

## Relatório 6 - Prática: Embedding (II)

Suele Sousa

### Descrição da atividade

Nessa atividade devemos aprender a transformar dados para vetores, utilizando o componente Qdrant do Langflow.

Na leitura inicial, “How To Get Started With Qdrant Cloud”, tanto no texto como no vídeo dentro do site, é ensinado como criar uma cloud para armazenar os **embeddings** criados. O exemplo abaixo é o exemplo da coleção criada no tutorial.

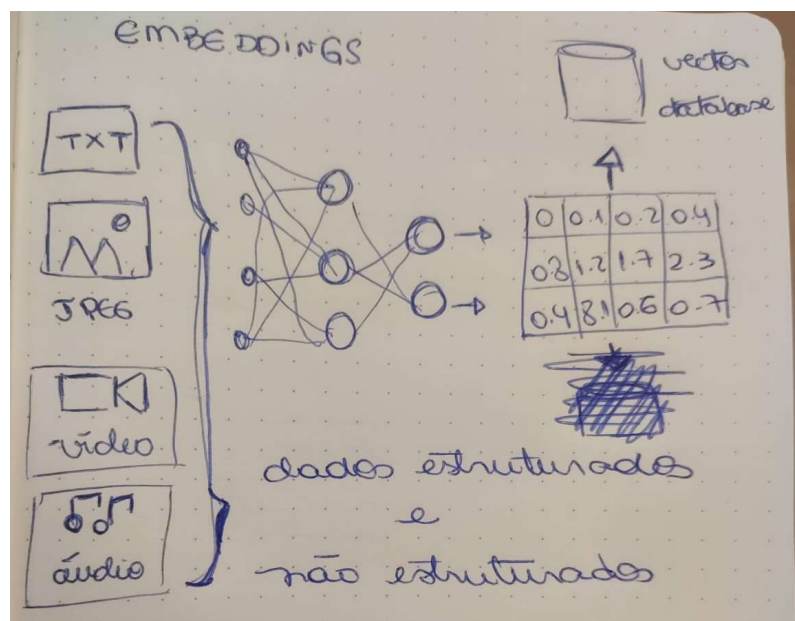
### Collections

UPLOAD SNAPSHOT

<input type="text" value="Search Collection"/>						
Name	Status	Points (Approx)	Segments	Shards	Vectors Configuration (Name, Size, Distance)	Actions
star_charts	● green	5	2	1	default 4 Dot	⋮

O que é um *embedding*?

Embedding basicamente é transformar uma imagem, texto, voz em uma lista de números, por exemplo:



Armazenar conteúdo em forma de *embeddings* é amplamente utilizado no contexto de IA, como por exemplo em NLP (*Natural Language Processing*), onde cada palavra, texto

ou imagem é representado por um token e, posteriormente, convertido em um vetor numérico que captura seu significado ou contexto. Em modelos de LLMs, por exemplo, o modelo gera uma resposta com base em vetores semelhantes àqueles associados a perguntas parecidas.

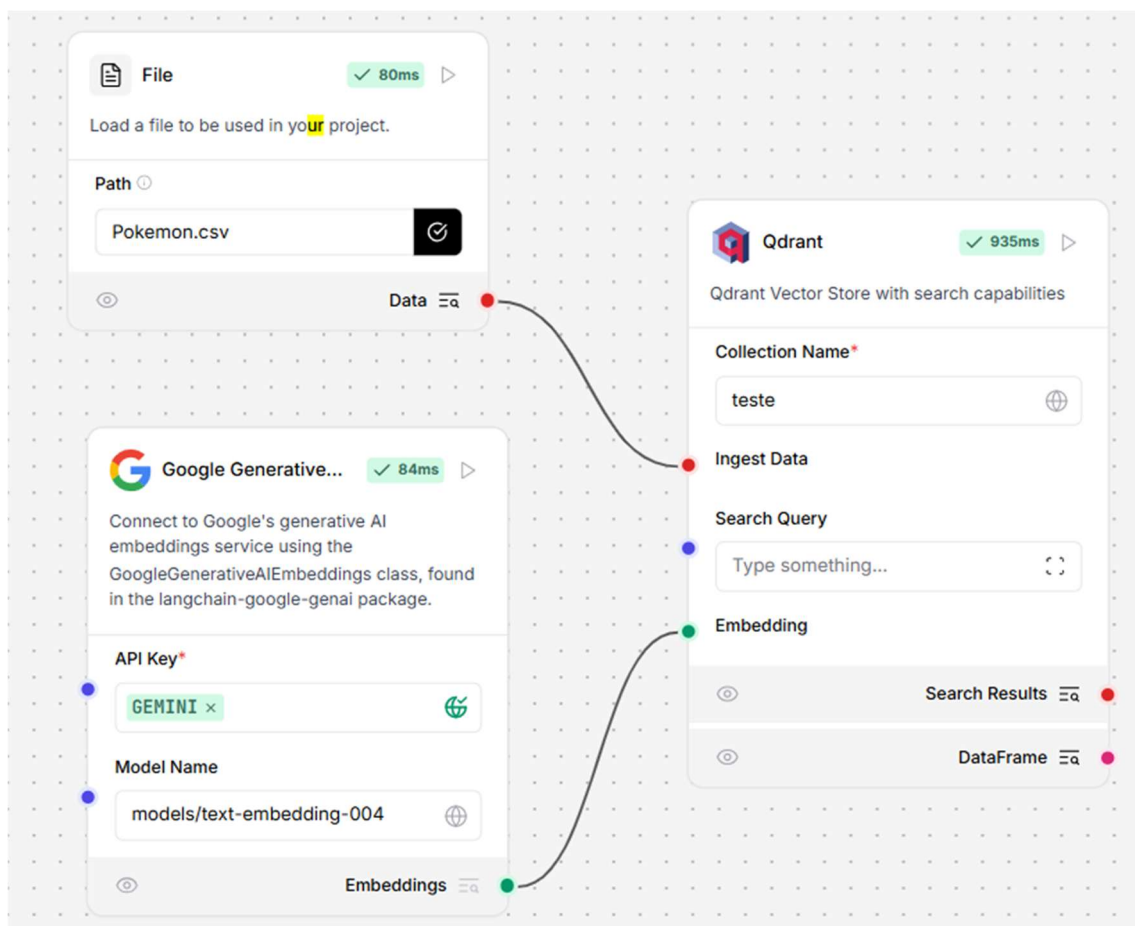
**Utilizei o Langflow via site, pois não consegui iniciar o localhost, havia tido o mesmo problema em um dos relatórios anteriores.**

**Os componentes utilizados foram:**

**File:** Pra subir um arquivo direto da máquina.

**GEMINI Embeddings:** Para gerar o embbeding do arquivo que estamos subindo.

**Qdrant:** Para subir os dados na cloud do qdrant.



Na primeira tentativa, os dados subiram todos juntos, como uma única coluna com apenas uma linha de dados.

drant

teste

POINTS INFO SEARCH QUALITY SNAPSHOTS VISUALIZE GRAPH

Find similar by ID or filter by payload key value pair. Example: name: John Doe, age: 25, id: c0847827-d005-4e46-b328-88772373d2d, id: 1234567890

Point 3ba2de1f-3ebc-4125-a489-57260e1990a3

Payload

page\_content

```
# Name, Type 1, Type 2, Total HP, Attack, Defense, Sp. Atk, Sp. Def, Speed, Generation, Legendary 1, Bulbasaur, Grass, Poison, 318, 45, 49, 65, 65, 45, 1, False
2, Ivysaur, Grass, Poison, 405, 60, 62, 63, 80, 80, 60, 1, False, 3, Venusaur, Grass, Poison, 525, 80, 82, 83, 100, 100, 80, 1, False, 3, VenusaurMega
Venusaur, Grass, Poison, 625, 80, 100, 123, 122, 120, 80, 1, False, 4, Charmander, Fire, 309, 39, 52, 43, 60, 50, 65, 1, False
5, Charmeleon, Fire, 405, 58, 64, 58, 80, 65, 80, 1, False, 6, Charizard, Fire, Flying, 534, 78, 84, 78, 109, 85, 100, 1, False, 6, CharizardMega Charizard
X, Fire, Dragon, 634, 78, 130, 111, 130, 85, 100, 1, False, 6, CharizardMega Charizard Y, Fire, Flying, 634, 78, 104, 78, 159, 115, 100, 1, False
7, Squirtle, Water, 314, 44, 48, 65, 50, 64, 43, 1, False, 8, Wartortle, Water, 405, 59, 63, 80, 65, 80, 58, 1, False, 9, Blastoise, Water, 530, 79, 83, 100, 85, 105, 78, 1, False
9, BlastoiseMega Blastoise, Water, 630, 79, 103, 120, 135, 115, 78, 1, False, 10, Caterpie, Bug, 195, 45, 30, 35, 20, 20, 45, 1, False
11, Metapod, Bug, 205, 50, 20, 55, 25, 25, 30, 1, False, 12, Butterfree, Bug, Flying, 395, 60, 45, 50, 90, 80, 70, 1, False
13, Weedle, Bug, Poison, 195, 40, 35, 30, 20, 20, 50, 1, False, 14, Kakuna, Bug, Poison, 205, 45, 25, 50, 25, 25, 35, 1, False
15, Beedrill, Bug, Poison, 395, 65, 90, 40, 45, 80, 75, 1, False, 15, BeedrillMega Beedrill, Bug, Poison, 495, 65, 150, 40, 15, 80, 145, 1, False
16, Pidgey, Normal, Flying, 251, 40, 45, 40, 35, 35, 56, 1, False, 17, Pidgeotto Normal Flying, 349, 63, 60, 55, 50, 50, 71, 1, False
18, Pidgeot, Normal, Flying, 479, 83, 80, 75, 70, 101, 1, False, 18, PidgeotMega Pidgeot, Normal, Flying, 579, 83, 80, 80, 135, 80, 121, 1, False
```

Isso gerava um erro de visualização, pois é como existisse um único ponto, conforme imagem abaixo:

Visualization Unsuccessful, error: cannot perform TSNE on single point DISMISS

```
1
2
3 // Try me!
4
5 RUN
6 {
7   "limit": 500
8 }
9
10 // Specify request parameters to select data for visualization.
11 //
12 // Available parameters:
13 // - 'limit': maximum number of vectors to visualize.
14 // *Warning*: large values may cause browser to freeze.
15 //
16 // - 'filter': filter expression to select vectors for visualization.
17 // See https://qdrant.tech/documentation/concepts/filtering/
18 //
19 // - 'color_by': specify score or payload field to use for coloring points.
20 //
```

Corrigindo esse problema com a ferramenta *split text*, utilizada anteriormente, foi possível arrumar os dados, com isso os vetores foram gerados e o gráfico de points foi gerado.

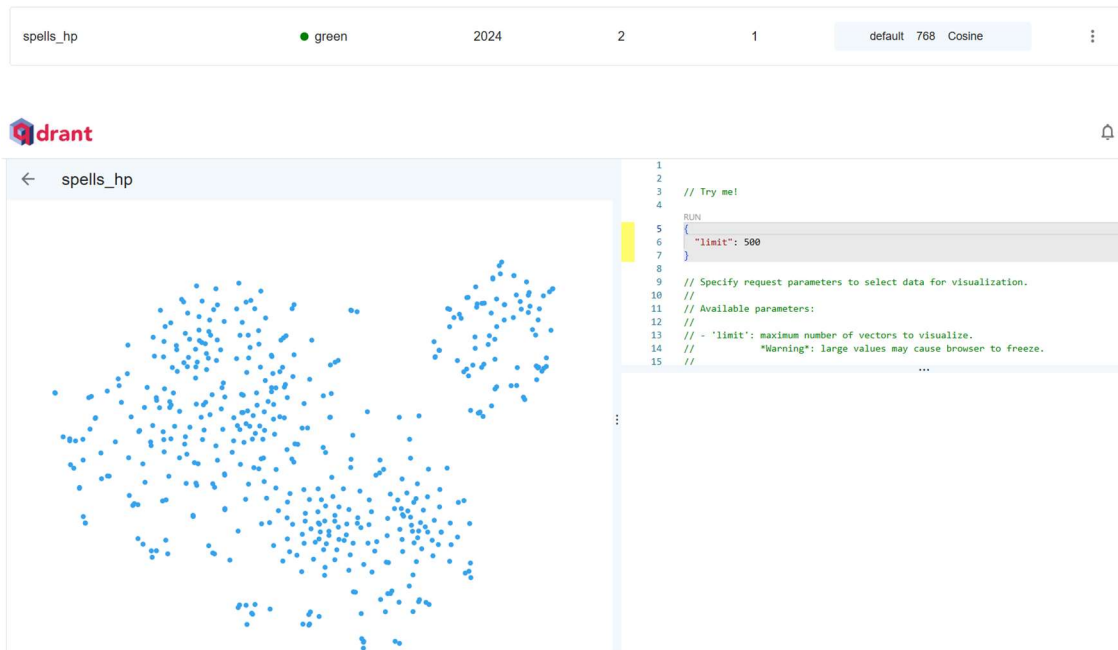
drant

← pokemon



```
1
2
3 // Try me!
4
5 RUN
6 {
7   "limit": 1000
8 }
9
10 // Specify request parameters to select data for visualization.
11 //
12 // Available parameters:
13 // - 'limit': maximum number of vectors to visualize.
14 // *Warning*: large values may cause browser to freeze.
15 //
16 // - 'filter': filter expression to select vectors for visualization.
17 // See https://qdrant.tech/documentation/concepts/filtering/
18 //
19 // - 'color_by': specify score or payload field to use for coloring points.
20 // How to use:
21 //
22 //   "color_by": {
23 //     "payload": "field_name"
24 //   }
25 //
```

Para praticar o conhecimento adquirido fiz o mesmo processo, desta vez usando a base de feitiços de Harry Potter.



## Dificuldades

Minha principal dificuldade foi rodar o localhost do langflow, não foi possível na minha máquina, tentei instalar e reinstalar, mas mesmo assim não foi possível. Por isso, decidi seguir pelo site. Encontrei algumas dificuldades para fazer a conexão do componente do Qdrant funcionar, mas após a configuração correta, rodou corretamente.

## Conclusões

O componente apresentado é muito eficiente no armazenamento dos embeddings, utilizando em conjunto com um componente para gerá-los, todo o fluxo fica fácil de executar.

## Referências

DataEX. (2022). OLTP e OLAP: O que são e quais as diferenças? Retrieved from [What Are Vector Embeddings: Types, Use Cases, & Models. \(2024\). Retrieved from <https://airbyte.com/data-engineering-resources/vector-embeddings>. Acesso em 06 abr 2025.](https://www.dataex.com.br/oltp-e-olap/O'Reilly, B. (2025). A Beginner's Guide to Building. Acesso em 05 abr 2025.</a></p></div><div data-bbox=)

Vectors. Qdrant. Retrieved from <https://qdrant.tech/documentation/>. Acesso em 07 abr 2025.

Using Langflow with Qdrant Vector Store. Dev Evangelista. (2024) Retrieved from <https://www.youtube.com/watch?v=blfWz8c2398>. Acesso em 05 abr 2025.