Encoding Initiative Consortium. 2025. TEI P5: Guidelines for Electronic Text Encoding and Interchange. https://tei-c.org/guidelines/. Accessed: 2026-07-11.

[23] Diamanto Tzanoulinou, Loukas Triantafyllopoulos, and Vassilios S. Verykios. 2025. Harnessing Language Models for Studying the Ancient Greek Language: A Systematic Review. *Machine Learning and Knowledge Extraction* 7, 3 (2025). doi:10.3390/make7030071

[24] Yannis Tzitzikas, Michalis Mountantonakis, Pavlos Fafalios, and Yannis Marketakis. 2022. CIDOC-CRM and Machine Learning: A Survey and Future Research. *Heritage* 5, 3 (2022), 1612–1636. doi:10.3390/heritage5030084

[25] Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, Yong Jiang, and Wenjuan Han. 2024. ChatIE: Zero-Shot Information Extraction via Chatting with ChatGPT. arXiv:2302.10205 [cs.CL] https://arxiv.org/abs/2302.10205

A Competency Questions and SPARQL Queries

Below follow the query (i) of the introductory section.

```sparql
PREFIX : <http://example.org/antigone#>
SELECT ?scene ?lineNo ?english ?mandarin
       (GROUP_CONCAT(DISTINCT STRAFTER(STR(?lawConcept), "#"); separator=", ") AS ?civicLawConcepts)
       (GROUP_CONCAT(DISTINCT STRAFTER(STR(?myth), "#"); separator=", ") AS ?mythRefs)
WHERE {
  VALUES ?lawConcept {
    :Theme_AuthorityAndLaw :Theme_BurialEdict
    :Theme_CivicStabilityAfterCrisis :Theme_LeadershipAndLaw
    :Theme_PatriotismAndCivicLoyalty :Theme_AuthorityAndInterrogation
    :Theme_SuspicionAndAuthority :Theme_DefianceOfEdict
    :Theme_HumanLaw :Theme_Authority
    :Theme_AbsoluteAuthority :Theme_CivicOrderVsAnarchy
    :Theme_CivicPurityAndMiasma :Theme_CivicPollution
    :Theme_LawAndCity :Theme_StateEdict :Theme_StateLaw
    :Theme_TyrannicalLaw :Theme_ObedienceToAuthority
  }
  :Antigone_Character_Creon :experiencesEmotion :Emotion_Anger .

  ?scene :hasSpeech ?speech .
  ?speech :speechBy :Antigone_Character_Creon ;
          :containsLine ?line .
  ?line :lineNumber ?lineNo .

  {
    ?speech :hasTheme ?lawConcept .
  }
  UNION {
    ?speech :developsTheme ?lawConcept .
  }
  UNION {
    ?line :hasTheme ?lawConcept .
  }
  UNION {
    ?line :relatedTo ?lawConcept .
  }

  ?enVariant :relatedTo ?line ;
             :text ?english .
  FILTER(STRENDS(STR(?enVariant), "_EN"))
  ?zhVariant :relatedTo ?line ;
             :text ?mandarin .
  FILTER(STRENDS(STR(?zhVariant), "_ZH"))

  OPTIONAL {
    { ?scene :referencesMyth ?myth . }
    UNION { ?speech :referencesMyth ?myth . }
    UNION { ?line :referencesMyth ?myth . }
  }
}
GROUP BY ?scene ?lineNo ?english ?mandarin
ORDER BY ?scene ?lineNo
```