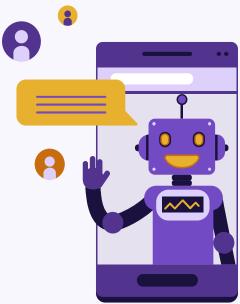


Mineração de dados - GRUPO 1

André Almeida, PG57866 Cyrill Koch, e12361 Francisco Lameirão, PG57542 Matilde Fernandes, PG57588 Maya Gomes, PG57891 Rui Cerqueira, PG57902







(role C)

Fontes específicas ao tema	EatRight, DietaryGuidelines (para além das fontes referidas pela role C)	
Key Words	Diet terms, eating habits terms, food terms, nutrients terms	
Query	"Eating habits "AND "nutrition" AND ("health outcomes" OR "dietary patterns")	
Obtenção dos dados	Web Scraping, Selenium, Keywords por Domínio	
Processamento	SpaCy para tokenização	
Chunking Semântico	Divisão preservando contexto e respeitando limites de tokens	
Enriquecimento de Metadados	Adição de informações estruturais (hierarchy, indexação para RAG)	

## Exemplos de documentos recolhidos





	Non Reasoning Model (Llama 3.3 70B)	Reasoning Model (Deepseek R1 70B)
Dietas	7.5	8.4
Hábitos Alimentares	6.7	8.4
Alimentos Específicos	8.2	8.4
Nutrientes	6.7	7.4
Média Global	7.3	8.2

Respostas classificadas usando ChatGPT 4.0, Claude Sonnet 4, Gemini 2.5 Falsh e Grok 3



Este conjunto de parâmetros define o comportamento do agente na seleção e utilização dos dados.

- reasoningModel Ativa modelo com capacidade de raciocínio;
- **topK** Número máximo de *chunks* recuperados;
- targetThreshold Valor ideal a partir do qual um chunk é considerado fortemente relevante;
- **minimumThreshold** Valor mínimo para considerar fonte complementar;
- maxHierarchyLevel Nível máximo de hierarquia nos dados.

Os parâmetros, exceto **reasoningModel** e **maxHierarchyLevel**, foram ajustados empiricamente, através de sucessivos testes e afinações, com o objectivo de obter melhores resultados observáveis na geração de respostas.

```
"reasoningModel": true,
"topK": 10.
"targetThreshold": 0.6,
"minimumThreshold": 0.2,
"maxHierarchvLevel": 4
```



Mineração de dados - GRUPO 1

André Almeida, PG57866 Cyrill Koch, e12361 Francisco Lameirão, PG57542 Matilde Fernandes, PG57588 Maya Gomes, PG57891 Rui Cerqueira, PG57902

