

LLM-Text2KG: Um *Pipeline* Orientado por Ontologias para Construção e Avaliação de Grafos de Conhecimento Extraídos de Documentos Institucionais

Giovanna Verissimo
Universidade Federal do Ceará
giovannaxavier04@gmail.com

Julio Cesar Marinho
Universidade Federal do Ceará
juliocesarliramarinho@alu.ufc.br

Marlon Duarte
Universidade Federal do Ceará
marlon493408@gmail.com

Amanda D. Pires Venceslau
Universidade Federal do Ceará
amanda.pires@ufc.br

Vania Vidal
Universidade Federal do Ceará
vaniap.vidal@gmail.com

Wellington Franco
Universidade Federal do Ceará
wellington@crateus.ufc.br

Carlos Caminha
Universidade Federal do Ceará
caminha@ufc.br

ABSTRACT

Institutional documents, such as meeting minutes, contain valuable organizational knowledge, but their unstructured nature hinders systematic analysis. This work presents LLM-Text2KG, an ontology-guided pipeline for constructing knowledge graphs from institutional documents. The pipeline was evaluated on ten meeting minutes from the Acaraú River Basin Committee using six large language models under *zero-shot* and *few-shot* strategies. The evaluation combined *Precision*, *Recall*, and *F1-score* against a manually annotated *gold standard*, ontology-conformity validation, competency questions expressed as SPARQL queries, and an ablation of the ontology representation. Gemma 4 31B in *few-shot* achieved the best performance, with *F1-score* of 0.650 for entities and 0.498 for relations, and 99.37% ontology conformity. The competency questions achieved 41.0% overall coverage. Replacing the Turtle ontology with a structured textual list reduced relation *F1-score* from 0.464 to 0.102 in *zero-shot* and from 0.498 to 0.265 in *few-shot*. The results indicate that ontology representation influences extraction quality and that semantic relation extraction remains the main challenge.

KEYWORDS

Knowledge Graph, Large Language Models, Prompt Engineering, Ontology-Guided Extraction, Competency Questions

1 INTRODUÇÃO

As organizações públicas e privadas produzem continuamente documentos narrativos, como atas, deliberações, relatórios e normativas, que registram o conhecimento histórico e o processo decisório institucional. Entretanto, sua natureza não estruturada dificulta a recuperação de informações e a análise

automatizada, o que representa um desafio para a gestão do conhecimento na Web [2, 22]. Uma alternativa para estruturar esse conteúdo é a construção de grafos de conhecimento, do inglês *Knowledge Graphs* (KGs), que representam entidades e relações em redes semânticas, viabilizando recuperação de informação e consultas estruturadas [9, 23].

De acordo com [9], a construção de KGs a partir de textos não estruturados envolve três desafios: extração semântica de entidades e relações em linguagem natural ambígua [13], alinhamento ontológico para garantir conformidade com um esquema semântico [20] e avaliação semântica, pois métricas baseadas na correspondência exata de triplas não capturam equivalências semânticas [13]. O avanço dos Modelos de Linguagem de Grande Escala (LLMs) ampliou as possibilidades para enfrentar esses desafios, reduzindo a dependência de anotação manual ao explorar sua capacidade de extrair relações semânticas diretamente a partir de instruções [12].

No semiárido brasileiro, os comitês de bacias hidrográficas, instituídos pela Lei nº 9.433/1997¹ para reunir o poder público, usuários e sociedade civil na deliberação sobre a alocação da água e na mediação de conflitos [4], produzem atas que registram atores, deliberações e ações em bacias vulneráveis à escassez hídrica [24]. Seu caráter narrativo e não estruturado torna esses documentos um caso representativo dos três desafios mencionados, o que motiva a investigação proposta neste trabalho.

Para enfrentar esses desafios, este trabalho apresenta o LLM-Text2KG, um *pipeline* orientado por ontologias para a construção automática de KGs a partir de documentos institucionais. Além do *pipeline*, realiza-se um estudo experimental controlado para avaliar a capacidade de diferentes LLMs de extrair entidades e relações alinhadas a um esquema ontológico por meio de instruções contextuais. A abordagem é avaliada em atas do Comitê da Bacia Hidrográfica do Acaraú [18].

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

¹<https://www.gov.br/ana/pt-br/assuntos/gestao-das-aguas/politica-nacional-de-recursos-hidricos>

As principais contribuições deste trabalho são: (i) o LLM-Text2KG, um *pipeline* orientado por ontologias que realiza a extração de entidades e relações em duas etapas e produz representações serializáveis em RDF; (ii) uma avaliação experimental de seis LLMs, de diferentes famílias e portes, sob estratégias *zero-shot* e *few-shot*, mantendo constantes o corpus, a ontologia e o protocolo de avaliação; e (iii) um protocolo de avaliação multidimensional que combina métricas de correspondência com o padrão-ouro, conformidade ontológica, questões de competência e uma ablação da representação da ontologia em Turtle e lista textual, avaliando a fidelidade das extrações, a aderência ao esquema, a utilidade semântica dos grafos e o efeito da representação da ontologia.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3, descreve o pipeline LLM-Text2KG; a Seção 4, o estudo de caso, protocolo experimental, a configuração experimental e os resultados de extração, conformidade ontológica, ablação, questões de competência e comparação exploratória com as *baselines*; a Seção 5, as limitações; e a Seção 6, as conclusões e as direções para trabalhos futuros.

2 TRABALHOS RELACIONADOS

A construção automática de grafos de conhecimento a partir de textos não estruturados tem avançado com o uso de LLMs, envolvendo estratégias de extração, consolidação de entidades, orientação ontológica e avaliação dos grafos produzidos. Em [25] investigam modelos generalistas na construção de grafos a partir de textos não estruturados, comparando o REBEL, especializado na extração conjunta de entidades e relações, ao ChatGPT com diferentes níveis de detalhamento de *prompt*. A avaliação manual mostrou que instruções mais detalhadas, incluindo a geração de uma ontologia OWL, melhoram a qualidade dos grafos produzidos. O KGGen [15], por sua vez, busca reduzir redundância e esparsidade por meio de extração, agregação e resolução de entidades, combinando agrupamento semântico e refinamento via LLM. No benchmark MINE-1, o método superou GraphRAG e OpenIE, alcançando 66,07% de retenção de informação e elevada validade das triplas em avaliação manual.

Abordagens recentes também buscam aumentar a consistência das informações extraídas. O ODKE+ [10] utiliza fragmentos dinâmicos de ontologia no *prompt* e integra recuperação de evidências, extração, validação e normalização. Um segundo LLM valida se os fatos produzidos são sustentados pelo contexto. Em ambiente de produção, o sistema alcançou 98,8% de *precision* após corroboração. O GraphJudge [8] combina *denoising* centrado em entidades com um LLM ajustado por *fine-tuning* que atua como juiz, filtrando triplas incorretas ou não fundamentadas. Em conjuntos gerais e especializados, apresentou ganhos de *precision* e F1 em relação a métodos como KGGen e iText2KG.

No contexto da avaliação, o Text2KGBench-LettrIA [21] revisa o Text2KGBench [14], benchmark para avaliação de sistemas Text-to-KG orientados por ontologias. A revisão reorganiza 19 ontologias e reúne mais de 14.000 triplas reanotadas

por especialistas. Além do *F1-score*, considera conformidade ontológica, restrições de domínio e alcance, validade estrutural e alucinação. Os resultados mostram que modelos menores com *fine-tuning* supervisionado superam os modelos proprietários avaliados em zero-shot, destacando a importância do alinhamento entre modelos, dados e esquema ontológico.

Em conjunto, esses trabalhos cobrem diferentes aspectos da construção de KGs, mas sob configurações distintas. O KGGen opera em domínio aberto, sem restrições ontológicas predefinidas; o ODKE+ emprega orientação ontológica, porém com módulos adicionais de recuperação e validação em cenários de larga escala; e o GraphJudge depende de *fine-tuning* para filtrar triplas. Já o Text2KGBench-LettrIA concentra-se na avaliação de diferentes LLMs para sistemas Text-to-KG em cenários de benchmark. Nesse contexto, o LLM-Text2KG investiga a extração orientada diretamente por uma ontologia fornecida no *prompt*, sem *fine-tuning* ou validadores aprendidos, aplicada a documentos institucionais em português e considerando diferentes estratégias de *prompting* e representações do esquema ontológico.

3 LLM-TEXT2KG

O LLM-Text2KG é um *pipeline* orientado por ontologias para a construção automática de grafos de conhecimento a partir de documentos institucionais. A abordagem integra documentos em linguagem natural, uma ontologia de domínio e estratégias de *prompting* para orientar a extração de entidades e relações e sua posterior representação em RDF.

O *pipeline* é organizado em três módulos sequenciais: (i) Pré-processamento e Configuração de *Prompt*, (ii) Inferência Semântica com LLMs e (iii) Construção do Grafo de Conhecimento. Essa organização permite sua adaptação a diferentes domínios por meio da substituição da ontologia de entrada, mantendo a mesma arquitetura de extração. A Figura 1 apresenta uma visão geral do processo.

3.1 Entradas do Sistema

O sistema recebe duas entradas principais: (i) documentos institucionais em linguagem natural, tais como atas de reunião e relatórios, cujo conteúdo pode ser obtido a partir de qualquer formato de arquivo, desde que seja possível sua conversão para texto plano; e (ii) uma ontologia de domínio serializada no formato *Turtle* (TTL)², que atua como esquema estrutural para a extração. A ontologia define formalmente o vocabulário do domínio por meio de classes (`owl:Class`) e propriedades de objeto (`owl:ObjectProperty`), estabelecendo os tipos de entidades e as relações semânticas admissíveis entre elas. Esses elementos constituem o esquema que orienta a extração do grafo de conhecimento ao longo de toda a *pipeline*, conforme detalhado nas subseções seguintes.

²<https://www.w3.org/TR/turtle/>

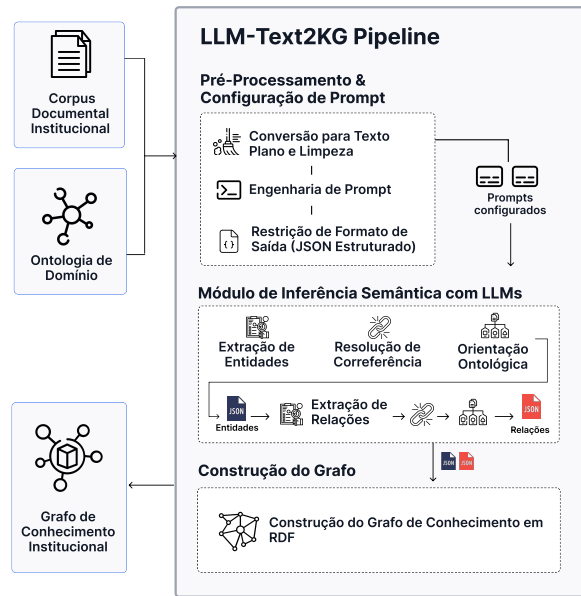


Figura 1: Arquitetura do LLM-Text2KG pipeline, que recebe como entrada um corpus documental e uma ontologia, realiza pré-processamento e configuração de prompts, executa a extração semântica com LLMs e, em seguida, constrói o grafo em RDF.

3.2 Módulo de Pré-processamento e Configuração de Prompt

3.2.1 Pré-processamento Textual. Essa etapa converte o conteúdo dos documentos para texto plano e aplica uma etapa de limpeza para remover artefatos tipográficos, quebras de linha desnecessárias e demais ruídos decorrentes da extração. No estudo de caso, os documentos originais são arquivos PDF e seu conteúdo é obtido por extração direta de texto.

3.2.2 Engenharia de Prompts. Os prompts foram elaborados iterativamente a partir do esquema CO-STAR [3], adotando os componentes *Context* (contexto da tarefa), *Objective* (objetivo da extração) e *Response* (formato rígido de saída). Versões preliminares foram aplicadas às atas para identificar problemas como saídas JSON inválidas, duplicação de entidades, ambiguidades na interpretação das instruções e inconsistências no formato esperado. Esses testes orientaram ajustes nas regras e na estrutura dos prompts, sem otimização baseada nas métricas finais de avaliação. Não foi realizado estudo sistemático de sensibilidade às diferentes formulações. As instruções foram redigidas em português, idioma dos documentos institucionais processados, uma vez que, em tarefas extrativas, manter o idioma das instruções consistente com o do texto de entrada resulta em desempenho superior ao do uso de prompts em inglês [16]. O processo de extração é decomposto em duas chamadas sequenciais ao LLM, cada uma com um prompt dedicado.

Listagem 1: Prompt base – Extração de entidades (Zero-Shot).

Você é um sistema avançado de extração de informações, especializado na identificação de entidades para a construção de grafos de conhecimento a partir de documentos institucionais.

Sua extração deve ser consistente, restrita ao esquema fornecido e reproduzível. Precisão e completude são igualmente importantes. Não invente informações sem suporte no texto e não produza explicações ou comentários.

ENTRADAS

ONTOLOGY_TTL:
Ontologia de domínio em Turtle (TTL), contendo classes, propriedades, hierarquias, domínios, alcances e demais elementos do esquema.

INPUT_TEXT:
Texto integral ou trecho de documento institucional.

A ontologia constitui a referência para os tipos de entidades permitidos. Utilize somente as classes (`owl:Class`) declaradas na ontologia para classificar as entidades identificadas.

EXTRACTION_RULES
...

FORMATO DE SAÍDA

Retorne apenas um objeto JSON válido, sem Markdown, comentários ou texto adicional, com a seguinte estrutura:

```
{
  "entidades": [
    { "id": "...", "tipo": "...", "propriedades": {} }
  ]
}
```

O *prompt* de relações compartilha a mesma organização geral, composta pelo papel atribuído ao modelo, pelas entradas, pelas regras de extração e pelo formato de resposta. Contudo, ele possui objetivos e restrições específicos: recebe, adicionalmente, as entidades da etapa anterior, limita a fonte e o alvo a esse conjunto e solicita uma estrutura JSON própria para a representação das relações.

3.2.3 Estratégias de Prompt. Foram desenvolvidas duas estratégias aplicadas simetricamente a ambas as etapas:

Zero-Shot (ZS): O *prompt* fornece contexto, ontologia completa e instruções de extração, sem exemplos. O modelo infere o comportamento esperado exclusivamente a partir do esquema ontológico.

Few-Shot (FS): O *prompt* é enriquecido com dois exemplos de pares entrada-saída em formato JSON, viabilizando aprendizado em contexto (*in-context learning*) e reduzindo ambiguidades na interpretação do esquema. O exemplo não foi extraído das atas avaliadas, trata-se apenas de uma representação genérica incluída para demonstrar o formato esperado de entrada e saída. Os prompts completos estão disponíveis no repositório público do projeto³.

3.3 Módulo de Inferência Semântica

O módulo de inferência opera em duas chamadas sequenciais ao LLM, ambas orientadas pela ontologia de domínio fornecida integralmente no *prompt*. Na primeira, o modelo

³<https://github.com/GiovannaXavier18/LLM-Text2KG.git>

identifica e classifica entidades segundo as classes `owl:Class` da ontologia, realizando correferência quando não há ambiguidade. Na segunda, extrai relações entre as entidades identificadas, restringindo-as às propriedades `owl:ObjectProperty` compatíveis com `rdfs:domain` e `rdfs:range`.

Em todos os *prompts* são indicadas regras de extração, que foram refinadas durante a elaboração das instruções para reduzir ambiguidades e manter as saídas consistentes com o esquema ontológico. As principais restrições, aplicadas às etapas de entidades e relações, são sintetizadas no Listagem 2.

Listagem 2: Principais regras de extração do LLM-Text2KG.

REGRAS DE EXTRAÇÃO ENTIDADES 1. Extraia um nó para cada entidade distinta explicitamente mencionada no texto. 2. Realize resolução de correferência quando claramente suportada pelo contexto: - Use a primeira ocorrência como id. - Resolva pronomes e referências contextuais quando não houver ambiguidade. 3. Os IDs devem ser legíveis para humanos (nunca identificadores numéricos). 4. Evite nós duplicados. 5. Se a ambiguidade for alta, prefira omitir. # REGRAS DE EXTRAÇÃO RELAÇÃO 1. Utilize apenas as entidades fornecidas em EXTRACTED_ENTITIES. 2. Não crie novas entidades. 3. Preserve exatamente os IDs e os tipos das entidades fornecidas. 4. Use apenas tipos de relacionamentos definidos na ontologia. 5. Respeite a compatibilidade entre fonte, relacionamento e alvo definida pelos domínios e imagens da ontologia. 6. Extraia relacionamentos explicitamente declarados ou claramente indicados pela estrutura linguística direta do texto. 7. Não invente relacionamentos especulativos ou sem suporte. 8. Evite relacionamentos duplicados. 9. Se a ambiguidade for alta, prefira omitir. 10. Mantenha a ordem da primeira ocorrência dos relacionamentos no texto.

Por exemplo, se uma pessoa já identificada pelo nome completo é posteriormente mencionada como “o presidente”, a resolução de correferência evita a criação de um segundo nó para a mesma entidade; sem essa regra, o grafo poderia conter entidades duplicadas representando o mesmo participante. De forma semelhante, uma relação semanticamente plausível, mas não sustentada pelo texto ou incompatível com as restrições de domínio e alcance da ontologia, é omitida, evitando a inclusão de triplas especulativas ou estruturalmente inválidas.

As chamadas foram configuradas com *temperature* = 0,8 e *top_p* = 0,9, valores definidos em testes preliminares para evitar respostas excessivamente conservadoras e favorecer a recuperação de entidades e relações semanticamente sustentadas pelo contexto. A configuração foi mantida constante entre todos os modelos e estratégias de *prompting*, garantindo condições uniformes de comparação. Entretanto, valores mais altos de *temperature* aumentam a aleatoriedade da geração e podem reduzir a reprodutibilidade entre execuções. Como não foram realizadas repetições independentes para estimar essa variabilidade, os resultados devem ser interpretados considerando essa limitação.

3.4 Construção do Grafo de Conhecimento

Após a inferência e o pós-processamento, as entidades e relações extraídas são convertidas para RDF. Cada entidade é instanciada como um recurso tipado segundo sua classe ontológica, enquanto as relações são representadas pelas propriedades definidas na ontologia, conectando os respectivos recursos de origem e destino. O grafo resultante é serializado em TTL, permitindo sua interoperabilidade com ferramentas de consultas SPARQL ⁴.

4 ESTUDO DE CASO COM GOVERNANÇA HÍDRICA

O LLM-Text2KG foi avaliado experimentalmente no domínio da governança hídrica, utilizando dez atas de reuniões ordinárias do Comitê da Bacia Hidrográfica do Acaraú, realizadas entre 2005 e 2015, com duas a três páginas cada. O experimento avaliou, em condições controladas, a capacidade de diferentes LLMs e de estratégias de *prompting* de identificar entidades e relações alinhadas à ontologia de domínio.

4.1 Construção da Ontologia de Domínio

A ontologia de domínio foi desenvolvida iterativamente em reuniões entre os autores e especialistas do domínio. Inicialmente, foram definidos os atores, temas, problemas e atividades essenciais presentes nas atas. Posteriormente, as classes e relações foram refinadas para representar informações associadas não apenas às pessoas, mas também à reunião, ao comitê e ao contexto institucional [19]. O objetivo foi produzir uma ontologia enxuta, em vez de modelar exaustivamente o domínio. O modelo resultante foi formalizado no Protégé⁵, com os conceitos representados como classes e suas associações como propriedades ontológicas. A Figura 2 apresenta a ontologia utilizada.

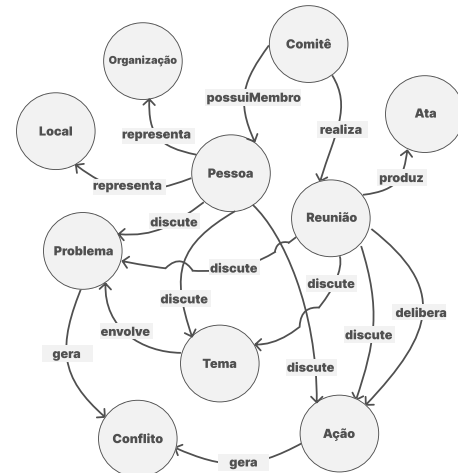


Figura 2: Modelagem da ontologia utilizada no estudo de caso.

⁴<https://www.w3.org/TR/sparql11-query/>

⁵<https://protege.stanford.edu/>

4.2 Seleção dos Modelos

Foram avaliados seis LLMs de pesos abertos (*open-weight*), selecionados segundo três critérios: (i) disponibilidade gratuita durante o período experimental, por meio da API da NVIDIA ou da plataforma Ollama; (ii) diversidade de famílias de modelos; e (iii) variação de porte, considerando diferentes números de parâmetros. Esses critérios buscaram ampliar a diversidade da comparação mantendo sua viabilidade computacional. Os modelos selecionados são apresentados na Tabela 1.

Tabela 1: Modelos avaliados e número de parâmetros.

Modelo	Parâmetros (B)
MiniMax-M2.5	230
Gemma 4 31B	31
GPT-OSS 120B	117
GPT-OSS 20B	21
GLM-5.1	754
DeepSeek V4 Flash	284

4.3 Avaliação Quantitativa

O *gold standard* foi construído a partir da anotação manual de dez atas, selecionadas de modo a contemplar diferentes anos entre 2005 e 2015. Um primeiro anotador identificou e tipou entidades e relações de acordo com o esquema ontológico. Em seguida, as anotações foram revisadas por um segundo anotador com experiência em grafos de conhecimento. As divergências foram discutidas e resolvidas por consenso, resultando na versão final do *gold standard*.

As extrações de entidades e relações foram comparadas separadamente ao *gold standard* por correspondência exata após normalização textual, com conversão para minúsculas, remoção de acentos e cedilhas e padronização dos espaços. Entidades foram avaliadas pelo par (identificador, tipo ontológico) e relações pela tripla (origem, tipo relacional, destino), sendo um elemento considerado correto somente quando todos os seus componentes coincidiam.

Foram utilizadas as métricas:

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

$$F1-score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Em que *TP* corresponde aos elementos extraídos que coincidem integralmente com o *gold standard*; *FP* aos elementos extraídos ausentes no *gold standard*; e *FN* aos elementos esperados não recuperados pelo modelo.

A variabilidade das métricas foi estimada por *bootstrap* [5], com 10.000 reamostragens com reposição das atas disponíveis para cada combinação de modelo e estratégia. Em cada iteração, recalculou-se a média das métricas, e os percentis 2,5% e 97,5% definiram o intervalo de confiança de 95%. Os resultados são reportados como média [IC 95%].

Além da correspondência com o padrão-ouro, avaliou-se a conformidade ontológica dos grafos finais. Foram verificadas: (i) a utilização exclusiva de classes declaradas na ontologia;

(ii) a utilização de propriedades previstas no esquema; (iii) a integridade referencial das relações; e (iv) a compatibilidade entre os tipos das entidades de origem e destino e as restrições de cada propriedade. A conformidade global corresponde à razão entre o número de relações integralmente conformes e o total de relações produzidas. Adicionalmente, utilizou-se SHACL (*Shapes Constraint Language*) para validar as restrições de domínio e alcance das propriedades ontológicas. Cada violação indica uma relação cuja entidade de origem ou destino apresenta tipo incompatível com o esperado pelo esquema.

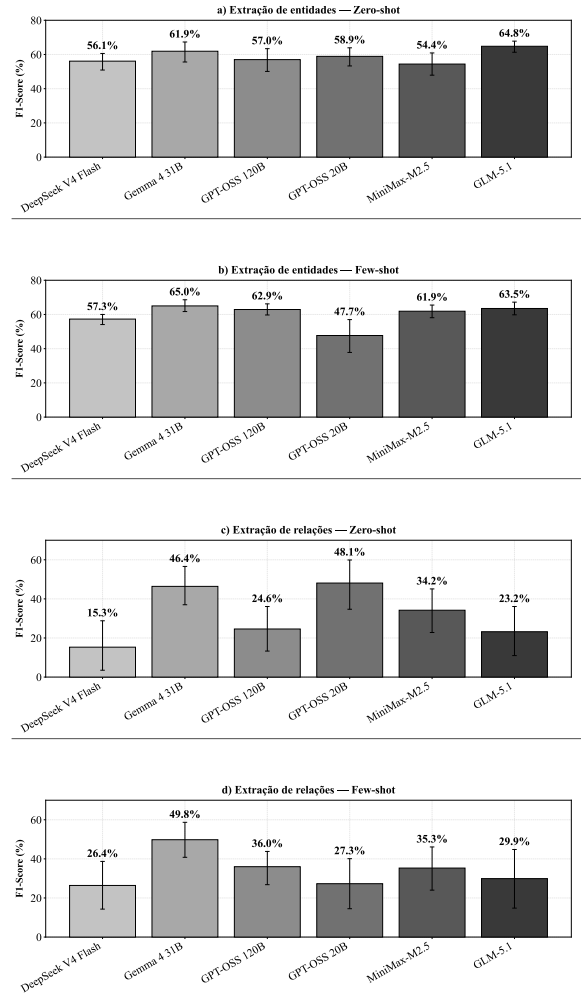


Figura 3: Comparação do desempenho dos modelos na extração de entidades e relações sob as estratégias zero-shot e few-shot. As barras representam o F1-score médio obtido nas atas avaliadas, as hastes indicam os intervalos de confiança de 95%.

4.4 Discussão dos Resultados Quantitativos

O DeepSeek V4 Flash apresentou resultados para dez atas em zero-shot e para nove em few-shot. Na ata 20, a requisição

de extração de relações em *few-shot* excedeu o tempo limite, o que impediu a obtenção de uma saída válida. Por esse motivo, a comparação entre as estratégias para esse modelo deve ser interpretada com cautela, pois as médias foram calculadas a partir de conjuntos de atas não idênticos. A análise comparativa principal contempla seis modelos. De modo geral, o efeito da inclusão de exemplos variou entre eles, resultando em aumento em algumas configurações e em redução em outras.

4.4.1 Extração de Entidades. O melhor resultado na extração de entidades foi obtido pelo **Gemma 4 31B em *few-shot*, com *F1-score* de 0,650**, seguido pelo GLM-5.1 em *zero-shot* (0,648) e *few-shot* (0,635) e pelo GPT-OSS 120B em *few-shot* (0,629). No GLM-5.1, a pequena diferença entre as estratégias e a sobreposição dos intervalos de confiança não indicam um efeito consistente dos exemplos. O impacto do *few-shot* variou entre os modelos, com maior aumento nominal no MiniMax-M2.5 (0,544 para 0,619), seguido por Gemma 4 31B (0,619 para 0,650) e GPT-OSS 120B (0,570 para 0,629). No DeepSeek V4 Flash, a variação foi discreta (0,561 para 0,573) e deve ser interpretada com cautela.

Em sentido contrário, o GPT-OSS 20B apresentou a maior redução de desempenho, com *F1-score* de 0,589 em *zero-shot* e 0,477 em *few-shot*. A queda em *Precision* (0,778 para 0,600) e *Recall* (0,483 para 0,397) indica menor cobertura das entidades esperadas e maior proporção de extrações incompatíveis com o padrão-ouro. Em contraste com o GPT-OSS 120B, esse resultado mostra que o efeito dos exemplos não é uniforme nem depende apenas do porte do modelo. Nas condições avaliadas, os exemplos em contexto beneficiaram a maioria dos modelos, mas prejudicaram outros, sugerindo que sua eficácia depende de como cada modelo interpreta e generaliza os exemplos fornecidos.

4.4.2 Extração de Relações. A extração de relações apresentou desempenho inferior ao da extração de entidades em todas as configurações. Isso reflete o caráter cumulativo da tarefa, que exige identificar corretamente as entidades, preservar seus identificadores, selecionar o predicado ontológico adequado e conectar corretamente os pares de entidades. Assim, erros ou omissões na etapa de entidades propagam-se para a avaliação relacional, mesmo quando a associação semântica é apenas parcialmente reconhecida.

O maior *F1-score* relacional foi obtido pelo **Gemma 4 31B em *few-shot* (0,498)**, seguido pelo GPT-OSS 20B em *zero-shot* (0,481) e pelo Gemma 4 31B em *zero-shot* (0,464). A melhor configuração apresentou *Precision* de 0,620 e *Recall* de 0,427, indicando maior correção das relações extraídas do que cobertura das relações esperadas. Assim, uma parte significativa do erro decorreu da omissão de conexões presentes no padrão-ouro.

A inclusão de exemplos aumentou, nominalmente, o *F1-score* relacional em cinco dos seis modelos, embora a sobreposição dos intervalos de confiança não permita afirmar uma vantagem consistente da estratégia *few-shot*. O GPT-OSS 20B apresentou comportamento oposto, com redução do *F1-score* relacional de 0,481 para 0,273 e do *F1-score*

de entidades de 0,589 para 0,477. Como a etapa relacional depende das entidades previamente extraídas, erros ou omissões nessa etapa comprometem a recuperação das triplas. Em contraste com o GPT-OSS 120B, esse resultado indica que o efeito dos exemplos varia mesmo entre modelos da mesma família e não depende apenas do número de parâmetros. Em conjunto, os resultados mostram que exemplos em contexto podem favorecer a extração, mas não eliminam o principal desafio da tarefa: estabelecer relações semânticas completas e consistentes entre as entidades.

4.4.3 Conformidade Ontológica dos Grafos. A conformidade ontológica complementa a avaliação baseada no padrão-ouro ao verificar se as relações utilizam predicados válidos e conectam entidades com tipos compatíveis ao esquema. A Tabela 2 apresenta a proporção global de relações conformes nos grafos finais e o número de violações identificadas pelo SHACL.

Tabela 2: Conformidade ontológica dos grafos finais. Conf. representa a proporção global de relações integralmente conformes, e Viol. corresponde ao número de violações SHACL.

Modelo	Zero-shot			Few-shot		
	<i>n</i>	Conf. (%)	Viol.	<i>n</i>	Conf. (%)	Viol.
DeepSeek V4 Flash	10	97,44	11	9	85,93	54
Gemma 4 31B	10	99,07	6	10	99,37	4
GPT-OSS 120B	10	93,98	18	10	100,00	0
GPT-OSS 20B	10	98,54	7	10	93,98	18
MiniMax-M2.5	10	94,51	29	10	87,99	62
GLM-5.1	10	89,15	53	10	66,67	190

A aderência estrutural foi elevada na maior parte das configurações: 11 das 12 combinações apresentaram conformidade superior a 85%. Entretanto, o efeito do *few-shot* não foi uniforme. Houve aumento no GPT-OSS 120B, de 93,98% para 100,00%, e pequena variação positiva no Gemma 4 31B, de 99,07% para 99,37%. Nos demais modelos, a conformidade diminuiu, com a maior redução no GLM-5.1, de 89,15% para 66,67%, acompanhada pelo aumento de 53 para 190 violações. A comparação do DeepSeek V4 Flash deve ser interpretada com cautela, a ausência da ata 20 pode ter influenciado a média agregada.

O contraste entre o GPT-OSS 120B e o Gemma 4 31B mostra que a conformidade estrutural e a fidelidade semântica são dimensões distintas. Embora o GPT-OSS 120B em *few-shot* tenha alcançado 100,00% de conformidade, seu *F1-score* relacional permaneceu inferior ao do Gemma 4 31B. Uma relação pode respeitar o predicado e os tipos permitidos pela ontologia, mas ainda conectar entidades incorretas, representar informação não sustentada pelo documento ou omitir associações presentes no padrão-ouro. O **Gemma 4 31B em *few-shot***, por sua vez, alcançou os maiores *F1-score* de entidades e relações, com 99,37% de conformidade, justificando sua seleção para a avaliação na etapa da Seção 4.5.

Os resultados fornecem evidências de que a disponibilização da ontologia no *prompt* contribuiu para orientar a geração

de grafos compatíveis com o esquema. Contudo, como a conformidade foi calculada após as etapas determinísticas de desserialização, deduplicação e filtragem, os valores representam o comportamento do *pipeline* completo, e não apenas a aderência bruta dos LLMs às instruções.

4.4.4 Análise de Ablação da Representação Ontológica. Para investigar o efeito da forma de apresentação do esquema, realizou-se uma ablação com o Gemma 4 31B, mantendo constantes a API, os documentos, as instruções, os exemplos, o formato de saída, os parâmetros de geração e o protocolo de avaliação. Na configuração original, a ontologia foi fornecida em Turtle (TTL); na condição de ablação, suas classes e combinações relacionais permitidas foram apresentadas em uma lista textual estruturada. A Tabela 3 compara o desempenho do Gemma 4 31B quando o esquema ontológico é apresentado na sua serialização original em TTL ou como uma lista textual estruturada.

Tabela 3: Resultados da ablação da representação do esquema ontológico no Gemma 4 31B.

Esquema	Prompt	n	F1 Ent.	F1 Rel.
TTL	ZS	10	0,619	0,464
Lista	ZS	10	0,521	0,102
TTL	FS	10	0,650	0,498
Lista	FS	10	0,553	0,265

A representação em TTL apresentou o maior *F1-score* em ambas as estratégias, com diferenças mais expressivas na extração de relações. Em *zero-shot*, o *F1-score* relacional caiu de 0,464 para 0,102 e, em *few-shot*, de 0,498 para 0,265. Para as entidades, as reduções foram menores: de 0,619 para 0,521 e de 0,650 para 0,553, respectivamente. Embora os exemplos tenham atenuado a perda da representação textual, seu desempenho permaneceu inferior ao da serialização em TTL.

A inspeção das saídas mostrou que a condição em lista foi acompanhada pela produção de predicados semanticamente plausíveis, mas ausentes da ontologia, como *vinculado_a*, *ocupou_cargo_em* e *apresentou_projeto_de*. Em sete das dez atas avaliadas em *zero-shot*, nenhuma relação extraída coincidiu integralmente com o padrão-ouro. Nas condições avaliadas, esses resultados sugerem que a serialização em TTL forneceu um sinal estrutural mais eficaz para restringir a seleção de predicados e de combinações permitidas entre origem e destino. O resultado não implica que o modelo execute raciocínio ontológico formal, mas evidencia que a forma de apresentação do esquema influencia a aderência das relações produzidas.

4.5 Avaliação por Questões de Competência

Como avaliação complementar à correspondência exata com o *gold standard*, empregaram-se questões de competência (CQs) [1, 6] para analisar a utilidade semântica dos grafos. As CQs foram elaboradas por especialistas após a definição da ontologia, com base no conteúdo das atas e sem acesso às saídas dos modelos, e posteriormente traduzidas em consultas

SPARQL orientadas pelo esquema ontológico e executadas no GraphDB⁶. Essa estratégia permite verificar se as triplas extraídas preservam informação suficiente para responder a questões relevantes ao domínio, reduzindo também o risco de viés circular [6, 17].

Essa avaliação complementa as métricas de correspondência exata, que podem penalizar representações semanticamente equivalentes com diferentes níveis de detalhamento. Por exemplo, na ata 2, o Gemma 4 31B em *few-shot* produziu a tripla *(Segunda Reunião Ordinária do Comitê de Bacia Hidrográfica do Acaraú, delibera, curso de extensão para os membros do CBH Acaraú)*, enquanto o *gold standard* registrou *(2ª Reunião Ordinária do Comitê da Bacia do Acaraú, delibera, Solicitar à Universidade Católica de Fortaleza a realização de um curso de extensão para os membros do CBH Acaraú)*. Embora apresentem diferentes níveis de especificidade, ambas preservam o mesmo núcleo semântico da deliberação.

A avaliação foi conduzida nos dez grafos produzidos pelo Gemma 4 31B em *few-shot*, configuração que apresentou os maiores *F1-score* para entidades e relações e conformidade ontológica de 99,37%. As respostas às consultas foram comparadas às respostas esperadas do *gold standard*. As dez questões de competência são apresentadas a seguir:

- (1) Qual reunião é descrita na ata analisada?
- (2) Quais pessoas são membros do Comitê da Bacia Hidrográfica do Acaraú na ata analisada e quais organizações ou locais elas representam?
- (3) Quais temas foram discutidos na reunião registrada na ata?
- (4) Quais problemas foram discutidos na reunião registrada na ata?
- (5) Quais ações foram deliberadas na reunião registrada na ata?
- (6) Quais temas, problemas ou ações foram discutidos por pessoas identificadas na ata?
- (7) Quais problemas estão associados aos temas identificados na ata?
- (8) Quais conflitos foram gerados por problemas ou ações identificados na ata?
- (9) Qual pessoa discutiu a maior quantidade de temas na ata analisada?
- (10) Quais corpos hídricos, como rios e açudes, foram identificados na ata analisada?

A Tabela 4 apresenta o número de respostas corretas recuperadas e o número esperado para cada questão de competência, considerando conjuntamente os dez grafos avaliados. No total, foram recuperadas 283 das 690 respostas esperadas, correspondendo a uma cobertura global de 41,0%.

Os melhores resultados ocorreram em consultas apoiadas por informações explícitas ou por relações locais: a CQ1 alcançou 100,0%, enquanto a CQ2 e a CQ10 obtiveram 68,9% e 72,1%, respectivamente. Os grafos foram, portanto, mais eficazes para recuperar a reunião descrita, os participantes, as organizações ou os locais representados e os corpos hídricos

⁶<https://graphdb.ontotext.com/>

Tabela 4: Resultados das CQs nos dez grafos gerados pelo Gemma 4 31B (*few-shot*). A cobertura corresponde à proporção de respostas esperadas que foram recuperadas.

CQ	Corretas	Esperadas	Cobertura (%)
CQ1	10	10	100,0
CQ2	157	228	68,9
CQ3	7	33	21,2
CQ4	4	8	50,0
CQ5	9	75	12,0
CQ6	60	213	28,2
CQ7	2	34	5,9
CQ8	0	35	0,0
CQ9	3	11	27,3
CQ10	31	43	72,1
Total	283	690	41,0

mencionados. Em contraste, as consultas que dependem de conexões semânticas entre conceitos apresentaram menor cobertura. A CQ5, relacionada às ações deliberadas, alcançou 12,0%; a CQ7, baseada na associação entre temas e problemas, 5,9%; e a CQ8, que exige relações entre problemas, ações e conflitos, não recuperou respostas.

Em diversos casos, as entidades **Tema**, **Problema**, **Acao** e **Conflito** estavam presentes, mas não conectadas por relações como **envolve**, **gera**, **discute** ou **delibera**. O contraste entre a conformidade ontológica de 99,37% e a cobertura de 41,0% evidencia que validade estrutural e utilidade semântica são dimensões distintas. Os grafos geralmente utilizaram predicados e tipos compatíveis com a ontologia, mas permaneceram incompletos quanto às conexões necessárias para responder às consultas. Esse resultado corrobora o observado no *F1-score* relacional: o principal gargalo está menos na produção de triplas formalmente admissíveis e mais na recuperação completa das relações semanticamente relevantes.

4.6 Comparação Exploratória com *Baselines*

Para fornecer uma referência externa, spaCy [7] e GraphRAG [11] foram avaliados sobre as mesmas dez atas e comparados ao Gemma 4 31B em *few-shot*. Como esses métodos operam com categorias genéricas, as classes ontológicas foram projetadas para tipos mais amplos: **Organizacao** e **Comite** para organização; **Local** para localização; e **Ata** e **Reuniao** para evento. Essa projeção não contempla classes específicas como **Tema**, **Problema**, **Acao** e **Conflito**; portanto, a comparação é exploratória e não estritamente equivalente.

Tabela 5: Comparação exploratória da extração de entidades.

Método	Precision	Recall	F1-score
spaCy	0,281 [0,242–0,319]	0,348 [0,315–0,383]	0,309 [0,275–0,345]
GraphRAG	0,561 [0,479–0,639]	0,321 [0,212–0,430]	0,395 [0,285–0,505]
Gemma 4 31B	0,723 [0,674–0,775]	0,591 [0,565–0,618]	0,650 [0,617–0,685]

O Gemma 4 31B apresentou maior correspondência às entidades do padrão-ouro. Parte dessa diferença, contudo, decorre da orientação ontológica, pois os *baselines* não receberam as classes específicas do domínio e concentraram as extrações em pessoas, organizações, locais e eventos. As relações não foram comparadas quantitativamente, pois spaCy produz dependências sintáticas e GraphRAG gera associações textuais sem predicados ontológicos diretamente comparáveis. Assim, os resultados constituem uma contextualização funcional, e não evidência isolada de superioridade do *pipeline*.

5 LIMITAÇÕES

O *gold standard* foi elaborado por dois anotadores, com resolução das divergências por consenso, embora relações dependentes de contexto permaneçam sujeitas a diferenças interpretativas. O corpus contém apenas dez documentos de um único comitê, limitando a generalização para outras instituições, domínios e ontologias. Além disso, cada ata foi processada como um grafo independente, sem resolução de entidades recorrentes ou integração temporal. O uso de APIs externas também limita a reprodutibilidade, pois alterações nos modelos ou na infraestrutura dos provedores podem afetar os resultados. Uma requisição do DeepSeek V4 Flash em *few-shot* não foi concluída, restringindo essa configuração a nove atas. As questões de competência foram avaliadas apenas para o Gemma 4 31B em *few-shot*, impedindo comparações entre configurações. Por fim, a comparação com spaCy e GraphRAG foi exploratória e não estritamente equivalente, devido à projeção das classes ontológicas em categorias genéricas e à ausência de relações diretamente comparáveis.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou e avaliou experimentalmente o LLM-Text2KG, um *pipeline* orientado por ontologias para a construção de grafos de conhecimento a partir de documentos institucionais. Foram comparados seis LLMs em estratégias *zero-shot* e *few-shot*. Nos experimentos com dez atas, o Gemma 4 31B em *few-shot* obteve o melhor desempenho, com *F1-score* de 0,650 para entidades e 0,498 para relações. Os resultados indicaram que o efeito dos exemplos varia entre os modelos e que a extração de relações permanece o principal desafio. A elevada conformidade ontológica não implicou maior fidelidade semântica: no Gemma 4 31B em *few-shot*, 99,37% de conformidade foram acompanhados de 41,0% de cobertura nas questões de competência. A ablação mostrou ainda que a representação da ontologia em TTL superou a lista textual estruturada em ambas as estratégias, sobretudo na extração de relações.

Como trabalhos futuros, pretende-se ampliar o *benchmark* com novos documentos, domínios, ontologias e modelos, comparar o *pipeline* com métodos baseados no mesmo esquema ontológico, aprimorar a extração de relações, avaliar questões de competência em diferentes configurações e integrar entidades recorrentes para construir um grafo institucional unificado com informações temporais e de proveniência.

CUIDADOS ÉTICOS

As atas utilizadas neste estudo foram disponibilizadas com autorização para uso em pesquisas acadêmicas e já haviam sido empregadas em investigações da área ambiental. Os documentos incluem os nomes dos participantes das reuniões, sem dados pessoais sensíveis adicionais, e foram utilizados exclusivamente para fins científicos.

ACKNOWLEDGMENTS

Este trabalho foi apoiado pelo Instituto Kunumi. Os autores agradecem à instituição pelo apoio financeiro e pelo compromisso com o avanço da pesquisa científica. O último autor agradece também à Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo apoio concedido ao desenvolvimento desta pesquisa.

Não antecipamos quaisquer preocupações sociais ou éticas imediatas decorrentes deste trabalho. Além disso, reconhecemos o uso de LLMs para auxiliar na verificação gramatical de partes do manuscrito.

REFERENCES

- [1] Camila Bezerra, Fred Freitas, and Filipe Santana. 2013. Evaluating Ontologies with Competency Questions. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Vol. 3. 284–285. <https://doi.org/10.1109/WI-IAT.2013.199>
- [2] Ricardo Campos, José Pedro Evans, José Miguel Isidro, Miguel Marques, Luís Filipe Cunha, Alípio Jorge, Sérgio Nunes, and Nuno Guimarães. 2026. NLP for Local Governance Meeting Records: A Focus Article on Tasks, Datasets, Metrics and Benchmark. arXiv:2602.08162 [cs.CL] <https://arxiv.org/abs/2602.08162>
- [3] Keyin Chen, Rui Yao, Zhichuan Tang, Anqi Yao, and Xiaoneng Jin. 2025. Bridging cultural gaps in product design: A cross-cultural supervisor approach using large language model. *Design and Artificial Intelligence* 1, 1 (2025), 100004. <https://doi.org/10.1016/j.daai.2025.100004>
- [4] Eric Mateus Dias, Zoraide Pessoa, and Rylanneive Teixeira. 2023. ADAPTIVE GOVERNANCE AND WATER SECURITY IN THE CONTEXT OF CLIMATE CHANGE IN THE SEMI-ARID. *Mercurator* 21 (2023). <https://doi.org/10.4215/rm2022.e21025>
- [5] Bradley Efron and Robert J. Tibshirani. 1994. *An Introduction to the Bootstrap* (1 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780429246593>
- [6] Michael Grüninger and Mark S. Fox. 1995. *The Role of Competency Questions in Enterprise Engineering*. Springer US, Boston, MA, 22–31. https://doi.org/10.1007/978-0-387-34847-6_3
- [7] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). <https://doi.org/10.5281/zenodo.1212303>
- [8] Haoyu Huang, Chong Chen, Zeang Sheng, Yang Li, and Wentao Zhang. 2025. Can LLMs be Good Graph Judge for Knowledge Graph Construction?. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng (Eds.). Association for Computational Linguistics, Suzhou, China, 10929–10948. <https://doi.org/10.18653/v1/2025.emnlp-main.554>
- [9] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems* 33, 2 (2022), 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>
- [10] Samira Khorshidi, Azadeh Nikfarjam, Suprita Shankar, Yisi Sang, Yash Govind, Hyun Jang, Ali Kasgari, Alexis McClimans, Mohamed Soliman, Vishnu Konda, Ahmed Fakhry, and Xiaoguang Qi. 2025. ODKE+: Ontology-Guided Open-Domain Knowledge Extraction with LLMs. arXiv:2509.04696 [cs.CL] <https://arxiv.org/abs/2509.04696>
- [11] Jonathan Larson and Steven Truitt. 2024. GraphRAG: Unlocking LLM Discovery on Narrative Private Data. Microsoft Research Blog. <https://www.microsoft.com/en-us/research/blog/graphrag-unlocking-llm-discovery-on-narrative-private-data/> Acesso em: jun. 2026.
- [12] Yancheng Ling, Zhenlin Qin, and Zhenliang Ma. 2026. A review of knowledge graph construction using large language models in transportation: Problems, methods, and challenges. *Transportation Research Part C: Emerging Technologies* 183, Article 105428 (Feb. 2026), 105428 pages. <https://doi.org/10.1016/j.trc.2025.105428>
- [13] Bingchen Liu, Yuanyuan Fang, Naixing Xu, Shihao Hou, Xin Li, and Qian Li. 2025. Large Language Models for Knowledge Graph Embedding: A Survey. *Mathematics* 13, 14 (2025). <https://doi.org/10.3390/math13142244>
- [14] Nandana Mihindukulasooriya, Sanju Tiwari, Carlos F. Enguix, and Kusum Lata. 2023. Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text. In *The Semantic Web – ISWC 2023*, Terry R. Payne, Valentina Presutti, Guilin Qi, María Poveda-Villalón, Giorgos Stoilos, Laura Hollink, Zoi Kaoudi, Gong Cheng, and Juanzi Li (Eds.). Springer Nature Switzerland, Cham, 247–265.
- [15] Belinda Mo, Kyssen Yu, Joshua Kazdan, Proud Mpala, Lisa Yu, Charilaos Kanatsoulis, and Sammi Koyejo. 2025. KGen: Extracting Knowledge Graphs from Plain Text with Language Models. In *Advances in Neural Information Processing Systems*, D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (Eds.), Vol. 38. Curran Associates, Inc., 30092–30115. https://proceedings.neurips.cc/paper_files/paper/2025/file/2b368455e832d2b1a60bcad8c4c6481f-Paper-Conference.pdf
- [16] Itai Mondshine, Tzuf Paz-Argaman, and Reut Tsarfaty. 2025. Beyond English: The Impact of Prompt Translation Strategies across Languages and Tasks in Multilingual LLMs. In *Findings of the Association for Computational Linguistics: NAACL 2025*, Luis Chiruzzo, Alan Ritter, and Lu Wang (Eds.). Association for Computational Linguistics, Albuquerque, New Mexico, 1331–1354. <https://doi.org/10.18653/v1/2025.findings-naacl.73>
- [17] Glaice Kelly Q. Monfardini, Jordana S. Salamon, and Monalissa P. Barcellos. 2023. Use of onbsp;Competency Questions innbsp;Ontology Engineering: A Survey. In *Conceptual Modeling: 42nd International Conference, ER 2023, Lisbon, Portugal, November 6–9, 2023, Proceedings* (Lisbon, Portugal). Springer-Verlag, Berlin, Heidelberg, 45–64. https://doi.org/10.1007/978-3-031-47262-6_3
- [18] Daniel Sanford Moreira. 2014. O Comitê de bacia como instrumento de democratização dos recursos hídricos. Um estudo de caso: a atuação do Comitê da Bacia do Acaraú – Ceará. , 91 pages. <http://repositorio.ufc.br/handle/riufc/11594>
- [19] Natalya F. Noy and Deborah L. McGuinness. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. Technical Report KSL-01-05 and SMI-2001-0880. Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, Stanford, CA. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- [20] Vayianos Pertsas and Panos Constantopoulos. 2023. Ontology-Driven Extraction of Contextualized Information from Research Publications.. In *KEOD*. 108–118.
- [21] Julien Plu, Oscar Moreno Escobar, Edouard Trouilleux, Axelle Gapin, Pasquale Lisena, Thibault Ehrhart, and Raphaël Troncy. 2025. Text2KGBench-LettrIA: A refined benchmark for Text2Graph systems. In *KBC-LM and LM-KBC Challenge (Knowledge Base Construction from Pre-trained Language Models) at ISWC 2025, 24th International Semantic Web Conference, 2-6 November 2025, Nara, Japan*, CEUR (Ed.). Nara. <https://ceur-ws.org/Vol-4041/paper3.pdf> CEUR.
- [22] André Regino and Julio Reis. 2025. Enhancing Knowledge Graphs with Large Language Models: Contributions to E-commerce Question Answering Systems. In *Anais Estendidos do XXXI Simpósio Brasileiro de Sistemas Multimídia e Web* (Rio de Janeiro/RJ). SBC, Porto Alegre, RS, Brasil, 21–22. https://doi.org/10.5753/webmedia_estendido.2025.16346
- [23] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. 2020. Knowledge Graphs: An Information Retrieval Perspective. *Found. Trends Inf. Retr.* 14, 4 (Oct. 2020), 289–444. <https://doi.org/10.1561/15000000063>
- [24] Marlene Yara Tenório Soares and Rodrigo Guimarães de Carvalho. 2022. Gestão dos Recursos Hídricos e Conflitos no Semiárido: Estudo do Caso do Aqude Lucrécia, Estado do Rio Grande do

- Norte. *Revista GeoInterações* 6, 1 (Dec. 2022), 57–76. <https://periodicos.apps.uern.br/index.php/RGI/article/view/4445>
- [25] Milena Trajanoska, Riste Stojanov, and Dimitar Trajanov. 2023. Enhancing Knowledge Graph Construction Using Large Language Models. arXiv:2305.04676 [cs.CL] <https://arxiv.org/abs/2305.04676>