



AI FOR INDUSTRY

The Italian Institute of
Artificial Intelligence

Dai documenti alla conoscenza: RAG e Knowledge Graph



Ministero delle Imprese
e del Made in Italy



Ministero
dell'Economia
e delle Finanze



DIPARTIMENTO
PER LA TRASFORMAZIONE
DIGITALE

Milano, 14.07.2026



An Integrated, Multi-Divisional Architecture



Struttura



Retrieval — recupero

- Data una **query testuale** dell'utente, si cercano nei documenti più simili per rispondere alla domanda.
- I documenti vengono selezionati via **similarità semantica**:



Augmented — arricchimento

I documenti significativamente vicini alla query di input vengono combinati assieme e strutturati per comporre un **contesto** ottimale di conoscenza



Generation — generazione

L'AI genera la risposta appoggiandosi al contesto generato precedentemente, non solo alla sua memoria.

Limiti principali

Contesto frammentato

Il chunking spezza relazioni tra entità che vivono in documenti — o paragrafi — diversi.

No ragionamento multi-hop

La similarità vettoriale non collega $A \rightarrow B \rightarrow C$: fallisce sulle domande che richiedono catene di inferenza.

Nessuna visione d'insieme

Domande globali ("quali temi ricorrono nel corpus?") non hanno un chunk che le risponda

Solo RAG vector search sulle trame

Film di Leonardo di Caprio

In base al contesto fornito, i film con Leonardo DiCaprio presenti nel sistema sono:

- **Inception** (2010) — regia di Christopher Nolan, genere Action/Thriller/Science Fiction/Mystery/Adventure, rating 8.1
- **Shutter Island** (2010) — regia di Martin Scorsese, genere Drama/Thriller/Mystery, rating 7.8
- **The Wolf of Wall Street** (2013) — regia di Martin Scorsese, genere Crime/Drama/Comedy, rating 7.9
- **The Revenant** (2015) — regia di Alejandro González Iñárritu, genere Western/Drama/Adventure/Thriller, rating 7.3
- **Django Unchained** (2012) — regia di Quentin Tarantino, genere Drama/Western, rating 7.8

Questi sono i titoli con Leonardo DiCaprio disponibili nel contesto a mia disposizione.



Il RAG costruisce le risposte basandosi sui testi semanticamente vicini alla domanda dell'utente

Il Knowledge Graph aiuta a collegare i documenti estratti per similarità semantica superando alcuni limiti del RAG

- Le **informazioni sono strutturate** in un network con entità e relazioni
- **L'AI naviga il network** per collezionare le informazioni più pertinenti
- **Risposte che spiegano** non solo citazioni di testi, ma il perché delle connessioni.

● Graph RAG vector search + knowledge graph

Film di Leonardo DiCaprio

Secondo le informazioni disponibili, Leonardo DiCaprio ha recitato nei seguenti film:

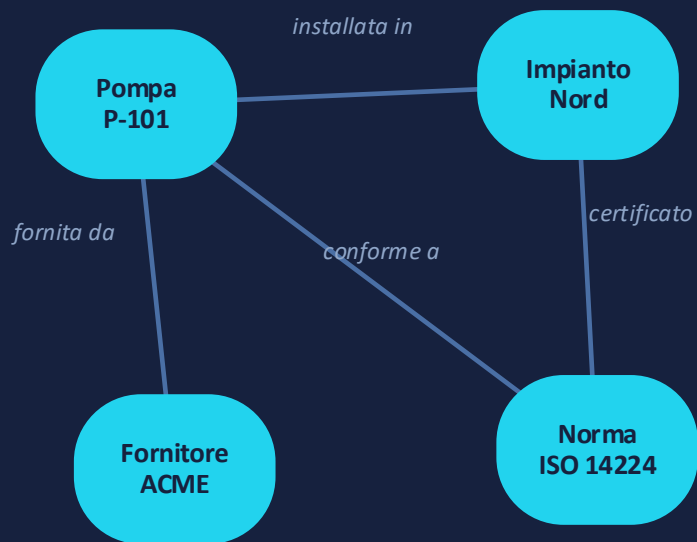
- **Inception** (2010) – regia di Christopher Nolan
- **Titanic** (1997)
- **Shutter Island** (2010) – regia di Martin Scorsese
- **The Wolf of Wall Street** (2013) – regia di Martin Scorsese
- **Django Unchained** (2012) – regia di Quentin Tarantino
- **The Revenant** (2015) – regia di Alejandro González Iñárritu

► Fonti: 5 documenti + knowledge graph

Il GraphRAG integra i chunk (usati dal RAG) con una rete di entità e relazioni.

Il knowledge graph modella relazioni esplicite tra entità e documenti, abilitando connessioni non derivabili dalla sola prossimità semantica

Entità + relazioni tipizzate



Query multi-hop: "Quali asset dell'Impianto Nord dipendono da fornitori non certificati ISO?"



Ragionamento multi-hop

Le domande attraversano più relazioni: il grafo viene percorso esplicitamente, hop dopo hop.



Query globali sul corpus

Community detection e summarization gerarchica rispondono a domande d'insieme, non solo puntuali.



Retrieval deterministico¹

Il processo di retrieval del contesto è basato su percorsi di ragionamento espliciti, molto più che sulla similarità semantica



Tracciabilità e spiegabilità

Ogni risposta è riconducibile a un percorso nel grafo: fonti verificabili, meno allucinazioni.

1. In un GraphRAG il retrieval può essere in tutto o in parte deterministico.

La qualità del GraphRAG dipende dalla qualità del knowledge graph — e costruirlo bene ha due colli di bottiglia



PAIN POINT 1

Definizione dell'ontologia

- **Schema design non banale**

Quali entità, quali relazioni, con che granularità? Ogni dominio richiede un modello dedicato.

- **Costi computazionali elevati**

L'estrazione di entità e relazioni richiede chiamate LLM su tutto il corpus: costo che scala con i token, non con i documenti.

- **Re-indexing costoso**

Se l'ontologia cambia, il grafo va in gran parte ricostruito: le iterazioni sullo schema sono care.



PAIN POINT 2

Human-in-the-loop (HITL)

- **Validazione degli esperti**

L'estrazione automatica produce entità duplicate, relazioni spurie, tipi errati: serve revisione di dominio.

- **Effort continuativo, non one-shot**

Il grafo va mantenuto: nuovi documenti, entity resolution, deduplicazione richiedono cicli di curation periodici.

- **Tempo esperto = risorsa scarsa**

Gli SME hanno poca disponibilità: il processo HITL va progettato per minimizzare il loro carico (review a campione, active learning).



Il RAG non sempre soddisfa requisiti dell'utente

Ottimo per Q&A puntuale su testo; fragile su relazioni, multi-hop e domande globali.



GraphRAG aggiunge struttura al retrieval, abilitando ricerche più complesse

Entità e relazioni esplicite abilitano ragionamento navigabile, spiegabile e verificabile.



Il grafo è un asset, e ha un costo

Ontologia e HITL vanno pianificati come progetto: budget computazionale + tempo degli esperti.



Partire in piccolo

Un sottodominio ad alto valore, schema minimo, cicli di curation brevi — poi estendere.



AI FOR INDUSTRY

The Italian Institute of
Artificial Intelligence

Demo session



The screenshot shows the 'Movies GraphRAG' interface. At the top, there are tabs for 'Chat' and 'Grafo'. On the right, there are tabs for 'Dataset' and 'Movies'. Below the tabs is a search bar with the placeholder 'Scrivi una domanda...' and an 'Invia' button. The interface is divided into two columns: 'Solo RAG' (vector search sulle trame) and 'Graph RAG' (vector search + knowledge graph). Both columns show the same query: 'quali sono i film nel dataset in cui recita di caprio?'. The 'Solo RAG' column lists four movies: Django Unchained (2012), Inception (2010), Shutter Island (2010), and The Wolf of Wall Street (2013). The 'Graph RAG' column lists five movies: Django Unchained, Inception, Shutter Island, The Revenant, The Wolf of Wall Street, and Titanic. Both columns include a note: 'Questi dati sono estratti dal catalogo completo del sistema e confermati dalle relazioni attore-film.' and a footer: 'Fonti: 5 documenti'.

Solo RAG vector search sulle trame

quali sono i film nel dataset in cui recita di caprio?

Leonardo DiCaprio recita nei seguenti film del dataset:

- Django Unchained (2012)
- Inception (2010)
- Shutter Island (2010)
- The Wolf of Wall Street (2013)

Questi dati sono estratti dal catalogo completo del sistema e confermati dalle relazioni attore-film.

Fonti: 5 documenti

Graph RAG vector search + knowledge graph

quali sono i film nel dataset in cui recita di caprio?

Leonardo DiCaprio ha recitato nei seguenti film nel dataset:

- Django Unchained
- Inception
- Shutter Island
- The Revenant
- The Wolf of Wall Street
- Titanic

Questi film sono confermati dalle relazioni "starring" e "acted_in" nel knowledge graph, che rappresentano la fonte autoritativa.

Fonti: 5 documenti + knowledge graph