

**UNIVERSIDADE DO VALE DO RIO DOS SINOS  
UNIDADE ACADÊMICA DE GRADUAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**ANDRÉ FELIPE DA SILVA HOFFMANN**

**PROPOSTA DE CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO  
EDUCACIONAL COM MODELOS DE LINGUAGEM ORIENTADOS PELA  
TAXONOMIA DE BLOOM**

São Leopoldo  
2025

André Felipe da Silva Hoffmann

**PROPOSTA DE CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO  
EDUCACIONAL COM MODELOS DE LINGUAGEM ORIENTADOS PELA  
TAXONOMIA DE BLOOM**

Artigo apresentado como requisito parcial para  
obtenção do título de Bacharel em Ciência da  
Computação, pelo Curso de Ciência da Compu-  
tação da Universidade do Vale do Rio dos Sinos  
(UNISINOS)

Orientador(a): Dra. Rosemary Francisco

São Leopoldo  
2025

# PROPOSTA DE CONSTRUÇÃO DE GRAFOS DE CONHECIMENTO EDUCACIONAL COM MODELOS DE LINGUAGEM ORIENTADOS PELA TAXONOMIA DE BLOOM

André Felipe da Silva Hoffmann<sup>1</sup>

Rosemary Francisco<sup>2</sup>

**Resumo:** Este trabalho propõe um framework para a construção automatizada de grafos de conhecimento educacional a partir de fontes não estruturadas como livros e artigos. O sistema utiliza modelos de linguagem (LLMs) e algoritmos de grafos para extrair, organizar e classificar unidades de conhecimento (UCs), com o objetivo de gerar um grafo detalhado, adaptativo e semanticamente estruturado. Para isso, a arquitetura inclui etapas de processamento de linguagem natural, como segmentação de texto, extração de entidades e detecção de comunidades, além da aplicação da Taxonomia de Bloom para classificar as UCs em diferentes níveis cognitivos, resultando em um grafo que representa as relações entre conceitos de forma clara e hierárquica. O framework também endereça o problema do cold start em sistemas de aprendizado adaptativo, realizando uma avaliação comparativa da dificuldade de cada UC ao aproveitar o conhecimento implícito dos LLMs para analisar a complexidade do conceito, o conhecimento prévio necessário e a clareza da linguagem. Com isso, o grafo gerado e a sua avaliação inicial estabelecem uma base para futuras extensões, como a geração automatizada de exercícios alinhados aos níveis da Taxonomia de Bloom e a recomendação de recursos complementares personalizados.

**Palavras-chave:** Grafos de Conhecimento. Aprendizado Adaptativo. Modelos de Linguagem. Taxonomia de Bloom. Unidades de Conhecimento. Educação Personalizada.

## 1 INTRODUÇÃO

Apesar do reconhecido potencial da educação personalizada, implementar sistemas de aprendizagem adaptativos na prática apresenta desafios significativos. Fechar a lacuna entre a pesquisa teórica e os aspectos práticos de criar sistemas escaláveis e adaptativos requer lidar com uma série de desafios complexos (PELÁNEK, 2024). Esses sistemas precisam gerenciar eficientemente grandes quantidades de dados, incorporar novos conhecimentos e atender às diversas necessidades de alunos e educadores.

Modelos de domínio, que representam a estrutura fundamental e as relações entre os conceitos dentro de uma disciplina, são elementos chave para a construção de sistemas de aprendizagem adaptativa eficientes. Um aspecto-chave da modelagem de domínio é a definição de unidades de conhecimento (UC). Essas unidades representam os conceitos, habilidades ou procedimentos fundamentais que os alunos precisam dominar em um determinado assunto. As unidades de conhecimento são os blocos essenciais para recursos de aprendizagem adaptativa, incluindo recomendações personalizadas, visualização do progresso e caminhos de aprendizagem personalizados (PELÁNEK, 2024). No entanto, a construção de modelos de domínio

---

<sup>1</sup>Graduando em Ciência da Computação pela Unisinos. Email: andresh@edu.unisinos.br

<sup>2</sup>Professora dos cursos de TI da UNISINOS. Email: rosemaryf@unisinos.br

eficazes apresenta desafios, como a necessidade de lidar com o problema de "cold start", que se refere à dificuldade em fornecer personalização no início da interação do aluno com o sistema devido à falta de dados sobre seu desempenho (VELDE et al., 2020).

Esta pesquisa propõe um sistema para gerar automaticamente modelos de domínio a partir de diversos recursos educacionais (por exemplo, livros didáticos e artigos), definindo unidades de conhecimento e suas inter-relações. O sistema utiliza modelos de linguagem de grande porte (LLMs) para criar uma estrutura inicial que pode ser progressivamente refinada com dados de interação dos alunos. Essa estratégia visa a oferecer suporte adaptativo desde o início da interação do aluno com a plataforma, mesmo quando dados de interação ainda são escassos.

Ao automatizar a etapa de modelagem de domínio, este trabalho busca oferecer uma contribuição prática para o campo, apresentando uma ferramenta projetada para auxiliar desenvolvedores e educadores na criação de sistemas de aprendizado adaptativo mais escaláveis e eficazes. O restante deste artigo está organizado da seguinte forma: A Seção 2 revisa a fundamentação teórica, abordando conceitos como modelos de domínio, a Taxonomia de Bloom e o papel dos LLMs. A Seção 3 analisa trabalhos relacionados, contextualizando a pesquisa. A Seção 4 detalha a arquitetura e a implementação do framework proposto. A Seção 5 apresenta os resultados obtidos em um estudo de caso e realiza uma análise qualitativa e quantitativa dos artefatos gerados. Finalmente, a Seção 6 conclui o trabalho, resumizando as contribuições, discutindo as limitações e apontando direções para pesquisas futuras.

## **2 FUNDAMENTAÇÃO TEÓRICA**

### **2.1 Dificuldade no Contexto de Aprendizado Adaptativo**

A definição e modelagem da dificuldade de unidades de conhecimento desempenha um papel de grande importância em sistemas de aprendizado adaptativo. Pelánek (2024) destaca a importância de definir adequadamente o nível de dificuldade dos componentes do conhecimento, pois isso afeta diretamente a personalização do aprendizado e a motivação do aluno. Imagine, por exemplo, um sistema de tutoria de matemática que apresenta a um aluno que ainda está aprendendo adição e subtração um problema envolvendo equações de segundo grau. Ao mesmo tempo, se o sistema apresentar a um aluno que já domina álgebra um problema como " $1 + 1 = ?$ ", ele pode se sentir entediado e desestimulado. Essa discrepância pode gerar frustração e desmotivação, levando o aluno a abandonar o sistema.

Em sistemas de aprendizado adaptativo, a dificuldade de cada unidade de conhecimento pode ser ajustada dinamicamente com base nos dados de interação dos alunos de forma contínua. Existem diferentes métodos para estimar a dificuldade, incluindo abordagens colaborativas, como a exemplificada em Velde et al. (2020), que utilizam dados de múltiplos alunos, e abordagens individualizadas, que concentram-se no desempenho individual de cada aluno. No entanto, mesmo as abordagens que utilizam dados de múltiplos alunos podem enfrentar o

problema de "cold start", já que ainda é necessário algum nível de interação para que o sistema colete dados suficientes e forneça sugestões pertinentes. Ao utilizar LLMs para estimar a dificuldade das unidades de conhecimento, podemos oferecer uma experiência personalizada desde o início da interação do aluno com o sistema, mesmo quando dados de interação ainda são escassos.

## 2.2 Modelo de Domínio e Grafo de Conhecimento

O modelo de domínio é um componente essencial em sistemas de aprendizado adaptativo, pois define a estrutura conceitual do conhecimento que deve ser aprendido. Ele descreve os conceitos principais de uma área de estudo, suas características e as relações entre eles, permitindo que o sistema de aprendizado compreenda a progressão lógica dos tópicos. Conforme discutido por Pelánek (2024), a construção de um modelo de domínio envolve uma série de desafios e trade-offs. O autor destaca que a criação de um modelo de domínio deve equilibrar precisão e generalidade, garantindo que os conceitos sejam suficientemente específicos para suportar o aprendizado eficaz, mas não tão detalhados que se tornem inflexíveis ou impraticáveis de aplicar em um ambiente dinâmico.

No contexto de aprendizado adaptativo, um modelo de domínio precisa também lidar com o problema da complexidade crescente, pois incluir muitos detalhes pode dificultar a aplicação prática dos conceitos. Pelánek (2024) argumenta que uma boa modelagem de domínio deve levar em consideração tanto a acessibilidade quanto a utilidade prática, sendo flexível o suficiente para se adaptar a diferentes perfis de alunos.

Por outro lado, um grafo de conhecimento pode ser entendido como uma implementação prática do modelo de domínio (PELÁNEK, 2024). Enquanto o modelo de domínio estabelece a visão teórica das relações conceituais, o grafo de conhecimento organiza essas relações de forma visual e computacional, permitindo uma representação intuitiva e estruturada. Em um grafo, cada nó representa um conceito e as arestas indicam as conexões, como pré-requisitos ou relações de dependência (CHEN et al., 2018). Dessa forma, o grafo de conhecimento não só materializa o modelo de domínio, mas também facilita o acesso e a navegação por essas informações, ajudando o sistema a compreender como guiar o aluno pelo conteúdo de forma eficaz.

Assim, enquanto o modelo de domínio oferece a base conceitual e estratégica sobre o que deve ser ensinado e como, o grafo de conhecimento fornece uma representação prática e navegável que pode ser utilizada pelo sistema de aprendizado para fornecer feedback e adaptar o conteúdo ao progresso do aluno.

## 2.3 Unidade de Conhecimento (UC)

A Unidade de Conhecimento é um componente central em modelos de domínio para aprendizado adaptativo. As UC representam blocos fundamentais do conhecimento que os alunos precisam dominar em um determinado campo. No contexto da teoria da aprendizagem e dos modelos educacionais, uma UC é definida como uma "unidade adquirida de função ou estrutura cognitiva", sendo inferida com base no desempenho do aluno em tarefas relacionadas (KOE-DINGER; CORBETT; PERFETTI, 2012; PELÁNEK, 2019).

Pelánek (2024) destaca que, na construção de modelos de domínio, a definição de UC é uma tarefa que envolve uma série de nuances e desafios. Em um sistema adaptativo, é crucial que as UCs sejam definidas de forma que representem unidades funcionais do conhecimento — suficientemente granulares para permitir uma personalização eficaz, mas também suficientemente amplas para evitar a fragmentação excessiva do conteúdo. Esse equilíbrio é um dos principais desafios mencionados pelo autor: a complexidade pode aumentar muito rapidamente se as UCs forem excessivamente detalhadas, dificultando a implementação prática e a navegação eficiente pelo conteúdo.

Além disso, Pelánek (2019) ressalta que as UCs não são apenas entidades isoladas; elas formam a base para determinar como o conhecimento pode ser representado e organizado em um modelo de domínio mais coeso. Por exemplo, cada UC pode ser conectada a outras UCs através de relações de pré-requisito, generalização-especialização ou até mesmo associações contextuais. No contexto do aprendizado adaptativo, isso significa que o sistema deve ser capaz de compreender não apenas quais UCs um aluno já dominou, mas também como essas UCs se relacionam entre si para facilitar a construção de um conhecimento progressivo e bem estruturado.

Portanto, ao definir UCs, o objetivo principal é garantir que o sistema possa utilizá-las efetivamente para guiar o aluno em um caminho de aprendizado coerente e adaptativo. Isso implica em considerar aspectos como (PELÁNEK, 2024):

- **Granularidade:** Cada UC deve ser granular o suficiente para capturar um aspecto relevante do conhecimento, mas não tão detalhada a ponto de complicar a estrutura do modelo de domínio.
- **Relacionamento entre UCs:** Os tipos de relação que existem entre as UCs, como relações de pré-requisito (o que precisa ser aprendido antes) ou relações de suporte (o que facilita o aprendizado de outra UC), desempenham um papel fundamental na forma como o sistema recomenda atividades ou conteúdos.

Assim, as Unidades de Conhecimento servem não só para agrupar itens de aprendizagem relacionados, como "adição de frações" ou "capitalização de nomes próprios", mas também como elementos fundamentais para garantir a eficiência do aprendizado adaptativo, permitindo que o sistema navegue entre diferentes níveis de complexidade do conhecimento.

## 2.4 Taxonomia de Bloom

A Taxonomia de Bloom, originalmente desenvolvida por Benjamin Bloom e colaboradores em 1956 (MARCHETI; BELHOT, 2010), é uma estrutura fundamental na educação, auxiliando na categorização de objetivos educacionais em diferentes níveis de complexidade cognitiva. Essa taxonomia organiza as habilidades em uma hierarquia que se estende do básico - como memorizar fatos - ao complexo - avaliar e criar novas ideias. Os níveis da Taxonomia de Bloom - Conhecimento, Compreensão, Aplicação, Análise, Síntese e Avaliação - são amplamente utilizados para definir a progressão de habilidades que um aluno deve desenvolver ao longo de sua jornada educacional.

Em 2001, a taxonomia passou por uma revisão, resultando em uma estrutura mais dinâmica e focada nos processos cognitivos. Nessa versão revisada, os substantivos originais foram transformados em verbos, como lembrar, entender, aplicar, analisar, avaliar e criar, para melhor representar as ações mentais envolvidas em cada nível. Essa mudança facilita a aplicação da Taxonomia de Bloom em contextos de aprendizado adaptativo, permitindo uma definição mais precisa das habilidades e uma melhor personalização do aprendizado (MARCHETI; BELHOT, 2010).

No contexto desta pesquisa, a Taxonomia de Bloom revisada desempenha um papel importante na estruturação das Unidades de Conhecimento que compõem o modelo de domínio. Ao classificar as UCs nos diferentes níveis da taxonomia, utilizando-a como uma heurística para guiar a expansão do grafo de conhecimento, pode-se criar um grafo que reflita a progressão natural das habilidades e personalize a experiência de aprendizado.

A utilização de LLMs para gerar UCs complementares para cada nível da Taxonomia de Bloom contribui para que o grafo de conhecimento seja abrangente e completo, mesmo quando a UC principal extraída do material didático se concentra em apenas um nível da taxonomia. O Quadro 1 apresenta exemplos de como o sistema pode gerar UCs complementares para a UC principal "Capital de Botswana", que se encaixa no nível "lembrar", expandindo para os outros níveis da Taxonomia de Bloom.

## 2.5 Modelos de Linguagem

Modelos de Linguagem (LLMs) são modelos de inteligência artificial que, após treinamento em conjuntos massivos de dados de texto, demonstram capacidade na geração de texto, tradução de idiomas, escrita criativa e resposta a perguntas (ZHAO et al., 2025). Essa versatilidade no processamento da linguagem natural oferece um vasto potencial para diversas aplicações, incluindo a área educacional. Estudos recentes (KOJIMA et al., 2023) demonstram que LLMs podem atuar como raciocinadores "zero-shot", operando com sucesso em tarefas complexas, como raciocínio aritmético e simbólico, sem necessidade de exemplos específicos. No presente trabalho, foram utilizados prompts detalhados para guiar os LLMs na geração de Unidades de

Quadro 1: Unidades de Conhecimento geradas utilizando "Capital de Botswana" como conceito base

Nível	Unidades de Conhecimento
Lembrar	<ul style="list-style-type: none"> <li>• Capital de Botswana</li> </ul>
Compreender	<ul style="list-style-type: none"> <li>• Compreender a importância da capital de Botswana no contexto do país.</li> <li>• Descrever as funções e características de uma capital.</li> </ul>
Aplicar	<ul style="list-style-type: none"> <li>• Ser capaz de localizar a capital de Botswana em um mapa.</li> <li>• Identificar a capital de Botswana em uma lista de cidades.</li> </ul>
Analisar	<ul style="list-style-type: none"> <li>• Comparar a capital de Botswana com outras capitais africanas, identificando semelhanças e diferenças.</li> <li>• Analisar os fatores que influenciaram a escolha da capital de Botswana.</li> </ul>
Avaliar	<ul style="list-style-type: none"> <li>• Justificar a localização da capital de Botswana com base em critérios geográficos, políticos e sociais.</li> <li>• Avaliar a importância da capital para o desenvolvimento de Botswana, considerando seus aspectos econômicos, sociais e culturais.</li> </ul>
Criar	<ul style="list-style-type: none"> <li>• Propor um projeto para melhorar a capital do país, considerando os desafios e oportunidades de desenvolvimento urbano.</li> <li>• Elaborar um mapa da capital de Botswana, incluindo seus principais pontos turísticos e áreas de interesse.</li> </ul>

Fonte: Elaborado pelo autor (2025).

Conhecimento (UCs) e na avaliação da dificuldade, buscando demonstrar a viabilidade dessa abordagem. É importante notar que, para garantir a geração de UCs completas e informativas, prompts mais extensos são necessários, o que pode resultar em um maior número de *tokens*. Isso não foi considerado um problema, visto que o foco principal da pesquisa reside na demonstração do potencial da aplicação de LLMs na construção de grafos de conhecimento educacional.

### 3 TRABALHOS RELACIONADOS

Esta seção tem como objetivo contextualizar a pesquisa e apresentar o contexto da pesquisa de construção de grafos de conhecimento educacionais. Para a seleção dos trabalhos aqui apresentados, foi utilizada uma abordagem exploratória, iniciando com a análise do artigo KnowEdu (CHEN et al., 2018), que serviu como principal inspiração para o desenvolvimento do framework proposto. A partir do KnowEdu, a pesquisa se expandiu para abranger trabalhos que abordassem os seguintes tópicos:

- Construção de grafos de conhecimento na área educacional;
- Uso da Taxonomia de Bloom na organização de informações educacionais;
- Modelos de linguagem para a extração e organização de informações;



- Estimativa de dificuldade de unidades de conhecimento;
- Mitigação do problema de "cold start" em sistemas adaptativos.

A seleção dos trabalhos se deu por meio de buscas em bases de dados acadêmicas, utilizando palavras-chave como "grafos de conhecimento", "educação", "Taxonomia de Bloom", "modelos de linguagem" e "aprendizado adaptativo". Além disso, foram considerados trabalhos que referenciavam o KnowEdu e outros artigos relevantes encontrados durante a pesquisa.

### **3.1 Detecção de Pré-requisitos**

O KnowEdu (CHEN et al., 2018) foi um dos primeiros esforços significativos para automatizar a criação de grafos de conhecimento educacionais, enfrentando o problema de como identificar relações de pré-requisito entre conceitos de forma eficiente. O objetivo principal do KnowEdu era construir automaticamente essas relações a partir de textos curriculares e dados de avaliação dos alunos, permitindo a criação de um grafo que refletisse as dependências instrucionais entre os tópicos de ensino. Para alcançar isso, os autores empregaram uma metodologia baseada em técnicas de mineração de regras e aprendizado de máquina, utilizando redes neurais para rotular sequências e inferir padrões entre os conceitos. A abordagem mostrou-se eficaz ao identificar padrões de conhecimento que eram difíceis de modelar manualmente. No entanto, o KnowEdu depende de um volume mínimo de dados educacionais para gerar inferências precisas, principalmente nos estágios iniciais do sistema, o que limita sua aplicabilidade em contextos de escassez de dados, criando o problema "Cold Start" explicado anteriormente. O trabalho atual busca superar essa limitação ao utilizar LLMs para fornecer estimativas iniciais robustas das relações de pré-requisito, permitindo a construção do grafo de conhecimento mesmo quando há pouca ou nenhuma informação prévia disponível, aumentando a escalabilidade e a aplicabilidade da abordagem em diferentes contextos educacionais.

### **3.2 Estimativa de Dificuldade**

O trabalho de Velde et al. (2020) abordou a questão da estimativa da dificuldade dos itens de aprendizagem, com o objetivo de melhorar a experiência adaptativa dos alunos, ajustando o conteúdo ao nível de habilidade de cada um. Garantir uma estimativa precisa da dificuldade é crucial para evitar que os alunos se sintam sobrecarregados com tarefas excessivamente complexas ou subestimados por atividades simplistas, promovendo um aprendizado mais eficaz e um maior engajamento.

Para resolver esse problema, os autores propuseram diferentes estratégias de estimativa, variando entre abordagens colaborativas e individualizadas, que utilizavam fontes de dados variadas para prever a dificuldade dos itens. Essas abordagens permitiram avanços na precisão da estimativa da dificuldade, demonstrando que mesmo uma estimativa preliminar era signifi-

cativamente mais eficaz do que a ausência de qualquer estimativa inicial ("Cold Start"). Esse resultado reforça a importância de fornecer estimativas iniciais para melhorar a adaptação dos conteúdos aos alunos.

A aplicação dessas estratégias é particularmente relevante para este trabalho, pois indica que estimativas iniciais, mesmo que imperfeitas, são fundamentais para criar um ponto de partida valioso para a personalização. Baseado nisso, o sistema proposto utiliza LLMs para fornecer essas estimativas preliminares de dificuldade, garantindo uma base adaptativa desde o início, mesmo em ambientes com escassez de dados de interação.

### **3.3 Definição de Dependências com Base na Taxonomia de Bloom**

Chystopolova et al. (2020) propuseram uma abordagem para identificar dependências entre resultados de aprendizagem utilizando a Taxonomia de Bloom como referência. O framework introduziu dois tipos fundamentais de relações: "REQUIRES", que estabelece hierarquias de pré-requisitos baseadas em níveis cognitivos, e "EXPANDS", que indica relações de complementaridade entre conceitos. Para identificar estas relações, os autores desenvolveram um processo de análise sintática e semântica dos resultados de aprendizagem, permitindo extrair automaticamente as conexões entre diferentes elementos curriculares.

A metodologia desenvolvida demonstrou como a Taxonomia de Bloom pode servir como base para estruturar relações entre componentes educacionais, estabelecendo tanto progressões hierárquicas de aprendizado quanto conexões laterais entre conceitos relacionados. O trabalho foi particularmente relevante ao mostrar que é possível automatizar parcialmente a identificação destas relações em currículos existentes, embora ainda exigindo considerável supervisão humana para validação.

No trabalho atual, adaptamos alguns destes conceitos, especialmente a utilização da Taxonomia de Bloom como estrutura organizadora e a distinção entre relações hierárquicas e de complementaridade, expandindo a abordagem para o contexto de geração automatizada de unidades de conhecimento.

### **3.4 Organização e Respostas a Consultas com Grafos**

O GraphRAG (EDGE et al., 2024) propõe uma solução para responder a perguntas complexas e globais em grandes coleções de documentos, abordando o problema de como gerar respostas abrangentes que vão além da recuperação de fragmentos específicos de texto. O objetivo do GraphRAG é estruturar e organizar o conhecimento contido nos documentos, de forma que perguntas amplas possam ser respondidas de maneira integrada e completa.

Para atingir esse objetivo, a metodologia empregada combina a geração aumentada por recuperação (RAG) com uma abordagem baseada em grafos para modelagem do conhecimento, utilizando o algoritmo Leiden para detectar comunidades dentro do grafo e aplicar uma abor-

dagem de map-reduce para sintetizar respostas. Além disso, LLMs são utilizados para gerar descrições das comunidades detectadas, permitindo uma visão coesa e hierárquica dos tópicos tratados.

O GraphRAG se destaca por criar uma estrutura modular de conhecimento, capaz de responder de forma eficiente a perguntas globais. No trabalho atual, foi feita a adaptação desses conceitos para estruturar e hierarquizar unidades de conhecimento. O algoritmo Leiden foi utilizado para identificar módulos de conceitos relacionados, com cada comunidade detectada representando um grupo coeso de tópicos educacionais.

Os resumos gerados para cada grupo de conceitos são tratados como unidades de conhecimento de granularidade superior, criando uma hierarquia de abstração que facilita a navegação e a personalização do aprendizado. Dessa forma, enquanto o GraphRAG foca na organização e resposta a questões globais, o presente trabalho adapta esses conceitos para construir um grafo educacional dinâmico e hierárquico, adequado para a personalização do aprendizado em diferentes níveis de granularidade.

3.5 Comparação dos Trabalhos Relacionados

O Quadro 2 apresenta uma análise comparativa dos trabalhos relacionados, destacando seus objetivos, métodos, contribuições e limitações. A análise crítica desses trabalhos fornece um panorama das lacunas existentes e direciona as contribuições da presente pesquisa.

Quadro 2: Comparação dos Trabalhos Relacionados

Autor e Ano	Objetivo	Abordagem	Principais Contribuições	Limitações
Chen et al. (2018)	Criar um grafo de conhecimento a partir de currículos escolares.	Algoritmos de mineração de regras e aprendizado de máquina.	Estruturação eficiente de conceitos educacionais; Identificação de padrões de conhecimento.	Depende de grandes volumes de dados; Problema de "Cold Start" em estágios iniciais.

Continuação na próxima página

Quadro 2 – Continuação da página anterior

Autor e Ano	Objetivo	Abordagem	Principais Contribuições	Limitações
Velde et al. (2020)	Estimar a dificuldade de unidades de conhecimento.	Modelos baseados em dados históricos e diferentes estratégias de estimativa (colaborativas e individualizadas).	Ajusta a dificuldade para melhorar a personalização; Demonstra a importância de estimativas iniciais.	A abordagem ao problema de "Cold Start" não pode ser adaptada para unidades de conhecimento recém criadas e sem nenhum dado de iteração de alunos
Chystopolova et al. (2020)	Definir dependências entre resultados de aprendizagem.	Taxonomia de Bloom Revisada; Relações "REQUIRES" e "EXPANDS".	Framework para categorizar dependências entre UCs; Abordagem estruturada para representar hierarquia e complementaridade.	Foco na definição de dependências, não na geração de UCs.
Edge et al. (2024)	Responder a perguntas complexas em grandes coleções de documentos.	Geração aumentada por recuperação (RAG); Abordagem baseada em grafos; Algoritmo Leiden; LLMs para gerar descrições.	Criação de uma estrutura modular de conhecimento; Respostas abrangentes a perguntas globais.	Foco na recuperação de informações, não na personalização do aprendizado.

Fonte: Elaborado pelo autor (2025).

4 ASPECTOS DE IMPLEMENTAÇÃO

Esta seção detalha a proposta para a construção automatizada de grafos de conhecimento educacional. O sistema resultante, desenvolvido no âmbito deste trabalho, visa extrair, organizar e classificar unidades de conhecimento a partir de diversos recursos textuais, utilizando Modelos de Linguagem de Grande Porte (LLMs) e algoritmos de grafos. A arquitetura foi concebida com foco em modularidade, orquestração de processos e gerenciamento de interações assíncronas

com serviços externos, como os LLMs.<sup>3</sup>

#### 4.1 Arquitetura Geral do Sistema

A arquitetura do sistema implementado, ilustrada na Figura 1 (Diagrama de Componentes de Alto Nível), é composta por dois módulos lógicos principais: o Sistema Central, que expõe a API de controle e gerencia a persistência dos dados, e a Camada de Orquestração e Processamento Especializado, responsável pela execução das tarefas assíncronas de análise e geração de conhecimento. O sistema interage com duas entidades externas principais: o Usuário, que inicia as execuções, e o Provedor LLM, que fornece os serviços de inteligência artificial.

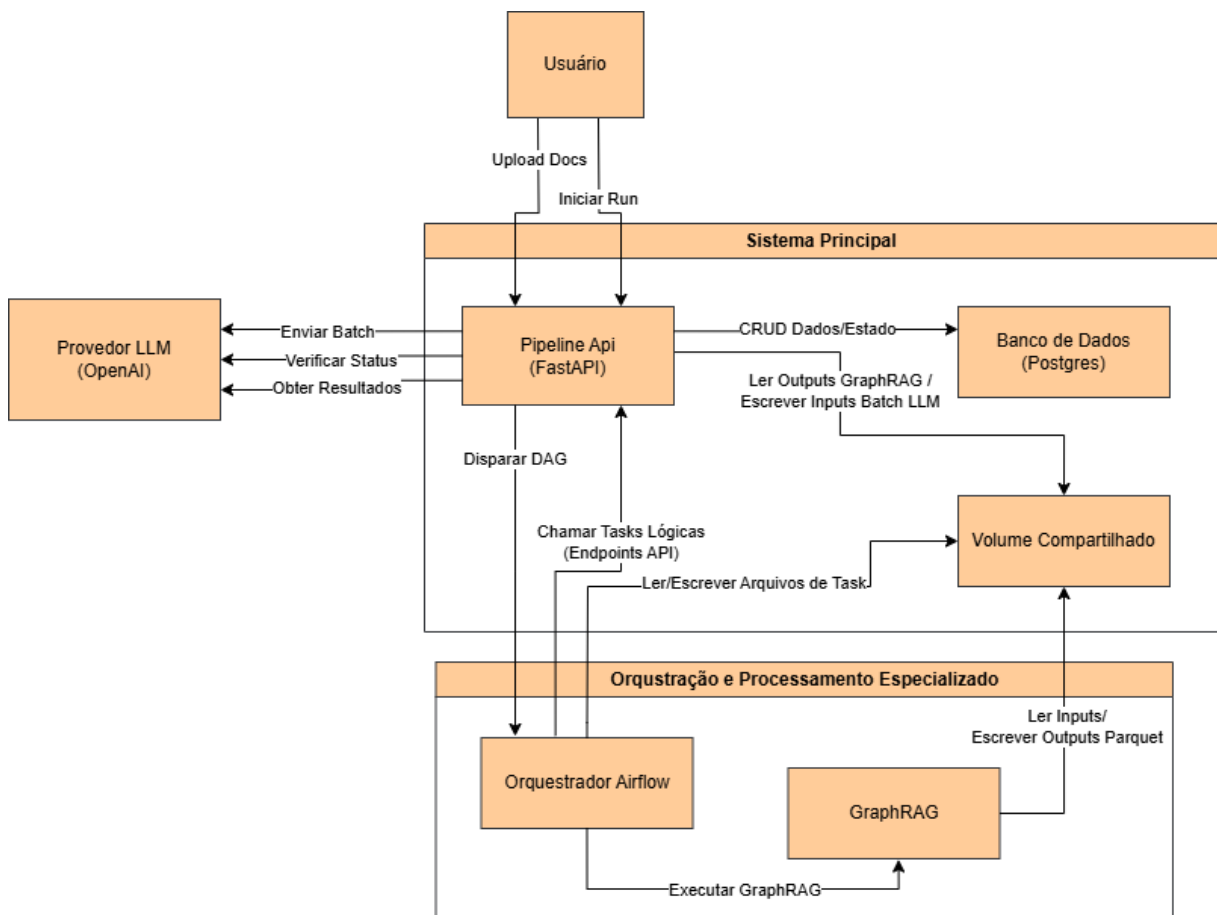


Figura 1: Diagrama de Componentes de Alto Nível do sistema proposto.

O **Usuário** interage com o sistema primariamente através da **Pipeline API** (desenvolvida em FastAPI), que serve como o ponto central de controle para o upload de recursos, inicialização de execuções do pipeline (denominadas *Pipeline Runs*) e consulta dos resultados. O **Sistema Central** também compreende o banco de dados **PostgreSQL**, responsável pela persistência de todos os dados de configuração, estado dos processos e artefatos gerados, e o acesso ao

<sup>3</sup>O código-fonte completo do sistema desenvolvido está disponível publicamente em um repositório no GitHub. Detalhes adicionais e o link de acesso podem ser encontrados no Apêndice I

**Provedor LLM** (OpenAI API) para as tarefas de geração e avaliação de UCs. Um **Volume Compartilhado** é utilizado para o intercâmbio de arquivos entre os diferentes componentes, como os textos de entrada e os outputs intermediários.

A camada de **Orquestração e Processamento Especializado** é composta pelo **Orquestrador Airflow**, que gerencia o fluxo de execução das etapas do pipeline, e por um contêiner Docker dedicado para o **GraphRAG**, responsável pelo processamento inicial dos textos e extração da estrutura base de conhecimento. A Pipeline API orquestra as chamadas ao Airflow, que por sua vez executa as tarefas lógicas (que são, em grande parte, chamadas a endpoints específicos da própria Pipeline API) e o contêiner GraphRAG.

## 4.2 Modelo de Dados e Persistência

O sistema utiliza um banco de dados PostgreSQL para armazenar metadados, dados intermediários e os resultados finais do pipeline. A estrutura do banco de dados foi definida e evoluída através de migrações gerenciadas pelo Alembic<sup>4</sup>. As principais entidades de dados estão detalhadas no Diagrama Entidade-Relacionamento (DER) apresentado no Apêndice A.

As tabelas centrais incluem:

- *PipelineRun*: Registra cada execução do pipeline, seu status, e os parâmetros de entrada.
- *Resource* e *PipelineRunResource*: Gerenciam os arquivos de entrada (materiais educacionais) e sua associação com cada *PipelineRun*.
- *PipelineBatchJob*: Tabela que desempenha um papel fundamental no gerenciamento do ciclo de vida das requisições em batch enviadas ao Provedor LLM. Armazena o ID do batch no provedor, o tipo de batch (geração de UCs ou avaliação de dificuldade), seu status atual (ex: *PENDING\_SUBMISSION*, *SUBMITTED*, *COMPLETED*, *FAILED*) e eventuais mensagens de erro. Esta tabela garante a idempotência e a rastreabilidade das interações assíncronas com o LLM.
- Tabelas *graphrag\_\**: Um conjunto de tabelas (*graphrag\_communities*, *graphrag\_entities*, etc.) que armazenam os resultados brutos extraídos pelo processamento do GraphRAG para cada *PipelineRun*.
- *KnowledgeUnitOrigin*: Armazena as "origens" das UCs, que podem ser entidades ou resumos de comunidades identificadas pelo GraphRAG. Contém informações como título, contexto e nível hierárquico.
- *GeneratedUcsRaw*: Persiste as UCs exatamente como foram geradas pelo LLM na etapa de geração, vinculadas à sua *KnowledgeUnitOrigin*.

---

<sup>4</sup><https://alembic.sqlalchemy.org/en/latest/>

- *DifficultyComparisonGroup* e *DifficultyGroupOriginAssociation*: Estruturas de dados que registram os grupos de UCs (via suas origens) que foram comparadas entre si durante a etapa de avaliação de dificuldade, juntamente com o nível de coerência do grupo.
- *KnowledgeUnitEvaluationsAggregatedBatch*: Armazena os resultados brutos da avaliação de dificuldade retornados pelo LLM para cada UC dentro de um grupo de comparação.
- *FinalKnowledgeUnit* e *FinalKnowledgeRelationship*: Representam o grafo de conhecimento educacional final, contendo as UCs processadas com seus *scores* de dificuldade agregados e as relações (REQUIRES, EXPANDS) entre elas.

Esta modelagem permite o rastreamento completo dos dados desde a entrada até a formação do grafo final, além de suportar a complexa lógica de interação com os serviços de LLM em batch.

### 4.3 Fluxo de Processamento do Pipeline

O pipeline de construção do grafo de conhecimento é orquestrado pelo Airflow, que coordena a execução de uma série de tarefas. A maioria dessas tarefas consiste em chamadas a endpoints específicos da Pipeline API, onde a lógica de processamento é efetivamente implementada. Este design centraliza a lógica de negócio na API, permitindo maior flexibilidade e testabilidade. O fluxo de dados e controle geral é ilustrado no Apêndice B.

#### 4.3.1 Inicialização do Pipeline e Preparação dos Inputs

Uma nova execução do pipeline (*PipelineRun*) é iniciada por uma chamada ao endpoint <POST/pipeline/{run\_id}/init> da Pipeline API. Nesta chamada, o usuário fornece os IDs dos *Resource* (materiais educacionais previamente carregados) que serão utilizados. A API registra a nova *PipelineRun* no banco de dados e dispara o DAG correspondente no Airflow.

A primeira tarefa do DAG, *prepare\_input\_files*, invoca um endpoint na API. A lógica deste endpoint é responsável por:

1. Localizar os arquivos originais (PDFs, TXTs) no Volume Compartilhado, com base nos *resource\_ids*.
2. Converter arquivos PDF para formato TXT utilizando a biblioteca *pypdfium2*. Arquivos TXT são mantidos como estão.
3. Organizar os arquivos TXT resultantes em um diretório de entrada específico para a *PipelineRun* corrente, dentro do Volume Compartilhado.
4. Atualizar o status e o caminho do arquivo processado na tabela *Resource*.

### 4.3.2 Processamento com GraphRAG

Esta tarefa é executada em um ambiente isolado, gerenciado pelo Airflow, que instancia dinamicamente um contêiner Docker configurado com a ferramenta GraphRAG utilizando o DockerOperator. O contêiner processa os arquivos TXT preparados na etapa anterior (do diretório de input da PipelineRun no Volume Compartilhado), executando o pipeline de indexação padrão do GraphRAG. Este pipeline inclui chunking de texto, extração de entidades e relações, resolução de entidades, detecção de comunidades (pelo algoritmo Leiden) e geração de resumos para estas. Os artefatos gerados pelo GraphRAG (arquivos Parquet como *communities.parquet*, *entities.parquet*, *relationships.parquet*, etc.) são salvos em um diretório de output específico da PipelineRun no Volume Compartilhado.

### 4.3.3 Preparação das Origens de Conhecimento

Após a conclusão do GraphRAG, a task *prepare\_origins* do Airflow chama o endpoint <POST/pipeline/{run\_id}/prepare-origins> da API, que realiza as seguintes ações:

1. Carrega os arquivos Parquet gerados pelo GraphRAG.
2. Processa o arquivo *communities.parquet* para identificar a relação pai-filho entre comunidades.
3. Enriquece os dados das entidades (do *entities.parquet*) com o ID da comunidade à qual pertencem.
4. Popula as tabelas *graphrag\_\** no banco de dados com os dados processados dos Parquets.
5. Finalmente, cria os registros na tabela *KnowledgeUnitOrigin*. As origens são derivadas tanto das entidades individuais quanto dos resumos das comunidades do GraphRAG. Cada *KnowledgeUnitOrigin* armazena o título, contexto (descrição da entidade ou resumo da comunidade) e informações hierárquicas.

Estas *KnowledgeUnitOrigin* servem como base para a subsequente geração de Unidades de Conhecimento.

### 4.3.4 Geração de Unidades de Conhecimento

A geração de UCs é iniciada pela task *submit\_generation\_batch* do Airflow, que chama o endpoint <POST/pipeline/{run\_id}/submit-batch/uc-generation> da API, que executa os seguintes passos:

1. Recupera as *KnowledgeUnitOrigin* do banco de dados para a *PipelineRun* atual.



2. Para cada origem selecionada, formata um prompt específico para o LLM (conforme Apêndice C.1, que instrui a geração de UCs para os seis níveis da Taxonomia de Bloom Revisada, utilizando o título e o contexto da origem).
3. Cria um registro na tabela *PipelineBatchJob* com *batch\_type='uc\_generation'* e status inicial *PENDING\_SUBMISSION*.
4. Envia as requisições formatadas como um batch para o Provedor LLM (OpenAI Batch API). O ID do batch retornado pelo provedor é armazenado no registro *PipelineBatchJob*.

A interação detalhada com a API de Batch do LLM, incluindo o upload do arquivo de requisições e a criação do job, é encapsulada pela classe *LLMClient* e segue o ciclo de vida ilustrado no Apêndice D.

Após a submissão, a task *wait\_generation\_batch\_completion* do Airflow (um *HttpSensor*) monitora periodicamente o status do batch chamando o endpoint `<GET/pipeline/{run_id}/batch-job-status/uc_generation>`. Este endpoint, por sua vez, consulta o Provedor LLM e atualiza o status do *PipelineBatchJob* no banco de dados (ex: para *SUBMITTED*, *COMPLETED*, *FAILED*).

Quando o batch é concluído pelo LLM, a task *process\_generation\_batch\_results* do Airflow chama o endpoint `<POST/pipeline/{run_id}/process-batch-results/uc_generation>`. Esta baixa o arquivo de resultados do Provedor LLM, processa cada linha (que contém a resposta para uma origem), extrai as UCs geradas para cada nível de Bloom, e salva esses dados na tabela *GeneratedUcsRaw*. O status do *PipelineBatchJob* é então atualizado para *COMPLETED*.

#### 4.3.5 Definição de Relações entre UCs

Com as UCs brutas geradas, a task *define\_relationships* do Airflow chama o endpoint `<POST/pipeline/{run_id}/define-relationships>` da API, que constrói as seguintes relações entre as UCs:

- **Relações REQUIRES:** São criadas entre UCs geradas a partir da mesma *KnowledgeUnitOrigin*. Especificamente, uma UC de um nível da Taxonomia de Bloom (ex: Entender) "requer" a UC do nível imediatamente anterior (ex: Lembrar) da mesma origem.
- **Relações EXPANDS:** São criadas para indicar conexões temáticas laterais entre UCs de diferentes *KnowledgeUnitOrigin*. Estas conexões são baseadas nas relações semânticas (e.g., "influencia", "é parte de") identificadas previamente pelo GraphRAG entre as *KnowledgeUnitOrigin* (sejam elas entidades ou comunidades). Embora a semântica específica desta relação original do GraphRAG seja persistida (ver Seção 4.2 tabela *graph-rag\_relationships*), optou-se, para o grafo educacional final desta iteração, por generalizá-las sob o rótulo EXPANDS. Esta decisão visa simplificar o grafo inicial, reduzindo sua

densidade visual e conceitual, e fornecer um indicador claro de exploração para tópicos relacionados.

No fim estas relações são salvas na tabela *KnowledgeRelationshipIntermediate*.

#### 4.3.6 Avaliação de Dificuldade das UCs

A avaliação da dificuldade das UCs segue um processo similar à geração de UCs, também utilizando a API de Batch do LLM. A task *submit\_difficulty\_batch* do Airflow chama o endpoint `<POST/pipeline/{run_id}/submit-batch/difficulty_assessment>`, que é responsável por:

1. Utilizar o *OriginDifficultyScheduler* para formar grupos de comparação. Este scheduler tenta agrupar *KnowledgeUnitOrigin* que são contextualmente próximas (mesma comunidade pai) e que ainda não atingiram um número mínimo de avaliações. O processo de formação de grupos é detalhado no Apêndice E.
2. Para cada grupo de comparação e para cada nível da Taxonomia de Bloom, são selecionadas as UCs correspondentes (da tabela *GeneratedUcsRaw*).
3. Formatar um prompt específico para o LLM (conforme Apêndice C.2), instruindo-o a avaliar comparativamente a dificuldade das UCs fornecidas (em uma escala de 0-100) e fornecer uma justificativa.
4. Criar um registro na tabela *DifficultyComparisonGroup* para cada conjunto de UCs enviado ao LLM, armazenando metadados como o nível de Bloom e o nível de coerência do grupo. As associações entre os grupos e as origens das UCs são salvas em *difficulty\_group\_origin\_association*.
5. Criar um registro na tabela *PipelineBatchJob* com *batch\_type='difficulty\_assessment'* e submeter o batch ao Provedor LLM.

Similarmente à geração de UCs, a task *wait\_difficulty\_batch\_completion* monitora o progresso, e *process\_difficulty\_batch\_results* (chamando *task\_process\_difficulty\_batch* na API) processa os resultados. As avaliações de dificuldade retornadas pelo LLM para cada UC (incluindo *score* e justificativa) são salvas na tabela *KnowledgeUnitEvaluationsAggregatedBatch*, vinculadas ao *comparison\_group\_id* correspondente. O ciclo de vida da interação com o LLM Batch API (Apêndice D) também se aplica aqui.

#### 4.3.7 Finalização e Geração dos Outputs

A última etapa do pipeline é executada pela task *finalize\_outputs* do Airflow, que chama o endpoint `</pipeline/{run_id}/finalize-outputs>`, que:

1. Carrega as UCs da *GeneratedUcsRaw* e todas as avaliações de dificuldade da *KnowledgeUnitEvaluationsAggregatedBatch*.
2. Para cada UC, calcula o *score* final de dificuldade como a média dos *scores* recebidos nas diversas avaliações em que participou. O número de avaliações e a concatenação das justificativas também são registrados.
3. Salva as UCs, agora enriquecidas com o *score* de dificuldade, na tabela *FinalKnowledgeUnit*.
4. Copia as relações da *KnowledgeRelationshipIntermediate* para a tabela *FinalKnowledgeRelationship*.
5. Atualiza o status da *PipelineRun* no banco de dados para *success*.

Ao final deste processo, o grafo de conhecimento educacional está construído e pronto para ser consultado através dos endpoints da API.

#### 4.4 Ferramentas e Tecnologias Utilizadas

O desenvolvimento do framework proposto envolveu o uso de um conjunto de ferramentas e tecnologias, incluindo:

- **Linguagem de Programação:** Python 3.10.
- **Framework API:** FastAPI.
- **ORM e Banco de Dados:** SQLAlchemy e PostgreSQL; Alembic para migrações.
- **Orquestração de Pipeline:** Apache Airflow.
- **Processamento de Texto e Grafos:** GraphRAG (Microsoft).
- **Modelos de Linguagem:** OpenAI API (Batch API, *gpt-4o*).
- **Manipulação de Dados:** Pandas.
- **Containerização:** Docker e Docker Compose.
- **Processamento de PDF:** Biblioteca *pypdfium2*.

## 5 RESULTADOS E ANÁLISE

Esta seção apresenta os resultados obtidos com a construção automatizada de grafos de conhecimento educacional. O sistema foi executado utilizando o livro *Uma Breve História do Mundo* de Blainey (2009) em formato PDF como material de entrada. A análise subsequente

visa avaliar a capacidade do sistema em gerar um grafo de conhecimento estruturado, a qualidade das Unidades de Conhecimento e das relações inferidas, bem como a plausibilidade das estimativas de dificuldade, considerando os objetivos propostos para este trabalho.

## 5.1 Configuração do Estudo de Caso

A execução do pipeline que gerou os resultados analisados nesta seção utilizou as seguintes configurações:

- **Material de Entrada:**

- Formato: PDF.
- Características Relevantes: O livro escolhido é uma obra de divulgação histórica que abrange desde a pré-história até eventos contemporâneos. Caracteriza-se por uma narrativa acessível e uma ampla gama de conceitos e eventos inter-relacionados, sendo adequado para a extração de conhecimento em diferentes níveis de granularidade e complexidade cognitiva, o que permitiu validar as principais funcionalidades do framework.

- **Modelo de Linguagem (LLM) Utilizado:**

- gpt-4o.

- **Parâmetros Chave do Pipeline:**

- Cada origem fez parte de no mínimo 3 grupos para comparação de dificuldade.

## 5.2 Análise Quantitativa do Grafo Gerado

A execução do pipeline sobre o material de entrada resultou na geração de um grafo de conhecimento com as seguintes características gerais.

- **Origens de Conhecimento (*KnowledgeUnitOrigin*):** Foram identificadas 2048 Origens, das quais 1718 (83.89%) são entidades (*entity*) e 330 (16.11%) são relatórios de comunidade (*community\_report*). Os relatórios de comunidade apresentaram uma estrutura hierárquica com 25 comunidades de Nível 0 (consideradas comunidades raiz ou sem pais explícitos no subgrafo analisado), 162 de Nível 1, 138 de Nível 2, e 5 de Nível 3, indicando diferentes graus de agregação temática.
- **Unidades de Conhecimento:** Foram geradas 12288 UCs finais (*FinalKnowledgeUnit*), correspondendo a 6 UCs (uma para cada nível de Bloom) por Origem de Conhecimento.

- **Estimativas de Dificuldade:** Das 12288 UCs finais, 12209 (99.36%) receberam uma estimativa de dificuldade. As 79 UCs restantes não foram avaliadas devido a falhas pontuais no processamento em batch com a API do LLM. As estimativas foram baseadas em uma média de 3.18 avaliações contextuais por UC (mínimo de 3, máximo de 4), realizadas em 7824 grupos de comparação distintos.
- **Relações Geradas:** Foram criadas 10240 relações REQUIRES, conectando consistentemente os níveis de Bloom dentro de cada Origem, e 8336 relações EXPANDS, estabelecendo conexões temáticas entre diferentes Origens.

Estes números indicam a formação de um grafo de conhecimento extenso e interconectado, propício para a exploração e aprendizado adaptativo. A Figura 2 apresenta um fragmento esquemático deste grafo, ilustrando a hierarquia entre diferentes tipos de Origens de Conhecimento (Comunidades e Entidades) e como as Unidades de Conhecimento, com seus respectivos níveis de Bloom e relações REQUIRES e EXPANDS, são estruturadas a partir dessas origens.

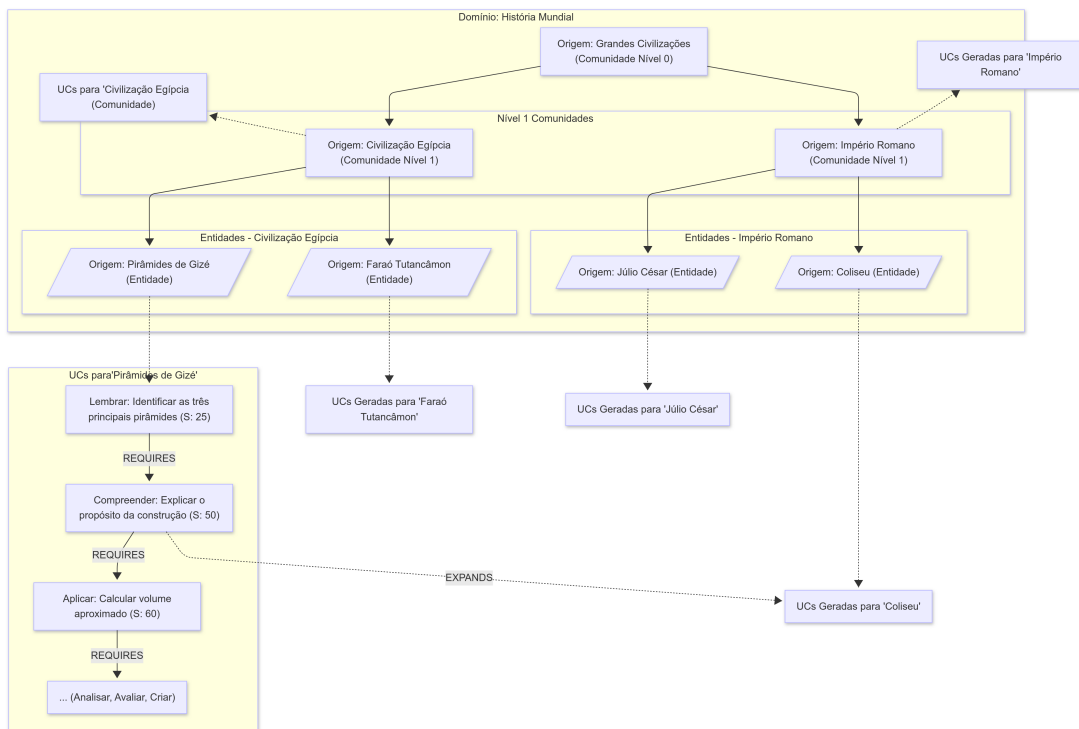


Figura 2: Fragmento representativo do Grafo de Conhecimento Educacional gerado. O diagrama ilustra a hierarquia das Origens de Conhecimento, a geração de Unidades de Conhecimento a partir delas, e as relações REQUIRES (dentro das UCs de uma mesma origem) e EXPANDS (conectando UCs de diferentes origens).

### 5.3 Análise Qualitativa das Unidades de Conhecimento e Estimativas de Dificuldade

Esta seção foca na avaliação qualitativa dos artefatos gerados pelo sistema, examinando a coerência, relevância e adequação das Unidades de Conhecimento e das estimativas de dificuldade.

dade, com base em exemplos representativos extraídos do grafo.

### 5.3.1 Geração de Unidades de Conhecimento e Alinhamento com a Taxonomia de Bloom

O sistema demonstrou capacidade de gerar UCs para os seis níveis da Taxonomia de Bloom Revisada para cada Origem de Conhecimento. As UCs geradas, em sua maioria, apresentaram clareza na formulação e relevância para o conceito da Origem e seu contexto.

- **Exemplo de Progressão Coerente em uma Origem:** A Origem “RIO NILO” exemplifica uma trilha com forte coesão temática:
  - Lembrar: “Identificar os principais afluentes do Rio Nilo e listar suas características.”
  - Entender: “Explicar a importância do Rio Nilo para a agricultura e o desenvolvimento das civilizações antigas no Egito.”
  - Aplicar: “Utilizar mapas para localizar o percurso do Rio Nilo e identificar as regiões que dependem de suas águas.”
  - Analisar: “Comparar o impacto das mudanças no fluxo do Rio Nilo com o desenvolvimento econômico das comunidades ribeirinhas.”
  - Avaliar: “Examinar as políticas de gestão da água do Rio Nilo e justificar a eficácia dessas políticas para a sustentabilidade das comunidades locais.”
  - Criar: “Desenvolver um projeto que proponha soluções inovadoras para os desafios enfrentados pelas comunidades que dependem do Rio Nilo.”

Neste caso, observa-se o alinhamento dos verbos de ação com os níveis de Bloom e uma progressão temática coesa, onde cada UC parece construir sobre a anterior, explorando o tema com crescente complexidade cognitiva.

- **Desafio na Coerência Temática Intra-Origem:** Em contraste, para Origens de Conhecimento com escopo muito amplo, como “ÁFRICA” (cujo contexto abrange desde berço da humanidade a desertos e desafios de desenvolvimento), a geração de uma única trilha de UCs de Bloom pode levar a “saltos temáticos” entre os níveis. As UCs geradas para esta origem foram:
  - Lembrar: “Identificar os principais desertos da África, incluindo o Deserto do Saara.”
  - Entender: “Explicar a importância da África como o berço da humanidade e como isso influenciou a migração humana.”
  - Aplicar: “Utilizar mapas geográficos para localizar as regiões tropicais e desérticas da África e discutir suas características.”

- Analisar: “Comparar as práticas religiosas na África com as da Europa, destacando as diferenças e semelhanças culturais.”
- Avaliar: “Examinar as condições sociais e econômicas na África ao longo dos últimos quatro milênios e justificar as razões para a falta de progresso.”
- Criar: “Desenvolver um projeto que proponha soluções para os desafios de governança e desigualdade social enfrentados pelos países africanos.”

Neste cenário, embora cada UC individual seja relevante para o tema mais amplo da “ÁFRICA”, a progressão entre elas (e.g., de “Lembrar desertos” para “Entender berço da humanidade”) ilustra um desafio conceitual. As UCs, ao abordarem facetas conceitualmente distantes da Origem “ÁFRICA” em níveis de Bloom sequenciais, não estabelecem um pré-requisito cognitivo tão direto e coeso quanto observado em origens mais delimitadas, como “RIO NILO”. As implicações dessa característica para a validade pedagógica das relações REQUIRES são discutidas em maior detalhe na Seção 5.4.1. Exemplos adicionais desta natureza são detalhados no Apêndice F.

### 5.3.2 Análise da Estimativa de Dificuldade

A metodologia de avaliação da dificuldade, detalhada na Seção 4.3.6, orientou o modelo de linguagem a estimar a complexidade relativa das Unidades de Conhecimento (UCs). Essa estimativa ocorreu por meio da comparação de UCs de *mesmo nível cognitivo* pertencentes a grupos de Origens semanticamente próximas, com *scores* atribuídos em uma escala de 0 a 100 para cada grupo, conforme o prompt no Apêndice C.2.

Os *scores* de dificuldade atribuídos indicam que o sistema diferenciou a complexidade das UCs com base em seu conteúdo e na natureza da tarefa. Por exemplo, no nível ‘Lembrar’, a UC “Identificar Tóquio como a capital do Japão” (*score* 14) foi considerada menos complexa que “Identificar os principais fatores que influenciam a dinâmica populacional, incluindo tecnologia, natureza e poluição” (*score* 75). Similarmente, no nível ‘Aplicar’, “Utilizar um telefone para realizar chamadas e enviar mensagens de texto” (*score* 23) recebeu um *score* inferior a “Utilizar a Teoria da Relatividade para resolver problemas relacionados à velocidade da luz e ao tempo” (*score* 85). Estes exemplos, e uma análise mais ampla dos extremos (Apêndice G), demonstram que o sistema captura nuances de dificuldade que transcendem o simples verbo da Taxonomia de Bloom associado.

A Tabela 1 resume a distribuição dos *score* de dificuldade para cada nível da Taxonomia de Bloom.

A análise dos *scores* revela uma considerável dispersão de dificuldade dentro de cada nível cognitivo. Este resultado é significativo, pois demonstra a capacidade do sistema em atribuir um amplo espectro de dificuldades com base no conteúdo específico de cada UC, em vez de depender unicamente do nível de da Taxonomia de Bloom. De fato, a metodologia buscava

Tabela 1: Distribuição dos Scores de Dificuldade por Nível Cognitivo de Bloom.

Nível de Bloom	Mín.	Máx.	Média	Mediana	N (UCs Aval.)
Lembrar	14	75	43,68	43,0	2039
Entender	37	84	64,22	65,0	2035
Aplicar	23	85	62,18	63,0	2030
Analisar	40	85	66,91	67,0	2034
Avaliar	40	85	69,90	70,0	2033
Criar	40	90	70,64	71,0	2038

Fonte: Elaborado pelo autor (2025).

exatamente essa diferenciação granular, comparando UCs de mesmo nível entre si.

Adicionalmente, observa-se uma tendência de aumento nos *scores* médios de dificuldade à medida que se avança nos níveis da Taxonomia de Bloom. Idealmente, como as UCs foram comparadas primariamente com outras de mesmo nível cognitivo, esperava-se que as médias globais por nível fossem mais similares, indicando uma efetiva normalização dos *scores* dentro de cada grupo de comparação. A tendência de médias crescentes sugere, no entanto, que a natureza inerente da tarefa cognitiva de cada nível da Taxonomia de Bloom (por exemplo, a maior complexidade em formular uma tarefa de ‘Criar’ em comparação com uma de ‘Lembrar’, conforme o prompt no Apêndice C.1) ainda influencia a percepção geral de dificuldade pelo LLM. Em outras palavras, tarefas cognitivas de ordem superior tendem a ser avaliadas como mais difíceis, mesmo quando a comparação é feita estritamente entre tarefas do mesmo nível cognitivo. Este é um ponto relevante para futuras investigações sobre a calibração das estimativas de dificuldade geradas por LLMs.

#### 5.4 Análise Qualitativa das Relações Geradas

O sistema gera dois tipos principais de relações entre UCs: REQUIRES (hierarquia de Bloom intra-origem) e EXPANDS (conexões temáticas inter-origens).

##### 5.4.1 Relações REQUIRES e a Hierarquia da Taxonomia de Bloom

As relações REQUIRES conectam Unidades de Conhecimento de níveis adjacentes da Taxonomia de Bloom que foram geradas a partir da mesma Origem de Conhecimento. Especificamente, uma UC de um determinado nível N da taxonomia é definida como “requerendo” a UC do nível N-1 imediatamente anterior, pertencente à mesma Origem.

A validade pedagógica dessas relações REQUIRES como pré-requisitos diretos e significativos é fortalecida quando a trilha de UCs gerada para uma Origem mantém uma forte coesão temática. Isso foi observado, por exemplo, na Origem “RIO NILO” (discutida na Seção 5.3.1), onde cada UC parece construir de forma lógica sobre a anterior, explorando diferentes aspectos do mesmo tema central de maneira progressiva.



No entanto, conforme antecipado na Seção 5.3.1 com o exemplo da Origem "ÁFRICA", em casos de *KnowledgeUnitOrigin* com escopo muito amplo, a imposição de uma única trilha da Taxonomia de Bloom pode levar a "saltos temáticos" consideráveis entre os níveis. Nestas situações, onde as UCs sequenciais abordam facetas conceitualmente distantes dentro da mesma Origem, a relação REQUIRES formalmente estabelecida perde parte de sua força como um pré-requisito pedagógico direto. Ela passa a ser interpretada mais como uma conexão estrutural ou "fraca", imposta pela progressão da Taxonomia de Bloom, em contraste com as relações REQUIRES "fortes" observadas em Origens coesas como "RIO NILO". Por exemplo, a transição da UC "Lembrar desertos" para "Entender berço da humanidade" dentro da Origem "ÁFRICA" ilustra como o aprendizado da primeira pode não constituir uma base cognitiva direta para a segunda.

Isso sugere que a utilidade das REQUIRES como guias efetivos para o aprendizado é altamente dependente da granularidade e do foco da *KnowledgeUnitOrigin* a partir da qual as UCs são derivadas. Consequentemente, embora o sistema estabeleça formalmente relações REQUIRES para todas as Origens, a validade dessas conexões como pré-requisitos cognitivos robustos é maior para Origens com escopo bem definido. Para Origens amplas, as relações REQUIRES "fracas" resultantes evidenciam uma limitação da aplicação de uma única trilha linear da Taxonomia de Bloom para representar a progressão em tópicos multifacetados, onde a interdependência entre diferentes aspectos do conhecimento pode ser mais complexa do que uma simples sequência hierárquica.

#### 5.4.2 Relações EXPANDS e a Conectividade do Grafo

Foram geradas 8336 relações EXPANDS, conectando UCs de 'Lembrar' e 'Entender' de Origens distintas, baseadas nas relações semânticas do GraphRAG. Uma análise quantitativa revelou que a média de expansões para Origens conectadas é de aproximadamente 2.68. Contudo, 24% das Origens não possuem expansões de saída, enquanto menos de 1% (cerca de 18 Origens, como "EUROPA" com 104 expansões) atuam como "hubs" com um número elevado de conexões. Adicionalmente, 69% das Origens que se expandem o fazem para 5 ou menos outras Origens, sugerindo uma estrutura de rede com conectividade heterogênea.

Qualitativamente, um exemplo da aplicação da relação EXPANDS ocorre entre as Origens "19TH CENTURY" e "RAINHA VITORIA". Neste caso, o processamento inicial com GraphRAG identificou uma relação semântica entre os conceitos, como por exemplo, que "RAINHA VITORIA" é uma figura chave associada ao "19TH CENTURY". Com base nessa conexão fundamental entre as Origens, o sistema estabelece que a UC "Lembrar: Identificar os principais eventos que marcaram o século XIX..." EXPANDS para a UC "Lembrar: Identificar os anos em que a Rainha Victoria reinou...". Esse tipo de relação facilita a exploração de tópicos interconectados. A decisão de limitar a materialização de EXPANDS aos níveis cognitivos inferiores ("Lembrar" e "Entender") e de usar um tipo genérico de relação foi uma estratégia para simpli-

ficar o grafo inicial e focar na conectividade básica.

### 5.4.3 Análise da Cobertura de Múltiplas Facetas do Conhecimento Através da Estrutura Hierárquica

O framework implementado resultou na construção de um grafo de conhecimento dotado de uma estrutura hierárquica, evidenciando a capacidade do sistema em representar conceitos em distintos níveis de granularidade. Conforme observado nos dados gerados (detalhados no Apêndice H), as *KnowledgeUnitOrigin* do tipo *community\_report* formam hierarquias: comunidades de nível superior (*level* mais baixo, como 0) representam temas de maior abrangência (menor granularidade), enquanto comunidades de nível inferior (*level* mais alto) detalham subtemas (maior granularidade). As *KnowledgeUnitOrigin* do tipo *entity* figuram como nós folha nesta hierarquia, representando conceitos ou entidades específicas encapsuladas pelo escopo de sua *community\_report* pai imediata.

Esta estrutura hierárquica é fundamental para entender como o sistema aborda a cobertura dos **diversos aspectos ou subtemas (aqui referidos como "facetas")** de um domínio de conhecimento. Conforme discutido na Seção 5.3.1 (“Desafio na Coerência Temática Intra-Origem”), a geração de Unidades de Conhecimento para uma única *KnowledgeUnitOrigin* segue uma progressão linear através dos níveis da Taxonomia de Bloom, o que pode levar a uma exploração focada em um aspecto principal, especialmente se a origem for muito ampla. No entanto, é a organização hierárquica das próprias *KnowledgeUnitOrigin* dentro do grafo gerado que constitui o mecanismo chave, permitindo a representação e a potencial exploração de **múltiplas facetas** do conhecimento em distintos nós do grafo.

Analisando, por exemplo, a *community\_report* de nível 0 (raiz) intitulada “Igreja e Renascimento: A Influência Cultural e Religiosa”, observa-se que ela representa um tema de baixa granularidade (amplo). As UCs geradas para esta origem, como a UC de nível Avaliar: “*Julgar a importância da Igreja como patrona das artes durante o Renascimento e seu impacto na evolução cultural da época.*”, demonstram uma tendência a exigir síntese e uma perspectiva panorâmica do tema.

Esta comunidade de nível 0, nos resultados, apresenta *community\_reports* filhas de nível 1, que representam temas de granularidade média (mais específicos) e, portanto, **diferentes facetas do tema principal**. Um exemplo é a comunidade “Florence and the Scientific Renaissance”. Esta, por sua vez, foca em uma **faceta particular** do tema “Igreja e Renascimento”. As UCs geradas para “Florence and the Scientific Renaissance”, como “*Entender: Explicar a importância de Florença como um centro de arte e ciência durante a Renascença. . .*”, exploram este subtema específico.

Finalmente, dentro da comunidade “Florence and the Scientific Renaissance” (que, como apresentado, é uma faceta de “Igreja e Renascimento”), o grafo inclui *entities* como “GALILEU”. Estas entidades representam conceitos de alta granularidade (altamente específicos) e

funcionam como componentes detalhados do tema da comunidade. As UCs para a entidade “GALILEU”, como “*Entender: Explicar como as descobertas de Galileu sobre a Lua e o Sol desafiaram as crenças científicas da época.*”, permitem um aprofundamento em **aspectos ainda mais específicos** do conhecimento, detalhando uma faceta da comunidade à qual pertencem.

Dessa forma, a análise da estrutura do grafo gerado revela que, enquanto uma única trilha de UCs para uma *KnowledgeUnitOrigin* específica explora aquele conceito de forma progressiva, a estrutura hierárquica geral do grafo possibilita a representação de um domínio de conhecimento de forma **multifacetada**. Ao transitar entre *KnowledgeUnitOrigin* de diferentes níveis hierárquicos e de granularidade, torna-se possível acessar tanto visões gerais de temas amplos quanto detalhes de suas **facetas constituintes**. Esta característica da estrutura gerada indica potencial para futuras aplicações em sistemas que busquem oferecer caminhos de exploração de conhecimento com amplitude e profundidade. A capacidade do GraphRAG de identificar estas comunidades e suas relações hierárquicas, combinada com a geração de UCs orientada pela Taxonomia de Bloom para cada *KnowledgeUnitOrigin*, resultou em uma base de conhecimento estruturada e de grande granularidade para análise e potenciais expansões.

## 5.5 Discussão Geral dos Resultados e Limitações

O framework implementado demonstra viabilidade para a construção automatizada de grafos de conhecimento educacional orientados pela Taxonomia de Bloom. O sistema gera UCs alinhadas aos níveis de Bloom e atribui estimativas de dificuldade sensíveis ao conteúdo, mitigando o problema de "cold start". As relações REQUIRES e EXPANDS, juntamente com a estrutura hierárquica de comunidades, fornecem uma base para navegação e aprendizado.

Os principais pontos fortes observados incluem a capacidade de gerar UCs com progressão cognitiva e de diferenciar a dificuldade com base no conteúdo. Esta última capacidade foi evidenciada pela ampla dispersão dos *scores*, correspondendo ao esperado para as UCs analisadas. Adicionalmente, as relações EXPANDS demonstraram potencial para conectar temas relevantes.

Contudo, foram identificados desafios e limitações. Muitas das limitações inerentes a sistemas baseados em LLMs derivam de sua natureza fundamental: são modelos probabilísticos treinados em vastos corpus de dados, o que lhes confere notável fluência, mas não um entendimento ou raciocínio causal profundo no sentido humano (KIPKEBUT, 2024).

A coerência semântica das relações REQUIRES pode ser comprometida por "saltos temáticos" em trilhas de UCs de origens muito amplas. Embora a geração de UCs para uma única *KnowledgeUnitOrigin* siga uma progressão linear pela Taxonomia de Bloom, a estrutura hierárquica do grafo como um todo, com diferentes níveis de *KnowledgeUnitOrigin* (comunidades e entidades), permite a representação e exploração de múltiplas facetas do conhecimento, como discutido anteriormente na seção 5.4.3. A literatura sugere que a melhoria da coerência semântica em textos gerados por LLM pode ser alcançada através do uso de ontologias e loops

de refinamento iterativo (VSEVOLODOVNA; MONTI, 2025), e métricas como "coerência lógica" e "relevância da resposta" (Confident AI, 2025) são importantes para avaliar a qualidade das UCs.

A qualidade geral é dependente das capacidades do LLM em interpretar os prompts e gerar conteúdo relevante, bem como da qualidade da estruturação inicial do conhecimento pelo GraphRAG. A literatura discute extensivamente os desafios e limitações dos LLMs na educação, incluindo questões de precisão, confiabilidade, vieses e a dificuldade em promover pensamento crítico (KASNECI et al., 2023). Casos documentados onde LLMs fornecem informações incorretas ou "alucinações", mesmo para tópicos proeminentes (LI; WU, 2025), reforçam a necessidade de cautela e de mecanismos de validação ao utilizar saídas de LLM em contextos educacionais. O "viés de automação", onde usuários tendem a aceitar respostas de LLMs acriticamente (LI; WU, 2025), também é uma consideração relevante.

Adicionalmente, este trabalho não incluiu uma avaliação com usuários finais, que seria um passo importante para validar a utilidade pedagógica e a experiência de uso do grafo gerado. A importância desta etapa é reconhecida na literatura, que avança na proposição de frameworks específicos para essa finalidade. Um exemplo notável é a proposta de sistemas de índice de avaliação para recursos educacionais digitais gerados por IA (AIGDERs), que considera dimensões como autenticidade, precisão e relevância (HUANG et al., 2025), ressaltando a complexidade e a necessidade de uma validação estruturada.

As 79 UCs (0.64%) que não receberam *score* de dificuldade também indicam uma área para melhoria na robustez do pipeline de avaliação, possivelmente relacionada a falhas pontuais no processamento em lote ou interações com a API do LLM, demandando atenção à robustez técnica do pipeline.

Apesar dessas limitações, o framework estabelecido oferece uma base sólida para futuras expansões, como o refinamento dos prompts de geração, a modelagem mais explícita de facetas, a validação de relações, e a eventual integração com sistemas de aprendizado adaptativo para avaliação empírica.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

A implementação de sistemas de aprendizado adaptativo em larga escala enfrenta o desafio fundamental de construir modelos de domínio ricos e estruturados, um processo tradicionalmente manual, lento e de difícil escalabilidade. Esta pesquisa abordou diretamente essa lacuna, focando no problema de "cold start", onde a ausência de dados de interação de novos alunos impede a personalização eficaz do ensino. O objetivo central foi propor e desenvolver um framework para a construção automatizada de grafos de conhecimento educacional a partir de fontes textuais não estruturadas, utilizando a capacidade de processamento de Modelos de Linguagem sob a orientação pedagógica da Taxonomia de Bloom.

Para enfrentar o desafio proposto, este trabalho apresentou um framework funcional que in-

tegra extração, geração e avaliação de conhecimento. A principal contribuição reside na criação de um grafo de conhecimento operacional desde sua concepção, mitigando o problema de "cold start" ao gerar estimativas iniciais de dificuldade e relações de pré-requisito (REQUIRES) e expansão temática (EXPANDS) entre as Unidades de Conhecimento (UCs). Adicionalmente, a integração da Taxonomia de Bloom como princípio norteador assegurou que o grafo gerado não fosse apenas uma estrutura de dados, mas um artefato estruturado segundo princípios pedagógicos, que respeita a progressão cognitiva natural da aprendizagem. A arquitetura hierárquica, resultante da detecção de comunidades e entidades, enriquece o modelo ao permitir a exploração do conhecimento em múltiplos níveis de granularidade.

Apesar dos resultados obtidos na pesquisa experimental, é importante reconhecer as limitações da abordagem atual. A coerência semântica das relações REQUIRES pode ser comprometida quando as UCs são geradas para tópicos de escopo muito amplo, resultando em "saltos temáticos" entre os níveis de Bloom. Observou-se também que a avaliação de dificuldade, embora capaz de diferenciar conceitos dentro do mesmo nível cognitivo, ainda demonstra uma tendência dos LLMs em atribuir *scores* mais altos para tarefas de ordem superior, indicando uma área para calibração futura. Finalmente, a qualidade do grafo é intrinsecamente dependente da capacidade do LLM e da estruturação inicial fornecida pelo GraphRAG, e o trabalho ainda carece de uma validação empírica com educadores e alunos para atestar sua utilidade pedagógica em um cenário real.

As limitações identificadas e a base sólida estabelecida por este trabalho abrem caminhos para diversas pesquisas futuras. Uma direção imediata é o refinamento da geração de UCs para origens amplas, explorando a criação de múltiplas trilhas de Bloom para diferentes facetas de um mesmo conceito. Futuramente, a estrutura do grafo pode ser aproveitada para a geração automática de atividades de aprendizagem alinhadas a cada nível cognitivo, bem como para a criação de um sistema de recomendação de recursos complementares. O passo decisivo para validar a eficácia do framework, contudo, será a integração deste grafo a um sistema de aprendizado adaptativo real para conduzir estudos com usuários. Tal validação permitirá não apenas medir o impacto da ferramenta na aprendizagem, mas também coletar dados para refinar e calibrar continuamente as estimativas de dificuldade, tornando o sistema uma ferramenta mais precisa e eficaz para o aprendizado personalizado.

## Referências

BLAINEY, G. **Uma Breve História do Mundo**. 2. ed. São Paulo: Editora Fundamento, 2009.

CHEN, P. et al. Knowedu: A system to construct knowledge graph for education. **IEEE Access**, v. 6, p. 31553–31563, 2018.

CHYSTOPOLOVA, Y. et al. Identification of dependencies between learning outcomes in computing science curricula for primary and secondary education – on the way to personalized learning paths. In: **Informatics in Schools. Engaging Learners in Computational Thinking: 13th International Conference, ISSEP 2020, Tallinn, Estonia, November**

**16–18, 2020, Proceedings.** Berlin, Heidelberg: Springer-Verlag, 2020. p. 185–196. ISBN 978-3-030-63211-3. Disponível em: <[https://doi.org/10.1007/978-3-030-63212-0\\_15](https://doi.org/10.1007/978-3-030-63212-0_15)>.

Confident AI. **LLM Evaluation Metrics: The Ultimate LLM Evaluation Guide.** 2025. Blog Post. Acessado em 8 de junho de 2025. Disponível em: <<https://www.confident-ai.com/blog/llm-evaluation-metrics-everything-you-need-for-llm-evaluation>>.

EDGE, D. et al. **From Local to Global: A Graph RAG Approach to Query-Focused Summarization.** 2024. Disponível em: <<https://arxiv.org/abs/2404.16130>>.

HUANG, Q. et al. Evaluating the quality of ai-generated digital educational resources for university teaching and learning. **Systems**, v. 13, n. 3, 2025. ISSN 2079-8954. Disponível em: <<https://www.mdpi.com/2079-8954/13/3/174>>.

KASNECI, E. et al. Chatgpt for good? on opportunities and challenges of large language models for education. **Learning and Individual Differences**, v. 103, p. 102274, 2023. ISSN 1041-6080. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1041608023000195>>.

KIPKEBUT, A. Large Language Models: Challenges and Opportunities in Education. **Iconic Research and Engineering Journals**, v. 8, n. 2, p. 918–924, aug 2024. ISSN 2456-8880. Acessado em 8 de junho de 2025. Disponível em: <<https://www.irejournals.com/formatedpaper/1706165.pdf>>.

KOEDINGER, K. R.; CORBETT, A. T.; PERFETTI, C. The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. **Cognitive Science**, v. 36, n. 5, p. 757–798, 2012. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1551-6709.2012.01245.x>>.

KOJIMA, T. et al. **Large Language Models are Zero-Shot Reasoners.** 2023. Disponível em: <<https://arxiv.org/abs/2205.11916>>.

LI, R.; WU, T. Delving into the practical applications and pitfalls of large language models in medical education: Narrative review. **Adv Med Educ Pract**, New Zealand, v. 16, p. 625–636, abr. 2025.

MARCHETI, A.; BELHOT, R. Taxonomia de bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. **Gestão & Produção**, v. 17, 01 2010.

PELÁNEK, R. Managing items and knowledge components: domain modeling in practice. **Educational Technology Research and Development**, v. 68, p. 1–22, 10 2019.

PELÁNEK, R. Adaptive learning is hard: Challenges, nuances, and trade-offs in modeling. **International Journal of Artificial Intelligence in Education**, 03 2024.

VELDE, M. et al. **Alleviating the Cold Start Problem in Adaptive Learning using Data-Driven Difficulty Estimates.** 2020.

VSEVOLODOVNA, R. I. M.; MONTI, M. **Enhancing Large Language Models through Neuro-Symbolic Integration and Ontological Reasoning.** 2025. ArXiv preprint arXiv:2504.07640.

ZHAO, W. X. et al. **A Survey of Large Language Models**. 2025. Disponível em: <<https://arxiv.org/abs/2303.18223>>.

## APÊNDICE A – DIAGRAMA ENTIDADE-RELACIONAMENTO (DER) DO SISTEMA

Para garantir a clareza e legibilidade, o Diagrama Entidade-Relacionamento (DER) do sistema foi dividido em três partes, cada uma representando uma fase lógica do pipeline de processamento de dados. Os diagramas a seguir ilustram: (1) a ingestão de dados e estruturação inicial com GraphRAG, (2) a geração e avaliação das Unidades de Conhecimento, e (3) a síntese do grafo de conhecimento final.

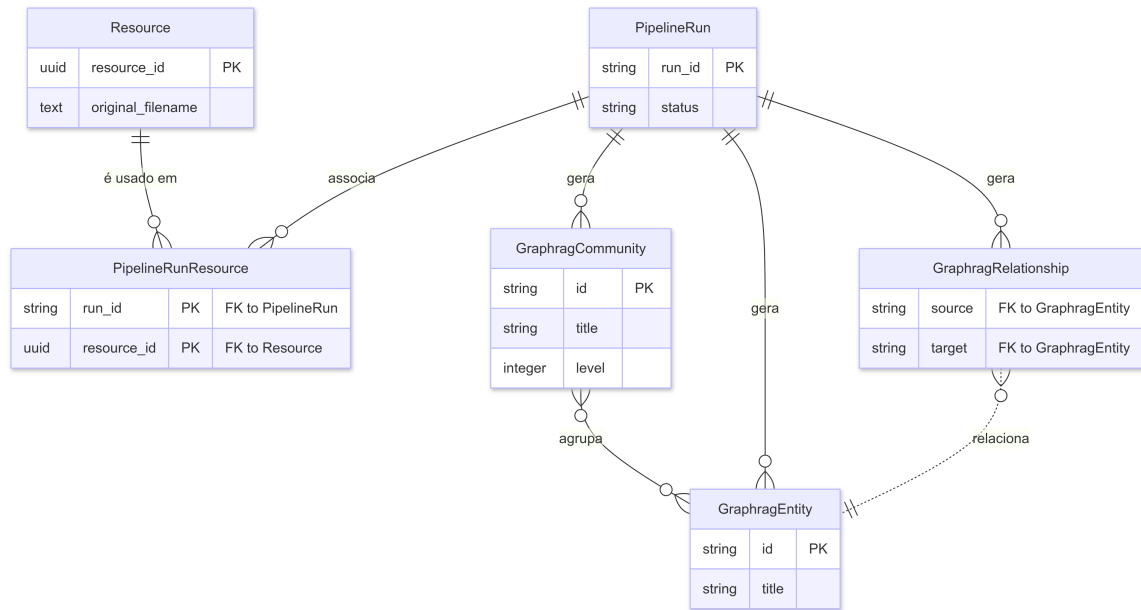


Figura A.1: Ingestão de Dados e Estruturação com GraphRAG.



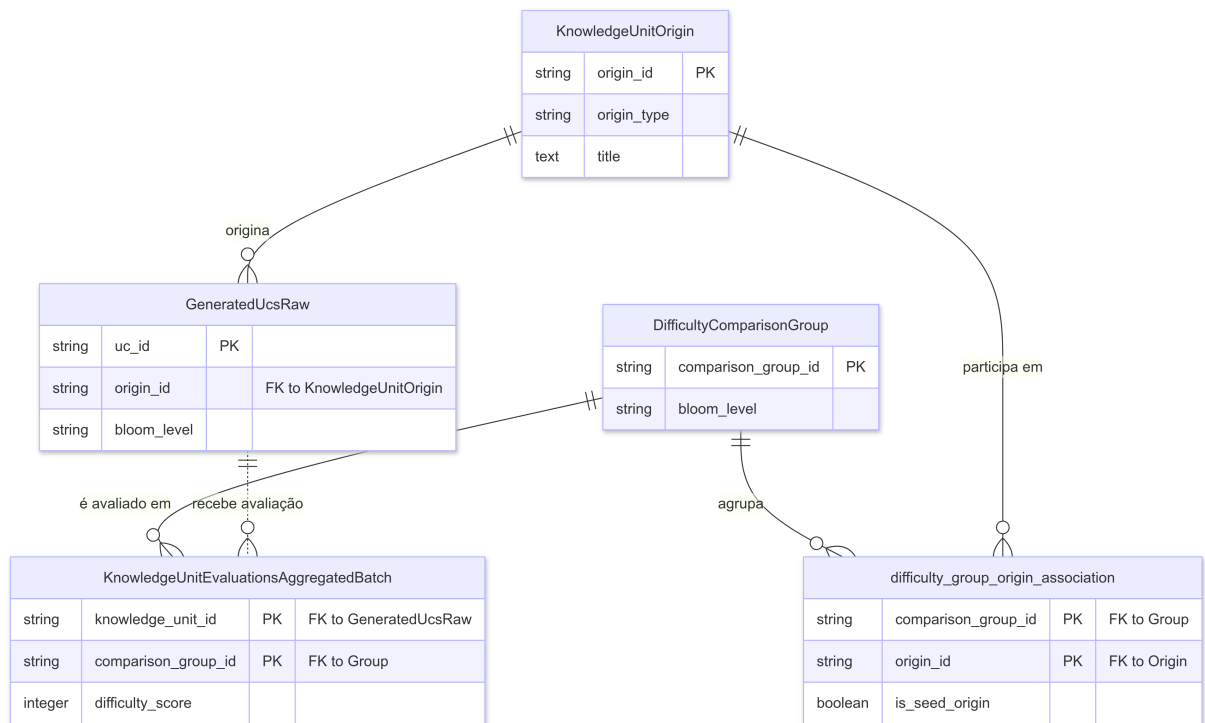


Figura A.2: Geração e Avaliação de Unidades de Conhecimento.

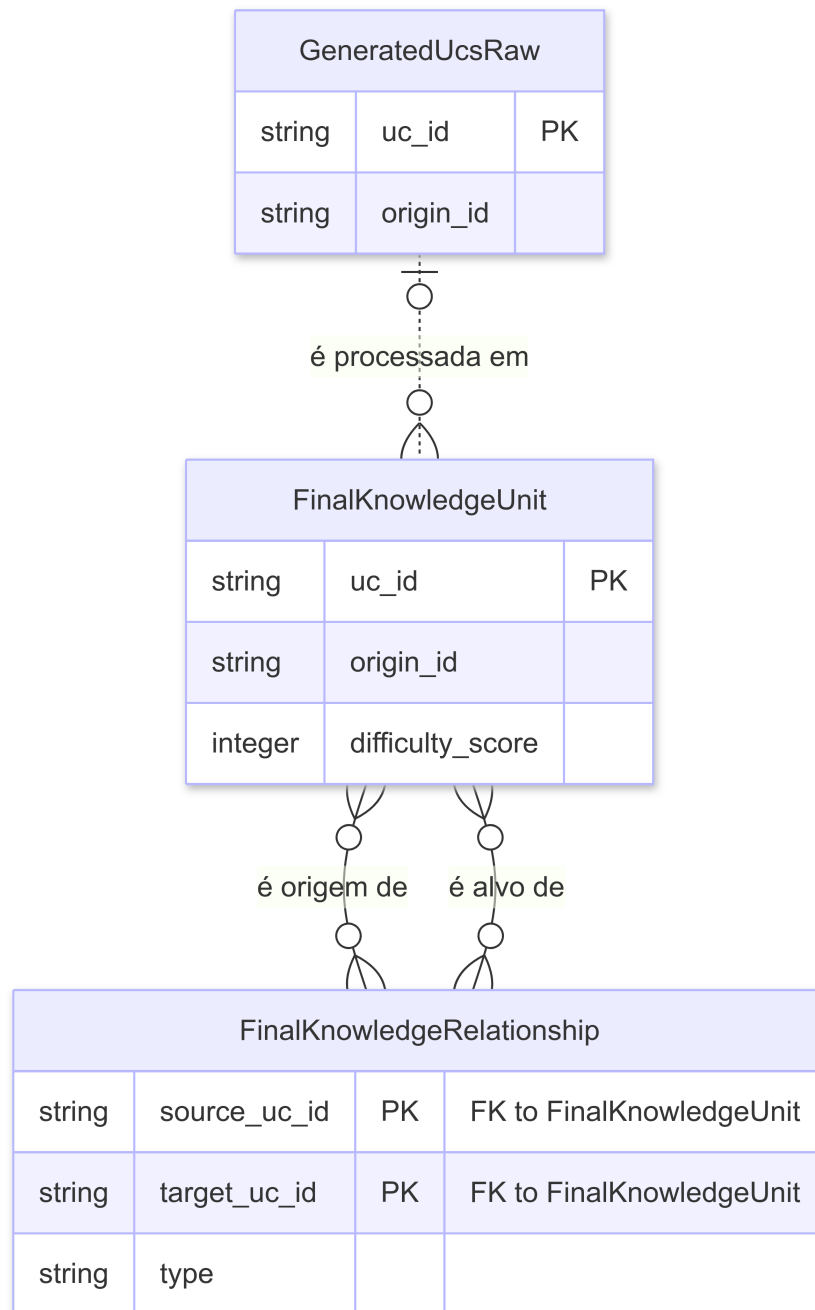


Figura A.3: Síntese do Grafo de Conhecimento Final.

## APÊNDICE B – DIAGRAMAS DE FLUXO DE PROCESSAMENTO DO PIPELINE

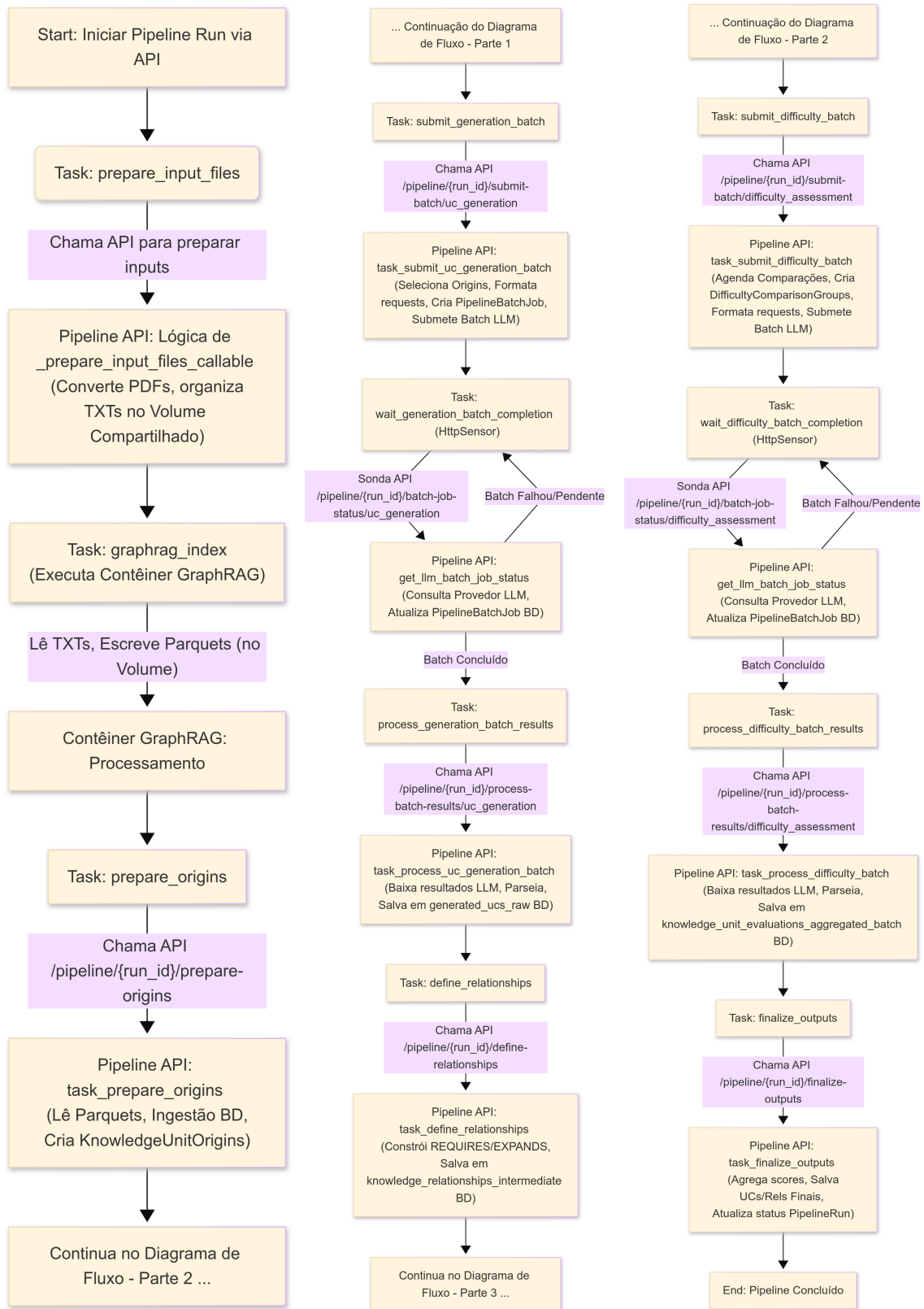


Figura B.1: Diagrama de Fluxo de Dados Detalhado do Pipeline

## APÊNDICE C – PROMPTS PARA GERAÇÃO DE UCS E AVALIAÇÃO DE DIFICULDADE

### C.1 Prompt para Geração de UCS

O prompt a seguir foi elaborado para guiar o modelo de linguagem na geração de UCS. Ele inclui uma breve explicação sobre o que são UCS e a Taxonomia de Bloom Revisada, além de instruções claras sobre a tarefa a ser realizada, com ênfase na utilização dos verbos da taxonomia e na relação entre as UCS e o conceito principal. A estrutura do prompt tenta maximizar a capacidade do modelo de linguagem de gerar UCS relevantes, completas e personalizadas para cada conceito. **Conforme discutido na seção sobre LLMs, a otimização do número de tokens não foi priorizada na elaboração deste prompt, visando, neste momento, apenas a validação da viabilidade da tarefa.**

PROMPT:

Você é um especialista em educação, capaz de gerar Unidades de Conhecimento (UCs) abrangentes e personalizadas com base na Taxonomia de Bloom Revisada.

#### **O que é uma UC?**

Uma UC representa um conceito, habilidade ou informação que um aluno deve aprender. As UCs podem variar em complexidade, desde a simples memorização de um fato até a aplicação do conhecimento em situações complexas e a criação de algo novo.

#### **O que é a Taxonomia de Bloom Revisada?**

A Taxonomia de Bloom Revisada é uma estrutura que classifica os objetivos de aprendizagem em seis níveis cognitivos:

**Lembrar:** reconhecer e reproduzir informações.

Verbos: identificar, listar, descrever, nomear, rotular, localizar, selecionar, definir, recitar, delinear, declarar, repetir, memorizar, enunciar, registrar.

2. **Entender:** interpretar, exemplificar, classificar, resumir, inferir, comparar e explicar informações.

Verbos: explicar, resumir, parafrasear, classificar, exemplificar, converter, traduzir, ilustrar, demonstrar, apresentar, fornecer exemplos, mostrar, categorizar, organizar, agrupar, diferenciar, distinguir, sintetizar, condensar, encurtar, recapitular, compendiar, concluir, deduzir, interpretar, extrapolar, interpolar, comparar, contrastar, relacionar, analisar, distinguir, descrever, discutir, esclarecer, interpretar, justificar.

3. **Aplicar:** executar e implementar o conhecimento em situações novas.

Verbos: realizar, implementar, aplicar, usar, empregar, manipular, operar, construir, desenvolver, simular, projetar, experimentar, solucionar.

4. **Analisar:** diferenciar, organizar, atribuir e concluir sobre as partes e relações de uma informação.

Verbos: discriminar, distinguir, separar, comparar, contrastar, estruturar,

sequenciar, integrar, classificar, delinear, esquematizar, determinar, conectar, relacionar, identificar, analisar, encontrar, deduzir, inferir, derivar, interpretar, justificar, explicar.

5. **Avaliar:** checar e criticar a informação com base em critérios.  
Verbos: testar, validar, verificar, controlar, monitorar, examinar, inspecionar, julgar, argumentar, justificar, defender, questionar, avaliar, recomendar.
6. **Criar:** generalizar, planejar e produzir algo novo a partir do conhecimento.  
Verbos: formular, construir, desenvolver, criar, projetar, produzir, compor, organizar, esquematizar, desenhar, estruturar, propor, estabelecer, idealizar, gerar, construir, inventar, compor, criar, desenvolver, originar.

**Sua tarefa:**

Com base no conceito fornecido, gere UCs complementares que explorem os seis níveis da Taxonomia de Bloom Revisada.

**Observações:**

Utilize os verbos da Taxonomia de Bloom Revisada para formular as UCs complementares.

- Assegure-se de que as UCs complementares estejam relacionadas ao conceito principal.

**Contexto:** (Necessário para garantir desambiguação entre conceitos similares)  
{ {resumo criado para à comunidade agrupada pelo algoritmo de leiden} }

**Conceito:**  
{ {UC avaliada} }

## C.2 Prompt para definição da dificuldade

O prompt a seguir foi elaborado para guiar o modelo de linguagem na avaliação da dificuldade das UCs. Ele inclui uma breve explicação sobre o que são UCs, os critérios a serem considerados na avaliação da dificuldade e instruções claras sobre a tarefa a ser realizada. A estrutura do prompt foi projetada para auxiliar o modelo de linguagem a fornecer avaliações consistentes com os objetivos do sistema. **Conforme discutido na seção sobre LLMs, a otimização do número de tokens não foi priorizada na elaboração deste prompt, visando, neste momento, apenas a validação da viabilidade da tarefa.**

PROMPT:

Você é um especialista em educação, capaz de analisar unidades de conhecimento (UCs) e determinar o seu nível de dificuldade.

Uma UC representa um conceito, habilidade ou informação que um aluno deve aprender. As UCs podem variar em complexidade, desde a simples memorização de um fato até a aplicação do conhecimento em situações complexas.

Sua tarefa é avaliar a dificuldade das UCs a seguir em uma escala de 0 a 100, onde 0 representa 'muito fácil' e 100 'muito difícil'. Leve em consideração os seguintes critérios ao determinar a dificuldade:

Complexidade do conceito: conceitos abstratos e multifacetados são mais difíceis de aprender do que conceitos simples e concretos.

- Habilidades cognitivas necessárias: UCs que exigem análise, síntese ou avaliação são mais difíceis do que aquelas que exigem apenas memorização ou compreensão.
- Conhecimento prévio necessário: UCs que exigem conhecimento prévio de outros conceitos são mais difíceis do que aquelas que podem ser aprendidas isoladamente.
- Clareza da linguagem: UCs escritas em linguagem clara e concisa são mais fáceis de aprender do que aquelas com linguagem complexa ou ambígua.

**UCs a serem avaliadas:**

{{ lista de UCs separadas por bullet points }}

**Resposta:**

Avalie a dificuldade de cada UC em uma escala de 0 a 100 e justifique sua resposta com base nos critérios fornecidos.

**Observações:**

Evite dar a mesma nota para todas as UCs.

- Utilize todo o espectro da escala de 0 a 100 para avaliar as UCs.
- Assegure-se de que suas avaliações sejam consistentes e imparciais.
- Lembre-se de que você é um especialista em educação e suas avaliações devem refletir sua expertise.

## APÊNDICE D – DIAGRAMAS DE SEQUÊNCIA DO CICLO DE VIDA DE BATCH LLM

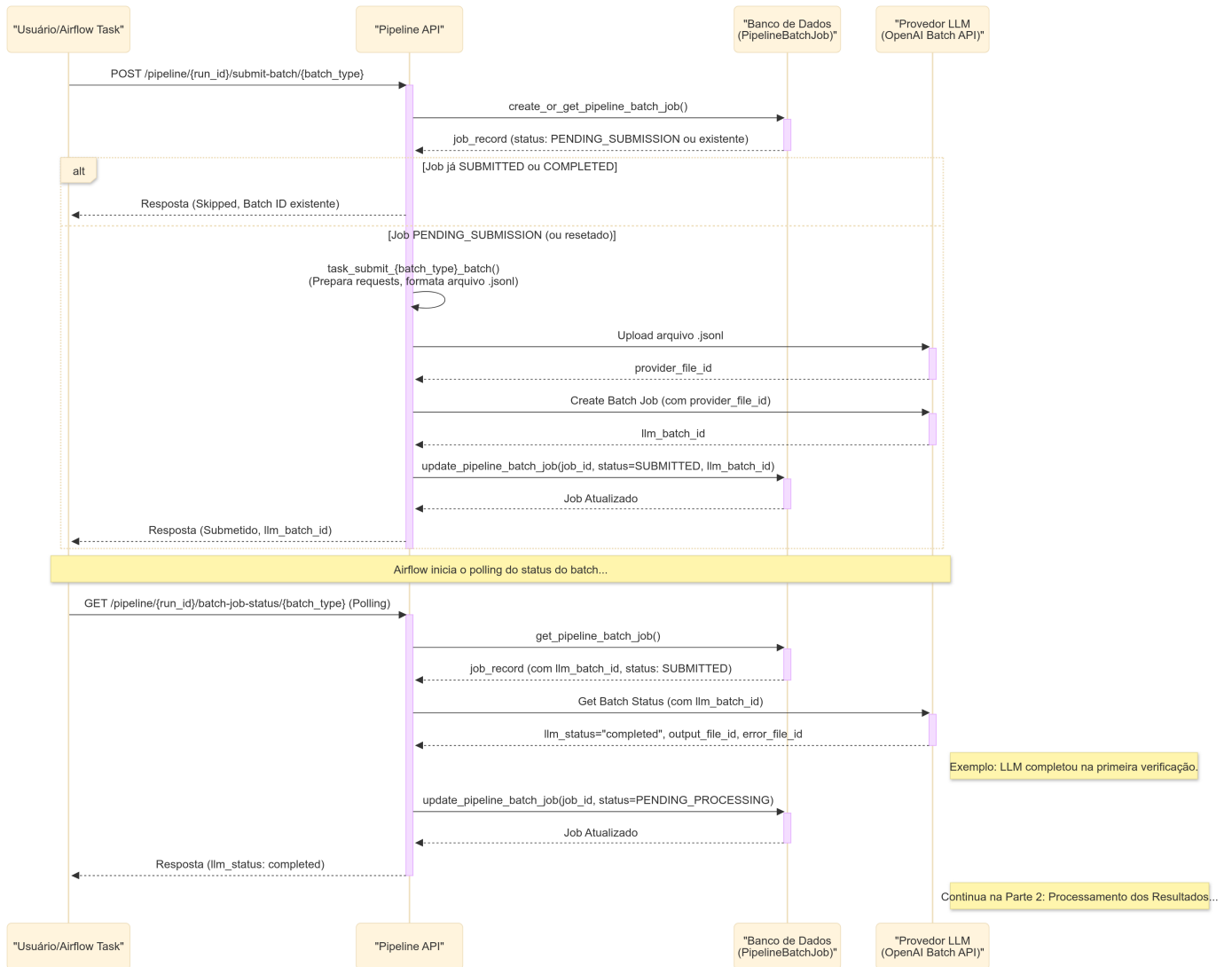


Figura D.1: Diagrama de Sequência do Ciclo de Vida de Batch LLM (Parte 1).

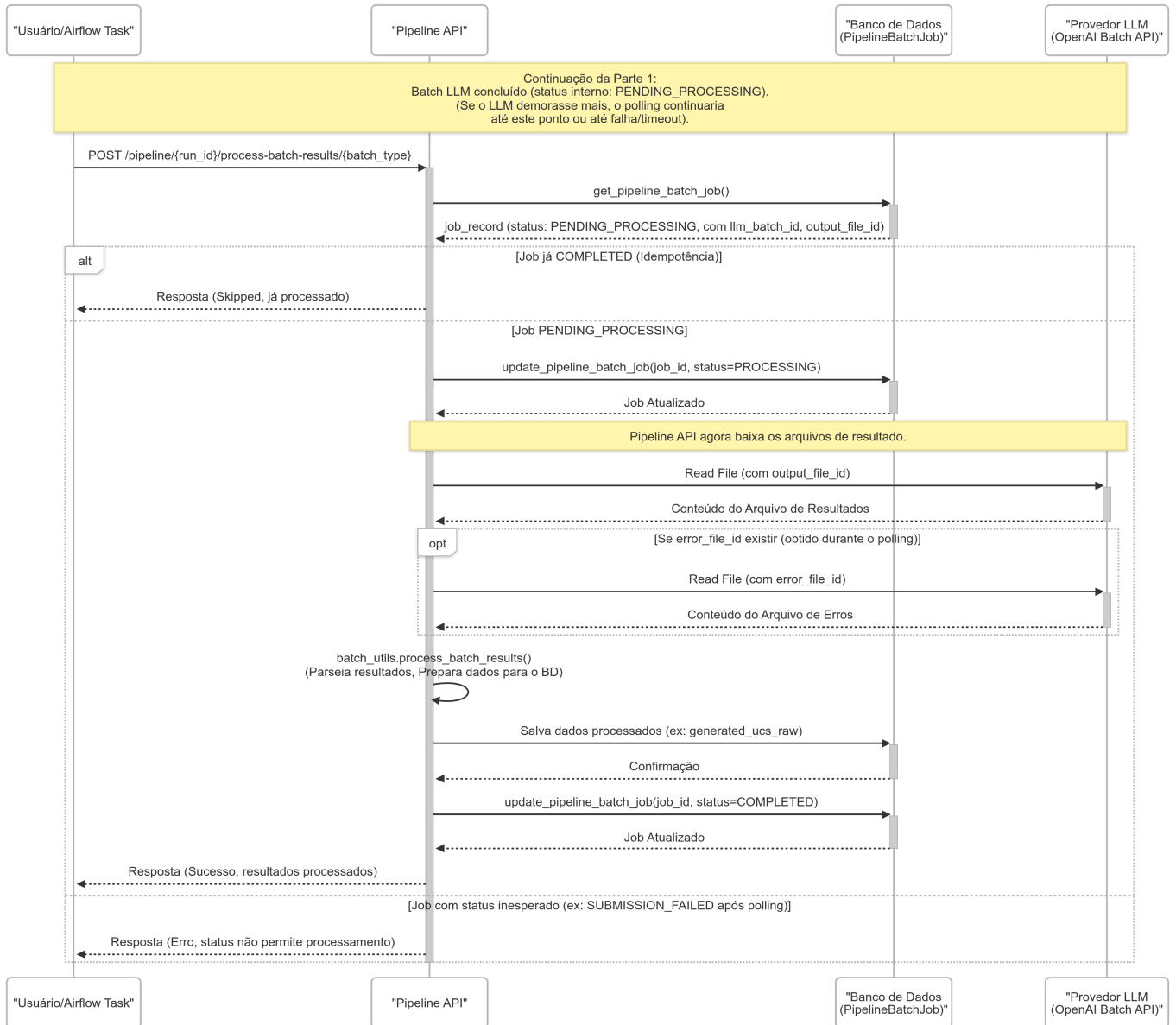


Figura D.2: Diagrama de Sequência do Ciclo de Vida de Batch LLM (Parte 2).



## APÊNDICE E – FLUXOGRAMA DO AGENDADOR DE DIFICULDADE DE ORIGENS

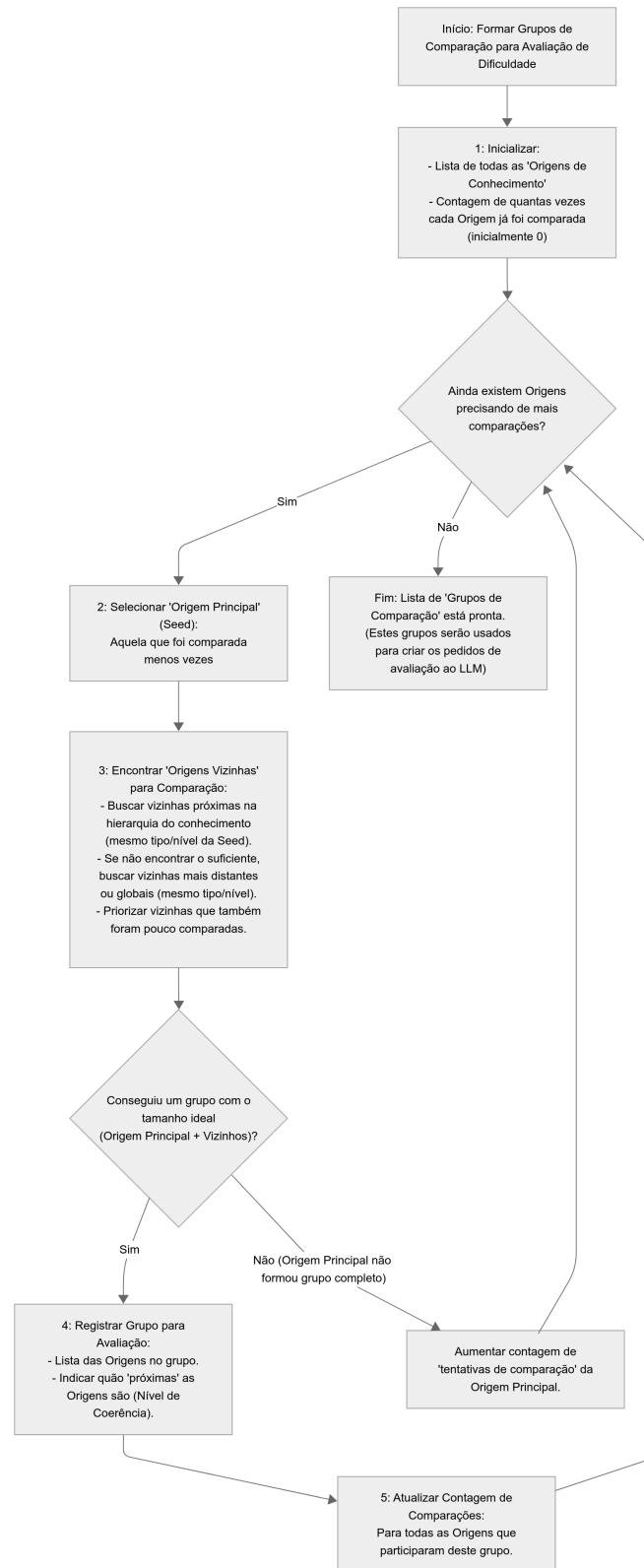


Figura E.1: Fluxograma da lógica do Agendador de Dificuldade de Origens.

## APÊNDICE F – EXEMPLOS UCS E ANÁLISE DE COESÃO TEMÁTICA

Este Apêndice apresenta exemplos de trilhas de Unidades de Conhecimento (UCs) geradas pelo sistema para diferentes Origens de Conhecimento, abrangendo os seis níveis da Taxonomia de Bloom Revisada. Os exemplos são categorizados em (1) trilhas com boa coesão temática na progressão e (2) trilhas que apresentam "saltos temáticos" ou menor coesão, ilustrando os desafios discutidos na Seção 5.3.1 e suas implicações para as relações REQUIRES (Seção 5.4.1). Para cada Origem, o contexto fornecido ao LLM é apresentado seguido pelas UCs.

### F.1 Trilhas com Boa Coesão Temática

Origem: REI DO GANGES (Entidade)

**Contexto:** Rei do Ganges é uma figura histórica ou mitológica associada ao rio Ganges, que é considerado sagrado na cultura indiana.

- **Lembrar:** Identificar as principais características do Rei do Ganges na mitologia indiana.
- **Entender:** Explicar a importância do rio Ganges na cultura indiana e como o Rei do Ganges está relacionado a essa reverência.
- **Aplicar:** Utilizar informações sobre o Rei do Ganges para criar uma apresentação sobre a mitologia indiana e sua relação com o rio Ganges.
- **Analisar:** Comparar as representações do Rei do Ganges em diferentes fontes históricas e mitológicas, identificando semelhanças e diferenças.
- **Avaliar:** Julgar a relevância do Rei do Ganges na cultura contemporânea indiana, considerando sua influência em rituais e festividades.
- **Criar:** Desenvolver uma narrativa fictícia que incorpora o Rei do Ganges como um personagem central, explorando suas interações com o rio e a cultura ao seu redor.

Origem: ASCENSÃO DE ROMA (Entidade)

**Contexto:** A Ascensão de Roma refere-se ao período em que Roma se tornou uma potência dominante na antiguidade, influenciando a política, cultura e sociedade.

- **Lembrar:** Identificar os principais eventos que marcaram a ascensão de Roma como potência na antiguidade.
- **Entender:** Explicar como a geografia de Roma contribuiu para seu crescimento e influência na política e cultura.

- **Aplicar:** Utilizar mapas históricos para demonstrar a expansão territorial de Roma durante sua ascensão.
- **Analisar:** Comparar as estratégias militares de Roma com as de outras civilizações contemporâneas durante o período de ascensão.
- **Avaliar:** Julgar a eficácia das políticas sociais e econômicas de Roma que contribuíram para sua ascensão como potência.
- **Criar:** Desenvolver uma apresentação multimídia que ilustre os fatores que levaram à ascensão de Roma e seu impacto duradouro na civilização ocidental.

#### Origem: GUERRAS MUNDIAIS (Entidade Abrangente)

**Contexto:** As Guerras Mundiais referem-se aos conflitos globais que ocorreram no século XX, impactando significativamente a história e a geopolítica mundial. Esses eventos históricos moldaram o século 20, caracterizando-se por uma sequência de conflitos entre nações e impérios. As duas guerras, a Primeira Guerra Mundial (1914-1918) e a Segunda Guerra Mundial (1939-1945), não apenas alteraram fronteiras e governos, mas também tiveram profundas consequências sociais, econômicas e culturais que ainda ressoam na sociedade contemporânea. As Guerras Mundiais são, portanto, marcos cruciais que definiram a dinâmica das relações internacionais e o desenvolvimento de ideologias políticas ao longo do século passado.

- **Lembrar:** Identificar as principais datas e eventos da Primeira e da Segunda Guerra Mundial.
- **Entender:** Explicar as causas e consequências das Guerras Mundiais, incluindo as mudanças geopolíticas resultantes.
- **Aplicar:** Utilizar mapas históricos para demonstrar as mudanças de fronteiras após as Guerras Mundiais.
- **Analisar:** Comparar e contrastar as estratégias militares utilizadas na Primeira e na Segunda Guerra Mundial.
- **Avaliar:** Julgar o impacto das Guerras Mundiais nas relações internacionais contemporâneas e nas ideologias políticas.
- **Criar:** Desenvolver uma apresentação multimídia que ilustre as consequências sociais e culturais das Guerras Mundiais.

### Origem: ABORÍGENES (Entidade)

**Contexto:** Os aborígenes são os povos nativos da Austrália, reconhecidos como os originários da região. Eles possuem um profundo conhecimento em botânica e zoologia, o que reflete sua conexão íntima com a terra e o meio ambiente. Essa sabedoria se manifesta em suas práticas de sobrevivência, que incluem técnicas de caça, coleta e uso do fogo, fundamentais para a adaptação e manutenção de suas comunidades ao longo dos milênios. A cultura aborígene é rica e complexa, englobando uma variedade de tradições, crenças e expressões artísticas que variam entre os diferentes grupos e regiões da Austrália. Os aborígenes têm uma relação espiritual profunda com a terra, que é central em suas narrativas e práticas culturais. Essa conexão é frequentemente expressa através de histórias, canções e danças que transmitem conhecimentos ancestrais e valores comunitários. Além de suas habilidades práticas, os aborígenes também são conhecidos por suas contribuições à compreensão da biodiversidade australiana, utilizando seu conhecimento tradicional para manejar os recursos naturais de maneira sustentável. A cultura aborígene, portanto, não é apenas um testemunho de sua história, mas também uma fonte de sabedoria que continua a influenciar a forma como a sociedade contemporânea interage com o meio ambiente. Em resumo, os aborígenes da Austrália são um povo com uma herança cultural rica, habilidades avançadas em botânica e zoologia, e práticas de sobrevivência que refletem uma profunda conexão com a terra.

- **Lembrar:** Identificar os principais grupos de aborígenes da Austrália e suas características culturais.
- **Entender:** Explicar a importância da conexão dos aborígenes com a terra e como isso se reflete em suas práticas culturais e de sobrevivência.
- **Aplicar:** Utilizar técnicas tradicionais de coleta e manejo de recursos naturais, inspiradas nas práticas aborígenes, em um projeto de sustentabilidade.
- **Analisar:** Comparar as práticas de sobrevivência dos aborígenes com as abordagens contemporâneas de manejo ambiental, identificando semelhanças e diferenças.
- **Avaliar:** Julgar a eficácia das técnicas de manejo sustentável dos aborígenes em relação às práticas modernas, considerando critérios de biodiversidade e preservação ambiental.
- **Criar:** Desenvolver um projeto que integre conhecimentos aborígenes sobre botânica e zoologia com práticas de conservação ambiental contemporâneas.

### Origem: Grande Buda de Nara and Emperor Shomu (Comunidade)

**Contexto:** The community centers around the Grande Buda de Nara, a monumental bronze statue symbolizing the influence of Buddhism in Japan, commissioned by Emperor Shomu

during a devastating epidemic. The relationships between the statue, the emperor, and the epidemic highlight the historical significance and cultural impact of these entities.

- **Lembrar:** Identificar o Grande Buda de Nara e nomear o imperador que o comissionou.
- **Entender:** Explicar a relação entre o Grande Buda de Nara, o imperador Shomu e a epidemia que afetou a comunidade.
- **Aplicar:** Utilizar informações históricas sobre o Grande Buda de Nara para discutir seu impacto na cultura japonesa contemporânea.
- **Analisar:** Comparar as motivações de Emperor Shomu para a construção do Grande Buda de Nara com outras iniciativas de liderança durante crises históricas.
- **Avaliar:** Julgar a eficácia do Grande Buda de Nara como símbolo de esperança durante a epidemia, com base em relatos históricos.
- **Criar:** Desenvolver uma proposta para um projeto educacional que explore a história do Grande Buda de Nara e seu significado cultural.

## F.2 Trilhas com Potenciais Saltos Temáticos ou Menor Coesão

Origem: CHINA (Entidade Muito Ampla)

**Contexto:** China is a vast and historically rich country located in East Asia, known for its significant contributions to global civilization, culture, and technology. With a history that dates back approximately 1.8 million years, China has been home to one of the world's oldest continuous civilizations. Throughout its history, China has experienced remarkable advancements in various fields, including agriculture, metallurgy, and communication, with notable inventions such as paper and gunpowder that greatly influenced the development of Europe during the Middle Ages.

The ancient Chinese society was characterized by a complex social structure, including the presence of millions of slaves and practices of slavery that impacted other cultures. Despite facing numerous challenges, including invasions, epidemics, and natural disasters, China maintained its independence, although it was weakened by wars and territorial concessions over the years. The Mongol invasions during the 12th and 13th centuries were particularly significant, leading to substantial changes in the region.

China's cultural landscape has been shaped by various religious influences, including Buddhism, which made significant strides in the country and coexisted with Confucianism and Taoism. The Silk Road played a crucial role in the exchange of ideas and goods, with China being a primary supplier of silk and other commodities, such as tea and porcelain, which were highly sought

after in Europe. The country also served as a source of new varieties of flowers and materials that were imported to Europe.

In the 18th century, China expanded its territory and became a dominant force in regional trade, although it later faced challenges from European powers during the early modern period. The Opium Wars and subsequent foreign interventions led to a period of humiliation and decline in China's influence. The revolution in 1949 marked a significant turning point, as the country sought to dismantle ancient practices and modernize its governance.

China's population has been significantly impacted by its historical practices, including the implementation of a one-child policy in the late 20th century to control rapid population growth. The country has also faced environmental challenges, such as land use intensity affecting bird populations and the impact of natural disasters on agriculture.

Overall, China is recognized for its rich cultural heritage, technological advancements, and historical significance, having played a pivotal role in shaping global trade and cultural exchanges throughout history. Its journey from an ancient civilization to a modern republic reflects a complex interplay of resilience, innovation, and adaptation in the face of numerous challenges.

- **Lembrar:** Identificar as principais invenções da China antiga, como papel e pólvora.
- **Entender:** Explicar a importância da Rota da Seda para o comércio e a troca cultural entre a China e a Europa.
- **Aplicar:** Utilizar informações sobre a história da China para analisar como as invasões mongóis influenciaram a estrutura social do país.
- **Analisar:** Comparar as práticas religiosas na China antiga, como o Budismo, Confucionismo e Taoísmo, e discutir suas interações.
- **Avaliar:** Examinar os efeitos da política do filho único na sociedade e economia chinesa e justificar sua implementação.
- **Criar:** Desenvolver um projeto que proponha soluções para os desafios ambientais enfrentados pela China, considerando sua história e cultura.

Origem: HUMANITY (Entidade Extremamente Ampla)

**Contexto:** Humanity encompasses the collective existence and evolution of human beings, highlighting significant periods of development throughout history. It refers to the entire human race, emphasizing the shared experiences and cultural practices that unite individuals across diverse regions. This concept reflects the interconnectedness of people, illustrating how cultural, social, and historical factors shape the human experience. Humanity is not only about biological and evolutionary aspects but also about the rich tapestry of traditions, beliefs, and interactions that define human societies. Through this lens, humanity can be understood as a dynamic and evolving entity, marked by both commonalities and unique differences among various groups.

- **Lembrar:** Identificar os principais períodos de desenvolvimento da humanidade ao longo da história.
- **Entender:** Explicar como as experiências culturais e sociais moldam a percepção da humanidade em diferentes regiões.
- **Aplicar:** Utilizar exemplos históricos para demonstrar a interconexão entre diferentes culturas e sociedades humanas.
- **Analisar:** Comparar e contrastar as tradições e crenças de diferentes grupos humanos para identificar semelhanças e diferenças.
- **Avaliar:** Julgar a importância das interações sociais na evolução da humanidade e seu impacto nas sociedades contemporâneas.
- **Criar:** Desenvolver um projeto que ilustre a evolução da humanidade, incorporando elementos culturais, sociais e históricos de diversas regiões.

Origem: 20TH CENTURY (Entidade Ampla com Múltiplos Eixos)

**Contexto:** The 20th century was a transformative era characterized by profound political, social, and cultural changes. One of the most notable developments was the decline of communism, particularly highlighted by its collapse in Russia and Eastern Europe. This marked a significant historical shift, reshaping global politics and ideologies as nations transitioned away from communist regimes towards various forms of governance, including democracy and capitalism.

In addition to political upheaval, the 20th century also presented significant challenges for religion. There was a notable decline in church attendance across Europe and the United States, reflecting broader societal shifts towards secularism and changing cultural values. This decline in religious participation indicated a move away from traditional beliefs and practices, as individuals increasingly sought alternative sources of meaning and community.

Overall, the 20th century was defined by its dynamic interplay of political ideologies and religious transformations, setting the stage for the complexities of the contemporary world.

- **Lembrar:** Identificar os principais eventos políticos que marcaram o século XX, como a queda do comunismo na Rússia e na Europa Oriental.
- **Entender:** Explicar como a transição de regimes comunistas para democracias e sistemas capitalistas afetou a política global no século XX.
- **Aplicar:** Analisar a influência das mudanças sociais do século XX na diminuição da participação religiosa em países ocidentais.

- **Analisar:** Comparar as diferenças entre o comunismo e a democracia, destacando como essas ideologias moldaram sociedades no século XX.
- **Avaliar:** Julgar o impacto da secularização no século XX sobre as tradições religiosas e a formação de novas comunidades de significado.
- **Criar:** Desenvolver um projeto que proponha soluções para os desafios sociais contemporâneos, inspirando-se nas transformações políticas e religiosas do século XX.

Origem: China and Its Historical Influences (Comunidade)

**Contexto:** The community centers around China, a historically rich nation with significant contributions to global civilization, and its key entities such as Confucius, Zheng He, and the Grande Pagode do Ganso Selvagem. These entities are interconnected through cultural, philosophical, and technological advancements that have shaped not only China but also influenced global trade and cultural exchanges.

- **Lembrar:** Identificar os principais influenciadores da história da China, como Confúcio e Zheng He.
- **Entender:** Explicar a importância do Grande Pagode do Ganso Selvagem na cultura e história chinesa.
- **Aplicar:** Utilizar os ensinamentos de Confúcio para analisar situações éticas contemporâneas na sociedade.
- **Analisar:** Comparar as contribuições de Zheng He para o comércio global com outros exploradores da época.
- **Avaliar:** Julgar a relevância das filosofias chinesas na formação de relações interculturais modernas.
- **Criar:** Desenvolver um projeto que propõe um intercâmbio cultural entre a China e outra nação, inspirado nas trocas históricas promovidas por Zheng He.

Origem: Cultural and Historical Dynamics of Europe (Comunidade)

**Contexto:** The community encompasses various entities that represent significant cultural, historical, and agricultural aspects of Europe, including Serbia, the Cold War, and key agricultural crops like potatoes and corn. These entities are interconnected through historical events, agricultural practices, and cultural developments that have shaped the European landscape over centuries.



- **Lembrar:** Identificar os principais países da Europa e suas contribuições culturais e históricas, como a Sérvia e os impactos da Guerra Fria.
- **Entender:** Explicar como a Guerra Fria influenciou as dinâmicas culturais e históricas na Europa, incluindo a relação entre os países do Leste e do Ocidente.
- **Aplicar:** Utilizar mapas históricos para demonstrar a evolução das fronteiras e a migração de culturas na Europa ao longo dos séculos.
- **Analisar:** Comparar e contrastar as práticas agrícolas de culturas como batatas e milho em diferentes regiões da Europa e seu impacto nas sociedades locais.
- **Avaliar:** Julgar a importância das dinâmicas culturais e históricas na formação da identidade europeia contemporânea, considerando eventos como a Guerra Fria.
- **Criar:** Desenvolver um projeto que proponha uma nova narrativa cultural que integre as histórias da Sérvia e de outros países europeus, destacando suas interconexões.

## APÊNDICE G – EXEMPLOS DE AVALIAÇÃO DE DIFICULDADE

Este Apêndice apresenta exemplos de Unidades de Conhecimento (UCs) que receberam os *scores* de dificuldade mais baixos e mais altos atribuídos pelo modelo de linguagem, dentro de cada nível da Taxonomia de Bloom. Para cada UC, são apresentados o texto da UC e seu respectivo *score* de dificuldade (escala 0-100).

### G.1 Nível: Lembrar

UCs com Menor Dificuldade (Lembrar)

- **UC:** Identificar Tóquio como a capital do Japão. (*Score:* 14)
- **UC:** Identificar as partes principais de um microscópio, como ocular, objetiva e platina. (*Score:* 17)
- **UC:** Identificar a data em que Neil Armstrong caminhou na Lua. (*Score:* 17)

UCs com Maior Dificuldade (Lembrar)

- **UC:** Identificar os principais fatores que influenciam a dinâmica populacional, incluindo tecnologia, natureza e poluição. (*Score:* 75)
- **UC:** Identificar os principais eventos históricos relacionados à escravidão em Cuba, Brasil e Mauritânia. (*Score:* 73)
- **UC:** Identificar os principais eventos e características da Idade Média, como o feudalismo e a influência da Igreja Católica. (*Score:* 72)

### G.2 Nível: Entender

UCs com Menor Dificuldade (Entender)

- **UC:** Explicar a diferença entre o presente simples e o presente contínuo em inglês. (*Score:* 37)
- **UC:** Explicar como a primavera está relacionada ao renascimento e ao crescimento das plantas. (*Score:* 37)
- **UC:** Explicar a importância da castanheira como fonte de alimento para os antigos japoneses. (*Score:* 40)

#### UCs com Maior Dificuldade (Entender)

- **UC:** Explicar as causas e consequências do Holocausto, incluindo o papel do regime nazista e a ideologia antissemita. (*Score:* 84)
- **UC:** Explicar as causas e consequências da Rebelião Taiping durante a Dinastia Qing. (*Score:* 82)
- **UC:** Explicar as causas e consequências das Guerras Mundiais, incluindo as mudanças geopolíticas resultantes. (*Score:* 82)

### G.3 Nível: Aplicar

#### UCs com Menor Dificuldade (Aplicar)

- **UC:** Utilizar um telefone para realizar chamadas e enviar mensagens de texto em diferentes contextos. (*Score:* 23)
- **UC:** Preparar uma receita simples de mingau de aveia utilizando água ou leite. (*Score:* 27)
- **UC:** Utilizar mapas para localizar Fiji e suas ilhas vizinhas no Oceano Pacífico. (*Score:* 30)

#### UCs com Maior Dificuldade (Aplicar)

- **UC:** Utilizar dados sobre a mortalidade infantil para simular cenários de sobrevivência de bebês gêmeos em uma tribo nômade. (*Score:* 85)
- **UC:** Utilizar a Teoria da Relatividade para resolver problemas relacionados à velocidade da luz e ao tempo. (*Score:* 85)
- **UC:** Implementar os princípios de redistribuição de terras e nacionalização da indústria em um estudo de caso sobre a transformação econômica da Rússia pós-revolução. (*Score:* 85)

### G.4 Nível: Analisar

#### UCs com Menor Dificuldade (Analisar)

- **UC:** Comparar e contrastar as estruturas de frases afirmativas e negativas em inglês. (*Score:* 40)

- **UC:** Comparar as diferentes atividades ao ar livre disponíveis nos Alpes Suíços, como esqui e caminhadas. (*Score:* 40)
- **UC:** Comparar as características do verão com as de outras estações do ano, como primavera e outono. (*Score:* 40)

#### UCs com Maior Dificuldade (Analisar)

- **UC:** Comparar e contrastar as representações de mulheres em diferentes culturas e mídias ao longo da história. (*Score:* 85)
- **UC:** Comparar e contrastar as estratégias militares utilizadas pelas potências do Eixo e pelos Aliados durante a Segunda Guerra Mundial. (*Score:* 85)
- **UC:** Comparar o tratamento de escravos africanos nos Estados Unidos com o de outros países, como o Brasil, e discutir as implicações sociais e políticas. (*Score:* 83)

### G.5 Nível: Avaliar

#### UCs com Menor Dificuldade (Avaliar)

- **UC:** Justificar a escolha de um determinado tempo verbal em uma redação em inglês. (*Score:* 40)
- **UC:** Julgar a qualidade de uma cerveja com base em critérios como aroma, sabor e aparência. (*Score:* 47)
- **UC:** Justificar a escolha de um determinado tipo de bola para jogar boliche de grama em diferentes condições de gramado. (*Score:* 47)

#### UCs com Maior Dificuldade (Avaliar)

- **UC:** Julgar o impacto das Guerras Mundiais nas relações internacionais contemporâneas e nas ideologias políticas. (*Score:* 85)
- **UC:** Julgar a eficácia das técnicas de cultivo e domesticação na Líbia em comparação com as práticas contemporâneas, argumentando sobre os impactos sociais e econômicos. (*Score:* 85)
- **UC:** Julgar a influência de Roma na cultura ocidental contemporânea, considerando aspectos políticos, religiosos e arquitetônicos. (*Score:* 85)

## G.6 Nível: Criar

### UCs com Menor Dificuldade (Criar)

- **UC:** Desenvolver um diálogo original entre dois personagens utilizando diferentes tempos verbais em inglês. (*Score:* 40)
- **UC:** Desenvolver uma nova receita de mingau de aveia que incorpore ingredientes locais e sazonais. (*Score:* 40)
- **UC:** Desenvolver uma nova receita de sobremesa utilizando iogurte como ingrediente principal. (*Score:* 43)

### UCs com Maior Dificuldade (Criar)

- **UC:** Desenvolver um novo teorema geométrico inspirado nos postulados de Euclides. (*Score:* 90)
- **UC:** Desenvolver um projeto de conservação que proponha medidas para proteger a biodiversidade da ilha de Réunion, incluindo a reintrodução de espécies extintas. (*Score:* 90)
- **UC:** Desenvolver um projeto que proponha soluções sustentáveis para a gestão da água em rios asiáticos afetados pelo degelo. (*Score:* 88)

## APÊNDICE H – DETALHAMENTO DA ESTRUTURA HIERÁRQUICA

Este Apêndice apresenta um fragmento da estrutura hierárquica de *KnowledgeUnitOrigin* (Origens de Unidade de Conhecimento) gerada pelo sistema, ilustrando como *community\_reports* (relatórios de comunidade) e *entities* (entidades) se relacionam para cobrir diferentes facetas de um domínio de conhecimento em variados níveis de granularidade. Para cada Origem, são fornecidos seu tipo, título, nível hierárquico (para comunidades, onde 0 é o nível mais alto/raiz), um resumo do contexto utilizado para a geração das UCs, e exemplos de Unidades de Conhecimento (UCs) com seus respectivos níveis da Taxonomia de Bloom e *scores* de dificuldade atribuídos.

### H.1 Comunidade Raiz (Nível 0): Tema Amplo – Igreja e Renascimento

**Tipo:** *community\_report*

**Título:** Igreja e Renascimento: A Influência Cultural e Religiosa

**Nível da Comunidade:** 0

**Contexto:** *A comunidade é centrada na Igreja, que desempenhou um papel crucial na formação do cristianismo e na cultura ocidental, especialmente durante o Renascimento. As entidades inter-relacionadas, como o Renascimento, a cidade de Florença, e figuras como Galileu e Leonardo da Vinci, refletem a complexidade e a interconexão entre religião, arte e ciência na história.*

Exemplos de UCs Geradas:

- **Lembrar:** Identificar as principais figuras do Renascimento, como Galileu e Leonardo da Vinci, e suas contribuições para a cultura e ciência. (*Score: 45*)
- **Entender:** Explicar como a Igreja influenciou a arte e a ciência durante o Renascimento, destacando a relação entre religião e criatividade. (*Score: 65*)
- **Avaliar:** Julgar a importância da Igreja como patrona das artes durante o Renascimento e seu impacto na evolução cultural da época. (*Score: 70*)

Comunidade Filha (Nível 1): Cristianismo Primitivo

**Tipo:** *community\_report*

**Título:** Antioquia and Early Christianity

**Nível da Comunidade:** 1 (Filha de "Igreja e Renascimento")

**Contexto:** *The community centers around Antioquia, a significant location in the early development of Christianity, where key figures such as Cristo and Paulo played pivotal roles in*

*shaping the faith. The relationships among these entities highlight the theological debates and transformations that occurred during the formative years of the Christian Church.*

Exemplos de UCs Geradas:

- **Lembrar:** Identificar os principais personagens da comunidade cristã em Antioquia, como Cristo e Paulo. (Score: 36)
- **Analisar:** Comparar as diferentes abordagens teológicas discutidas por Cristo e Paulo na formação da doutrina cristã em Antioquia. (Score: 75)
- **Criar:** Desenvolver um projeto que proponha uma nova interpretação das interações entre os líderes cristãos em Antioquia e seu impacto na igreja primitiva. (Score: 76)

Entidade (Componente de "Antioquia and Early Christianity"): ANTIOQUIA

**Tipo:** *entity*

**Contexto:** *Antioquia is a city in southern Turkey where the early debate about who could become a member of the Christian Church took place.*

**Exemplos de UCs Geradas:**

- **Lembrar:** Identificar os principais eventos que ocorreram em Antioquia relacionados ao início da Igreja Cristã. (Score: 45)
- **Aplicar:** Utilizar informações históricas sobre Antioquia para descrever como as decisões tomadas lá influenciaram a expansão do Cristianismo. (Score: 65)

Entidade (Componente de "Antioquia and Early Christianity"): CRISTO

**Tipo:** *entity*

**Contexto:** *Cristo é uma figura central no cristianismo, considerado o filho de Deus, cuja ascensão aos céus é um evento fundamental na fé cristã.*

**Exemplos de UCs Geradas:**

- **Entender:** Explicar o significado da ascensão de Cristo e sua importância na fé cristã. (Score: 63)
- **Avaliar:** Examinar e justificar a relevância da figura de Cristo na sociedade contemporânea e em debates éticos. (Score: 77)

## H.2 Comunidade Filha (Nível 1): Renascimento Científico – Ligado a "Igreja e Renascimento"

**Tipo:** *community\_report*

**Título:** Florence and the Scientific Renaissance

**Nível da Comunidade:** 1 (Filha de "Igreja e Renascimento")

**Contexto:** *The community centers around Florence, a pivotal city in the Renaissance, and its notable figures such as Galileo Galilei and Carolus Linnaeus. The relationships among these entities highlight Florence's historical significance in art, science, and trade, particularly during the Mongol era.*

Exemplos de UCs Geradas:

- **Lembrar:** Identificar os principais cientistas associados à Renascença em Florença, como Galileo Galilei e Carolus Linnaeus. (*Score: 30*)
- **Entender:** Explicar a importância de Florença como um centro de arte e ciência durante a Renascença, destacando suas contribuições para o comércio e a cultura. (*Score: 70*)

Entidade (Componente de "Florence and the Scientific Renaissance"): GALILEU

**Tipo:** *entity*

**Contexto:** *Galileu foi um cientista que melhorou o telescópio e fez descobertas significativas sobre a Lua e o Sol, desafiando as crenças da época sobre o universo.*

**Exemplos de UCs Geradas:**

- **Analisar:** Comparar as teorias de Galileu com as crenças predominantes sobre o universo antes de suas descobertas. (*Score: 80*)
- **Criar:** Desenvolver um projeto que utilize princípios de astronomia baseados nas descobertas de Galileu, propondo novas formas de observação do céu. (*Score: 65*)

Entidade (Componente de "Florence and the Scientific Renaissance"): CAROLUS LINNAEUS

**Tipo:** *entity*

**Contexto:** *Carolus Linnaeus, também conhecido como Lineu, foi um botânico e físico sueco que desenvolveu um método de classificação de plantas e animais.*

**Exemplos de UCs Geradas:**

- **Aplicar:** Utilizar o sistema de classificação de Linnaeus para categorizar diferentes espécies de plantas e animais em um projeto de pesquisa. (*Score: 65*)



- **Avaliar:** Julgar a eficácia do sistema de nomenclatura binomial de Linnaeus em relação a sistemas de classificação mais modernos, argumentando a favor ou contra sua utilização. (Score: 75)

Entidade (Componente de "Florence and the Scientific Renaissance"): FLORENÇA

**Tipo:** *entity*

**Contexto:** *Florence, known as "FLORENÇA" in Portuguese, is a prominent city in Italy celebrated as the birthplace of the Renaissance, a pivotal period in art and culture. This historical city is renowned for its significant contributions to the arts, architecture, and humanism, making it a central hub for cultural development during the Renaissance era.*

*In addition to its artistic legacy, Florence has a complex historical background that includes its involvement in the slave market during the Mongol era. This aspect of its history highlights the city's multifaceted role in trade and commerce, reflecting the broader socio-economic dynamics of the time.*

*Florence is also notable for its connection to the famous scientist Galileo Galilei, who spent the final years of his life under house arrest in the city. This period was marked by Galileo's significant contributions to science, despite the challenges he faced due to the prevailing religious and political climate.*

*Overall, Florence stands out not only for its artistic and cultural heritage but also for its intricate historical narratives that encompass both its achievements and its darker chapters.*

#### **Exemplos de UCs Geradas:**

- **Aplicar:** Utilizar informações sobre a história de Florença para criar uma linha do tempo dos eventos significativos do Renascimento. (Score: 61)
- **Criar:** Desenvolver um projeto multimídia que ilustre a evolução cultural de Florença, incluindo suas contribuições artísticas e os aspectos sombrios de sua história. (Score: 71)

## **APÊNDICE I – CÓDIGO-FONTE**

O projeto de software desenvolvido como parte deste trabalho, incluindo a API, os scripts de orquestração do pipeline e os artefatos para a configuração do ambiente, foi disponibilizado como código-fonte aberto sob a licença MIT.

O acesso ao projeto pode ser feito através do seguinte link:

- <https://github.com/Andresho/TaxoGrafo>