

Grupo SEXTO

Recipe Analytics



Antônio Hideto Borges Kotsubo - 236041

Gabriel Alves de Arruda - 248132

Guilherme Brentan de Oliveira - 252764

Gustavo Ferreira Gitzel - 223559

Maria Eduarda Elias Rocha - 248408

Pedro Sanchez Bitencourt - 231133

Tópicos



Relembrando

Resumo projeto 1
Mudanças



Modelos



Análises



Conclusão

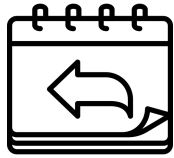
Resumo projeto 1

- Inclusão de todas as tabelas do **FoodDB**: Compound, Ontology, Flavor
- Perguntas iniciais propostas:
 - Quais são as receitas mais nutritivas?
 - Qual receita é mais rica em X vitamina/nutriente?
 - **Qual o ingrediente mais comum nas receitas?**
 - **Quão nutritiva são as receitas de determinada região do globo?**
 - Quais receitas são mais consumidas em determinada região do globo?
 - Quão nutritivo é o ingrediente mais consumido em determinadas receitas?

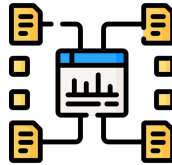
Mudanças

- Remoção de tabelas do **FooDB**: CompoundOntologyTerm, CompoundsFlavor, CompoundsHealthEffect, OntologySynonym, OntologyTerm, HealthEffect, CompoundSubstituent, Flavor.
- Separação do Content em **ContentCompound** e **ContentNutrient**.
- Perguntas **mudaram**

Tópicos



Relembrando



Modelos

Conceitual
Lógico Relacional
Lógico de grafos

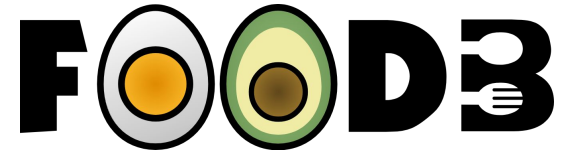
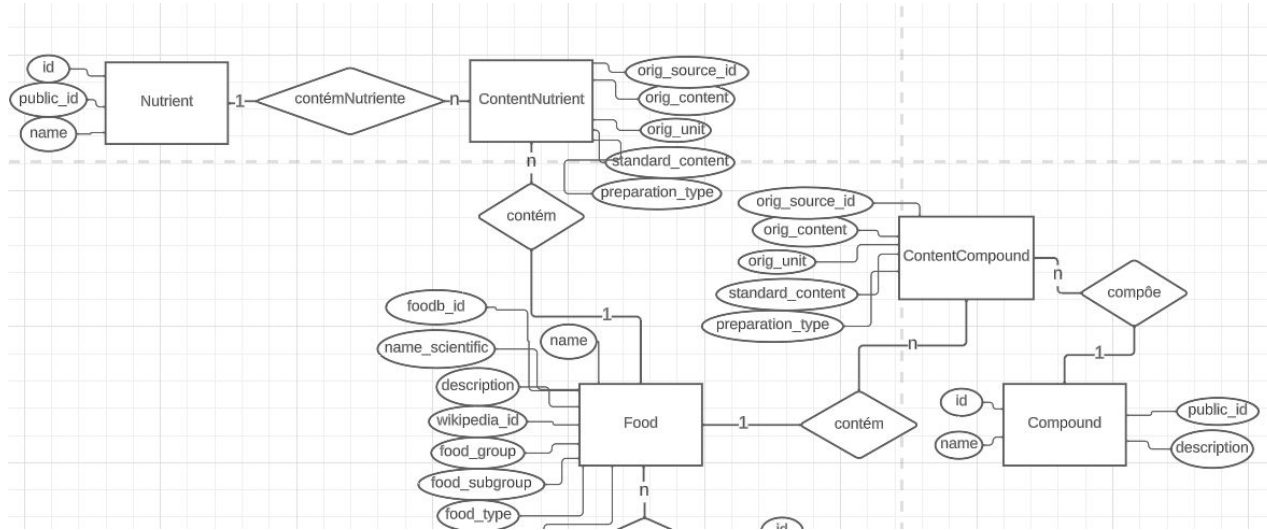


Análises

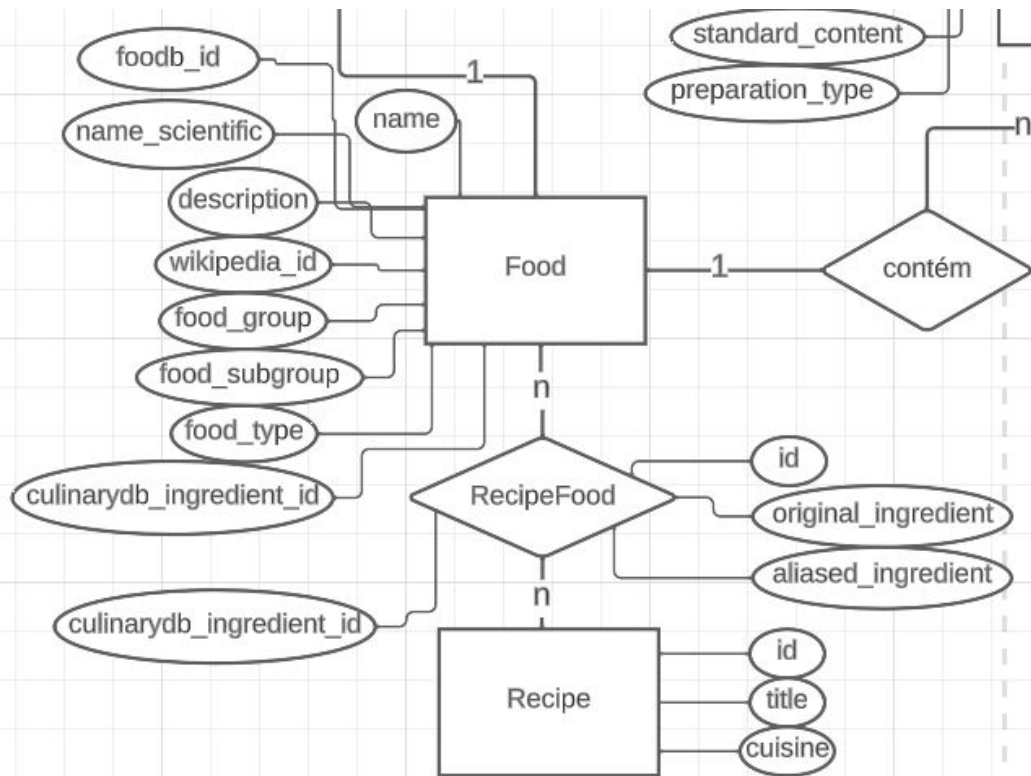


Conclusão

Conceitual



Conceitual



Lógico Relacional

ContentCompound(id, source_id, food_id, orig_source_id, orig_content, orig_unit, standard_content, preparation_type)

source_id chave estrangeira - Compound(id)

food_id chave estrangeira - Food(id)

ContentNutrient(id, source_id, food_id, orig_source_id, orig_content, orig_unit, standard_content, preparation_type)

source_id chave estrangeira - Nutrient(id)

food_id chave estrangeira - Food(id)

Compound(id, public_id, name, state, description)

Nutrient(id, public_id, name)

Lógico Relacional

Food(id, name, name_scientific, description, wikipedia_id, food_group, food_subgroup, food_type, entity_id, food_id)

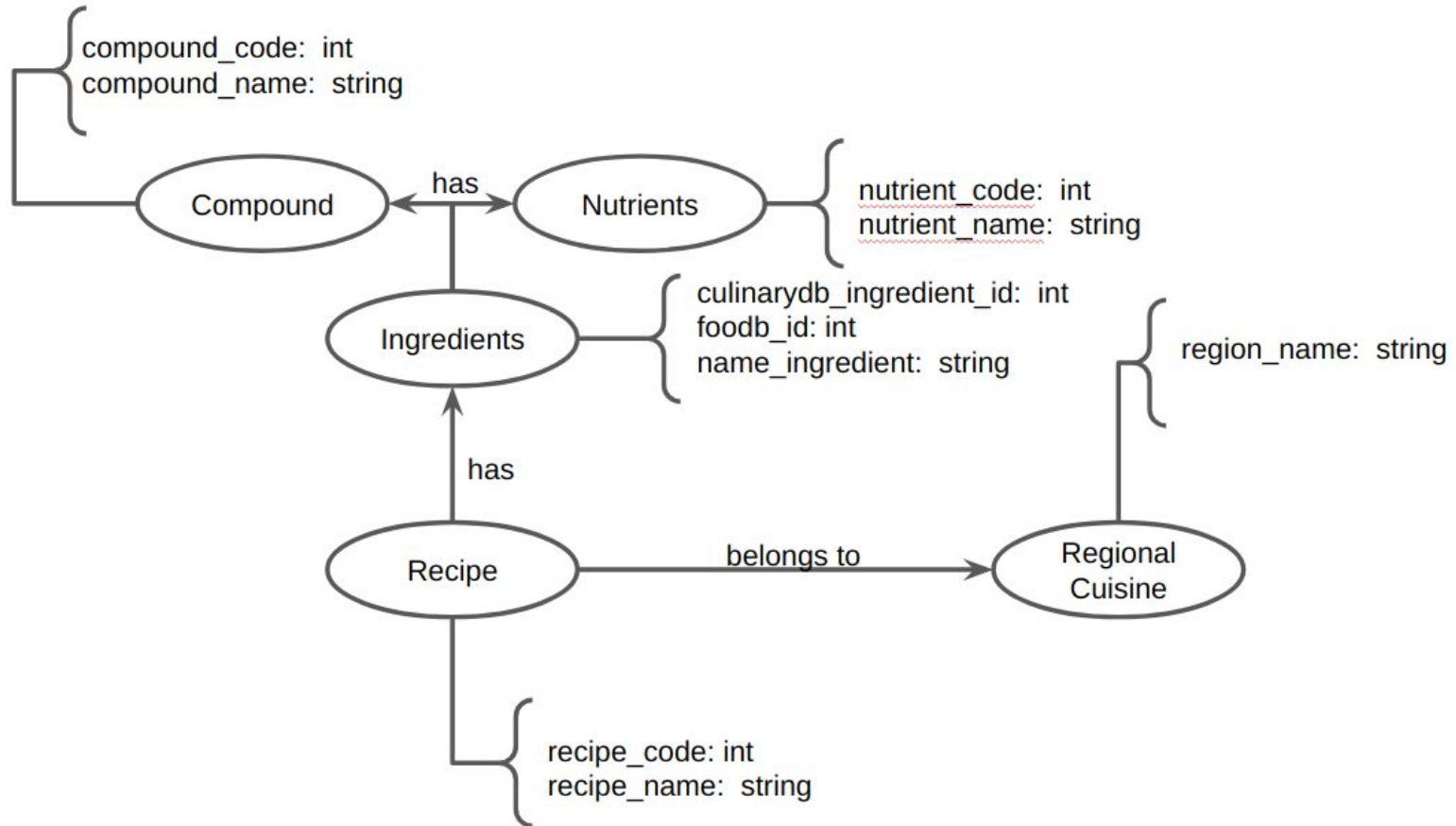
Recipe(id, title, cuisine)

RecipeFood(id, original_ingredient, aliased_ingredient, entity_id, recipe_id)

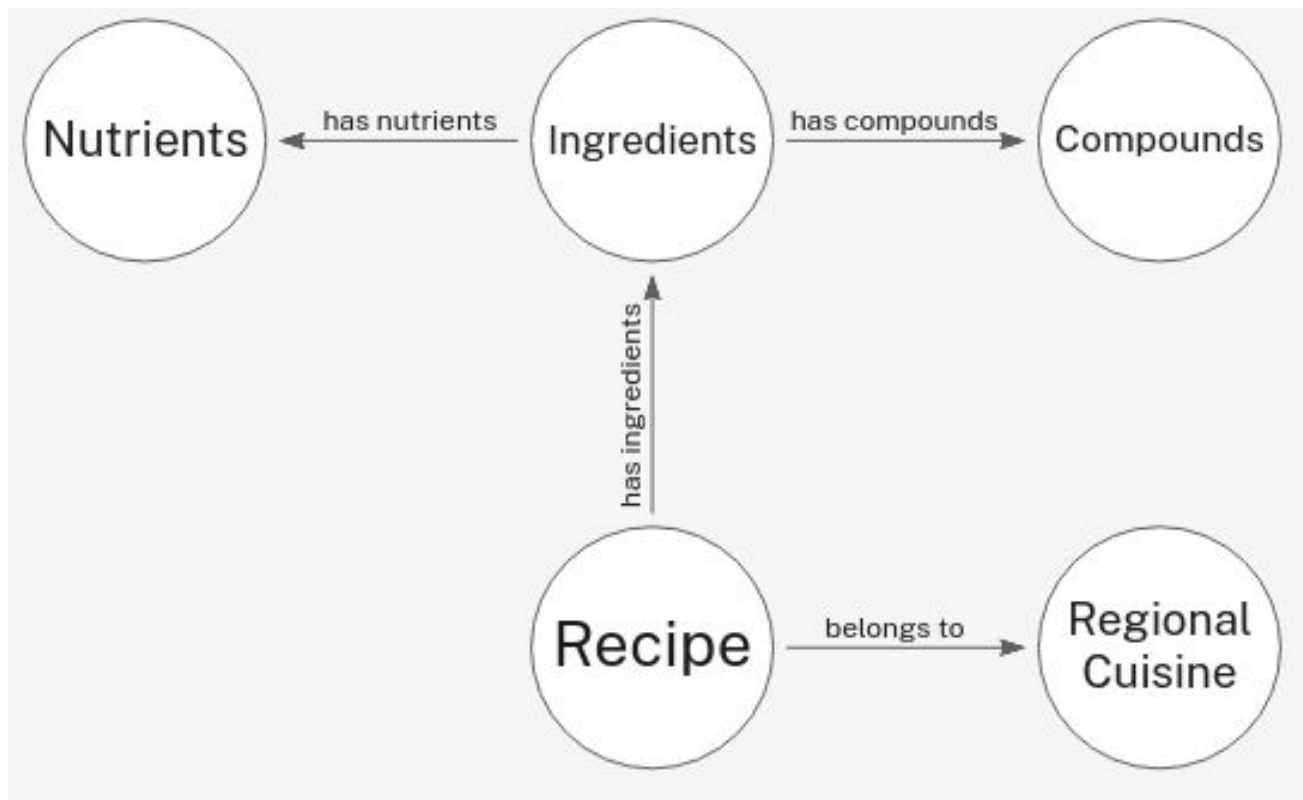
entity_id chave estrangeira - Food(entity_id)

recipe_id chave estrangeira - Recipe(id)

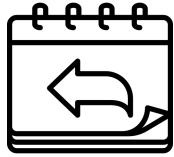
Lógico de Grafos



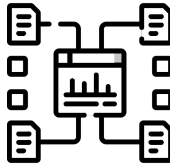
Lógico de Grafos



Tópicos



Relembrando



Modelos



Análises

Integração
Perguntas
Análise SQL
Análise Grafo



Conclusão

Integração

- Python foi usado para fazer a filtragem dos dados e junção entre as bases de dados para ambos os modelos
- Biblioteca para comparar os nomes dos alimentos: **fuzzywuzzy**
- **Jupyter Notebook**: base de dados do SQL
- **Neo4j**: modelo de grafos

Perguntas

- Perguntas SQL:
 - Quais são os 5 alimentos mais ricos em um determinado nutriente?
 - Qual é a receita mais complexa (ou seja, a que tem mais ingredientes)?
 - Quais são os 5 compostos mais comuns encontrados nos alimentos?
- Perguntas Grafos:
 - Qual região tem a maior variedade nutritiva?
 - Quais os ingredientes mais comuns de uma região?
 - Quais são as combinações mais comuns de Componentes e Nutrientes?

Análise SQL

- Análises SQL:

- Quais são os 5 alimentos mais ricos em um determinado nutriente?

```
SELECT f.name AS food, SUM(cn.orig_content) AS total_nutrient FROM Food f JOIN ContentNutrient cn ON f.id = cn.food_id
WHERE cn.nutrient_id = (SELECT id FROM Nutrient WHERE name = 'Carbohydrate')
GROUP BY f.name ORDER BY total_nutrient DESC LIMIT 5;
```

- Qual é a receita mais complexa (ou seja, a que tem mais ingredientes)?

```
SELECT r.title AS recipe, COUNT(*) AS ingredient_count FROM Recipe r JOIN RecipeFood rf ON r.id = rf.recipe_id
GROUP BY r.title ORDER BY ingredient_count DESC LIMIT 1;
```

- Quais são os 5 compostos mais comuns encontrados nos alimentos?

```
SELECT c.name AS compound, COUNT(*) AS frequency FROM Compound c JOIN ContentCompound cc ON c.id = cc.compound_id
GROUP BY c.name ORDER BY frequency DESC LIMIT 5;
```

Análise SQL

- Quais são os 5 alimentos mais ricos em um determinado nutriente?

index	NAME	TOTAL_NUTRIENT
0	Breakfast cereal	28929560
1	Pastry	10055600
2	Biscuit	8477470
3	Corn	5946590.45
4	Soup	5133620

- Qual é a receita mais complexa (ou seja, a que tem mais ingredientes)?

index	Key	Value
0	TITLE	Shepherd's Pie
1	INGREDIENT_COUNT	68

- Quais são os 5 compostos mais comuns encontrados nos alimentos?

index	NAME	FREQUENCY
0	alpha-Tocopherol	7476
1	Pyridoxine	5212
2	L-Alanine	3621
3	Glycine	3616
4	L-Proline	3572

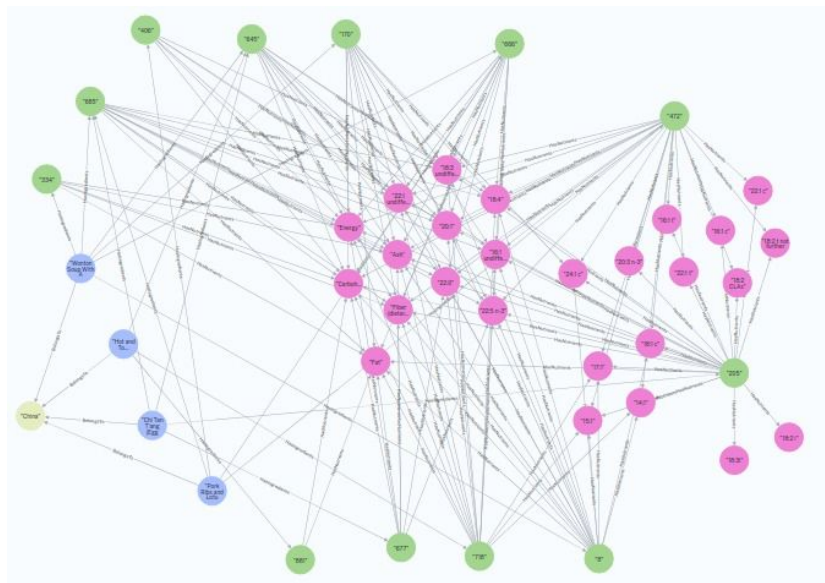
Análise Grafos

- Análises Grafos:
 - Qual região tem a maior variedade nutritiva?

```
MATCH (
  c:RegionalCuisine)-[:BelongsTo]-
  (r:Recipe)-[:HasIngredients]-
  (i:Ingredients)-[:HasNutrients]-
  (n:Nutrients)
RETURN c.region_name AS cuisine, count(n)
AS nutrientCount
ORDER BY nutrientCount DESC
```

	cuisine	nutrientCount
1	"USA"	1460000
2	"Italy"	680192
3	"Indian Subcontinent"	307136
4	"Mexico"	250169
5	"France"	248966
6	"Caribbean"	109265
7	"Middle East"	100251
8	"Greece"	90749
9	"China"	90326

```
MATCH (
  c:RegionalCuisine {region_name:
  'China'})-[b:BelongsTo]-
  (r:Recipe)-[j:HasIngredients]-
  (i:Ingredients)-[h:HasNutrients]-
  (n:Nutrients)
RETURN n, h, i, b, r, c, j
LIMIT 200
```



Análise Grafos

- Análises Grafos:
 - Quais os ingredientes mais comuns uma região?

```
MATCH (
  c:RegionalCuisine {region_name:
'USA'})-[:BelongsTo]-
  (r:Recipe)-[:HasIngredients]-
  (i:Ingredients)
RETURN i.name_ingredient AS ingredient,
COUNT(i) AS frequency
ORDER BY frequency DESC
LIMIT 10;
```

```
MATCH (
  c:RegionalCuisine {region_name:
'China'})-[:BelongsTo]-
  (r:Recipe)-[:HasIngredients]-
  (i:Ingredients)
RETURN i.name_ingredient AS ingredient,
COUNT(i) AS frequency
ORDER BY frequency DESC
LIMIT 10;
```

	ingredient	frequency
1	"Salt"	8296
2	"Sugar"	7916
3	"Pepper"	7611
4	"Butter"	6831
5	"Garlic"	5235
6	"Olive"	4668
7	"Water"	3776
8	"Cream"	2923
9	"Chicken"	2815

	ingredient	frequency
1	"Ginger"	532
2	"Garlic"	526
3	"Sesame"	497
4	"Sugar"	477
5	"Salt"	435
6	"Water"	395
7	"Chicken"	373
8	"Welsh onion"	353
9	"Corn"	330

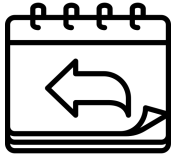
Análise Grafos

- Análises Grafos:
 - Quais são as combinações mais comuns de Componentes e Nutrientes?

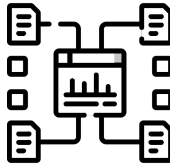
```
MATCH
(i:Ingredients)-[:HasNutrients]->(n:Nutrients)
MATCH (i)-[:HasCompound]->(c:Compounds)
RETURN n.nutrient_name AS nutrient,
c.compound_name AS compound, COUNT(i) AS
frequency
ORDER BY frequency DESC
LIMIT 10;
```

	nutrient	compound	frequency
1	"Carbohydrate"	"Pyridoxine"	551
2	"Energy"	"Pyridoxine"	550
3	"Fiber (dietary)"	"Pyridoxine"	523
4	"Ash"	"Pyridoxine"	519
5	"22:5 n-3"	"Pyridoxine"	472
6	"Carbohydrate"	"L-Alanine"	471
7	"Carbohydrate"	"Glycine"	471
8	"Carbohydrate"	"L-Proline"	470
9	"Energy"	"Glycine"	469

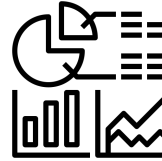
Tópicos



Relembrando



Modelos



Análises



Conclusão

Conclusão

- Dificuldades:
 - **Limitações** do Neo4J e do Jupyter Notebook

ERROR

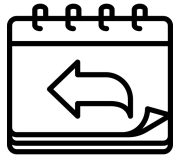
Neo.TransientError.General.MemoryPoolOutOfMemoryError

The allocation of an extra 39,0 MiB would use more than the limit 12,0 GiB. Currently using 12,0 GiB.

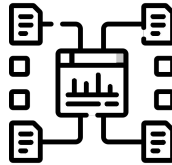
Ingredientes - 671
Receitas - 45522
Componentes - 393
Nutrientes - 27
Regiões - 26

- Muitos dados **não tratados** nas bases de dados
 - Mais de 5 milhões de linhas em alguns arquivos (700MB) -> Filtrado para 500 mil linhas (87MB)

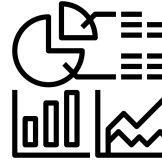
Tópicos



Relembrando



Modelos



Análises



Conclusão



Obrigado!