Софийски университет “Св. Климент Охридски”

Факултет по математика и информатика

Проект

по

Бази от знания

на тема

Работа с база от знания за храни (проект тип А)

Изготвен от: **Кирил Русев, ф.н. 2MI3400203, спец. ИИ, I курс**

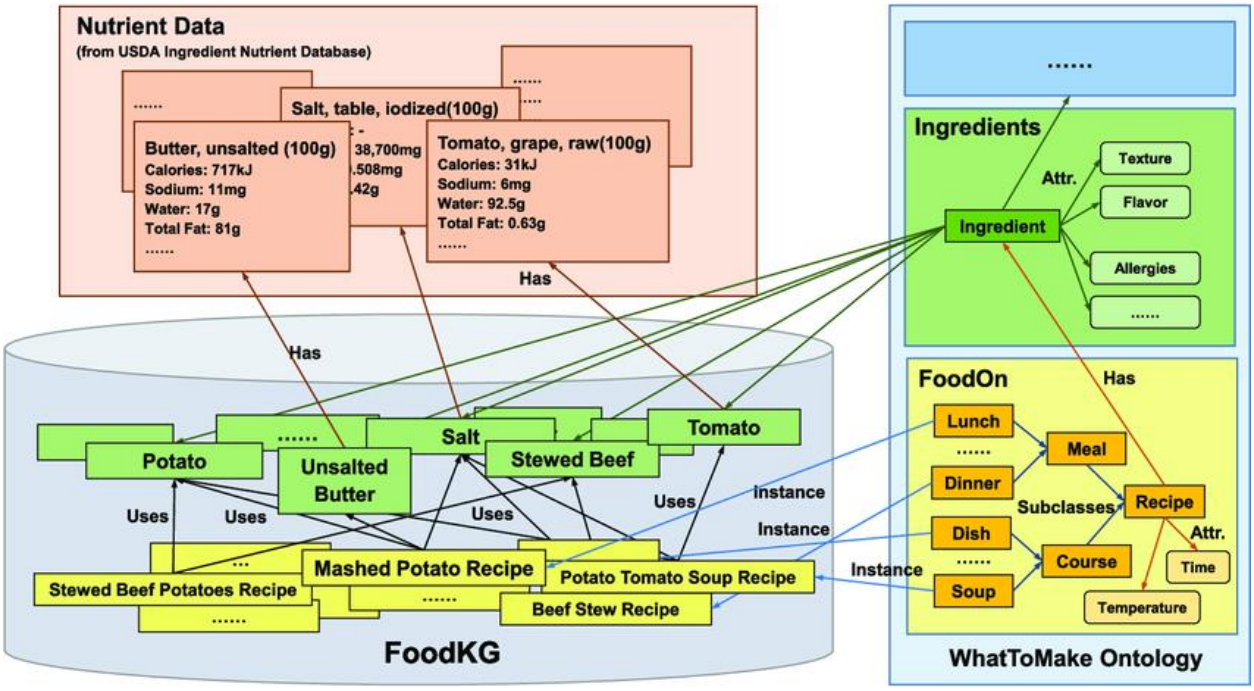
12.06.2023 г.

Описание на данните, използвани в проекта

FoodKG е граф от знания, съдържащ рецепти, съставки и хранителни вещества, който обхваща над милион рецепти и 67 милиона тройки. За изграждането му са използвани подмножества на обекти от онтологията FoodOn, като е обогатена информацията за избраните обекти.

Подробно описание за това как е създадена базата от знания има в [1], а документацията е налична в [2].

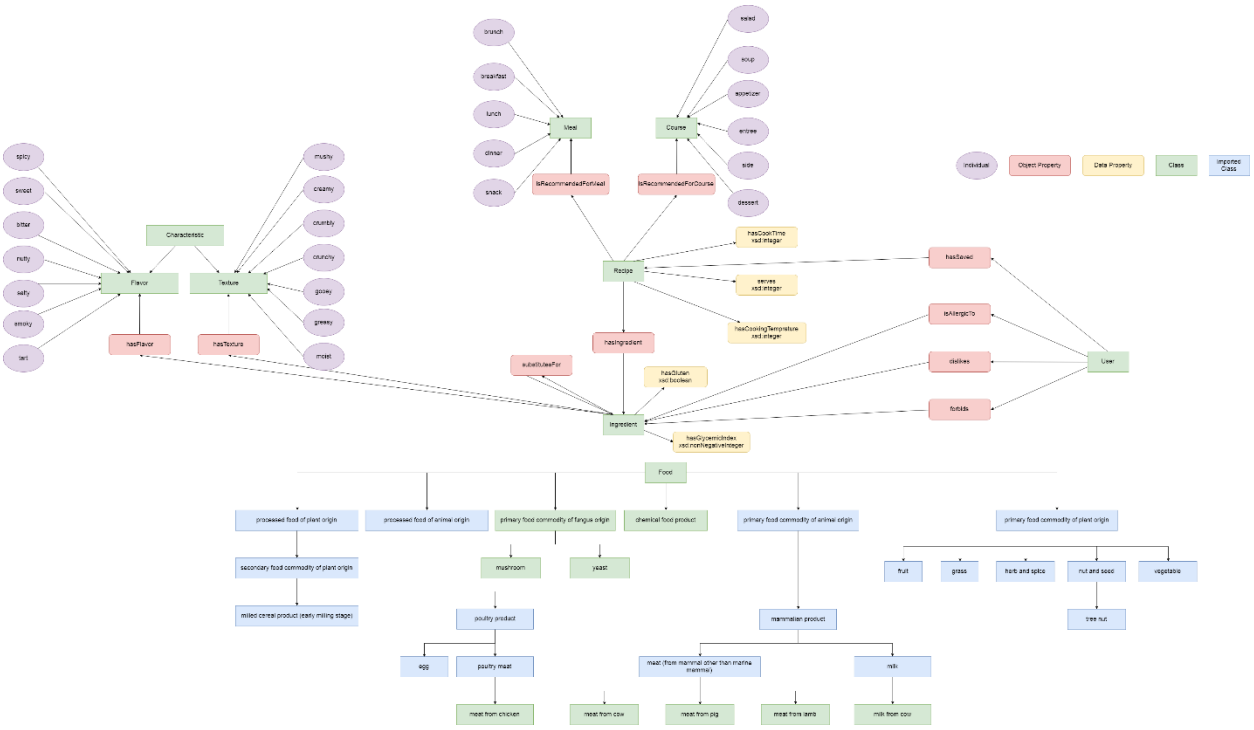
Основните обекти в FoodKG са част от онтологията WhatToMake, която е използвана за целите на проекта.



Фигура 1 Структура на FoodKG

Онтологията WhatToMake е съставена от три части: Food, FoodOn и Ingredient. Food съдържа всички базови класове, които са създадени за онтологията. FoodOn съдържа всички импортирани класове от FoodOn (тази част не е използвана за заявките в проекта). Ingredient съдържа всички индивиди, които представляват различни рецепти и съставки. Документацията и .rdf файловете, описващи частите на онтологията, могат да бъдат намерени на следните линкове:

Food	Документация	Онтология
FoodOn	Документация	Онтология
Ingredient	Документация	Онтология



Фигура 2 Визуализация на основните обекти в онтологията WhatToMake

Описание на принципите и методите за взимане на решения и правила за извод на GraphDB

GraphDB използва прав извод на правилата, за да извършва разсъждения [3]. Той следва стратегия, известна като пълна материализация, при която правилата за извод се прилагат многократно към експлицитно заявените факти, докато не могат да бъдат изведени други имплицитни факти. Този подход гарантира, че изведеното затваряне на базата знания е изчислено и достъпно за оценка или изпълнение на заявки към базата.

GraphDB предоставя възможност за включване и изключване на разсъжденията при изпълнение на заявки. Ако разсъждението е изключено, резултатът от заявката се базира само на знания и данни, които са експлицитно включени в онтологията.

В Приложение А са представени 3 примерни заявки, които демонстрират начина на взимане на решения от базата знания.

Първата от тях е за извличане на съставките, които са заместители на други съставки (използвайки симетричната релация **wtm:substitutesFor**).

Когато разсъждението е изключено, има 12 резултата, защото в онтологията връзката **wtm:substitutesFor** между съставките не е дефинирана експлицитно и в двете посоки.

Когато разсъждението е включено, има 16 резултата от изпълнението на заявката.

Втората от тях е за извличане на храните които могат да бъдат ядени преди вечеря (използвайки транзитивната релация **wtm:isEatenBefore**).

Когато разсъждението е изключено, има 10 резултата, защото в онтологията връзката **wtm:isEatenBefore** може да бъде транзитивна. Например *ind:BananaBread* е храна, за която е дефинирано *wtm:isRecommendedForMeal wtm:Breakfast*, а за *wtm:Breakfast* е дефинирано, че *wtm:isEatenBefore wtm:Lunch*, както и *wtm:Lunch wtm:isEatenBefore wtm:Dinner*, но няма експлицитна връзка *wtm:Breakfast wtm:isEatenBefore wtm:Dinner*, поради което *ind:BananaBread* не се включва в резултатите.

Когато разсъждението е включено, има 16 резултата от изпълнението на заявката. В тези резултати се включва и *ind:BananaBread*.

Третата от тях е за извличане на храните които имат сол в тях (използвайки обратната на *wtm:hasIngredient* релация **wtm:isIngredientIn**).

Когато разсъждението е изключено, няма резултати, понеже в онтологията връзката между храните и съставките е дефинирана със свойството *wtm:hasIngredient*, което е обратно на **wtm:isIngredientIn**.

Когато разсъждението е включено, има 16 резултата от изпълнението на заявката.

Проблеми и предизвикателства при разработката

За целите на проекта са използвани .rdf файловете на Food и Ingredient частите от WhatToMake. В последствие се оказа, че няма достатъчно сложни релации в онтологията, поради което тя беше разширена със следните релации:

Domain	Property	Range	Characterisitics
Ingredient	isIngredientIn	Recipe	inverse property (of hasIngredient)
Meal	isEatenBefore	Meal	Transitive property, inverse property (of isEatenAfter)
Meal	isEatenAfter	Meal	Transitive property, inverse property (of isEatenBefore)

За целта е използван **owlready2** [4]. Кодът е наличен в ***update_ontology.ipynb*** файла в документацията на проекта.

Първоначалните .rdf файлове са налични в папката ***food-repository/init,*** а файловете след добавяне на новите релации – в папката ***food-repository/updated***.

Бъдещо развитие

Проектът може да бъде разширен, като се разработи приложение, което да връща на потребителите резултатите от заявките. Също може да бъде добавена възможност потребителите сами да дефинират заявки към базата от знания. Не на последно място, базата от знания може да бъде разширена с още индивиди както и връзки между тях.

Приложение А: SPARQL заявките и резултатите от изпълнението им в GraphDB

Заявките са приложени във файла ***queries.txt***.

- **Заявка 1:** Извличане на всички храни, които съдържат банани, но не съдържат орехи

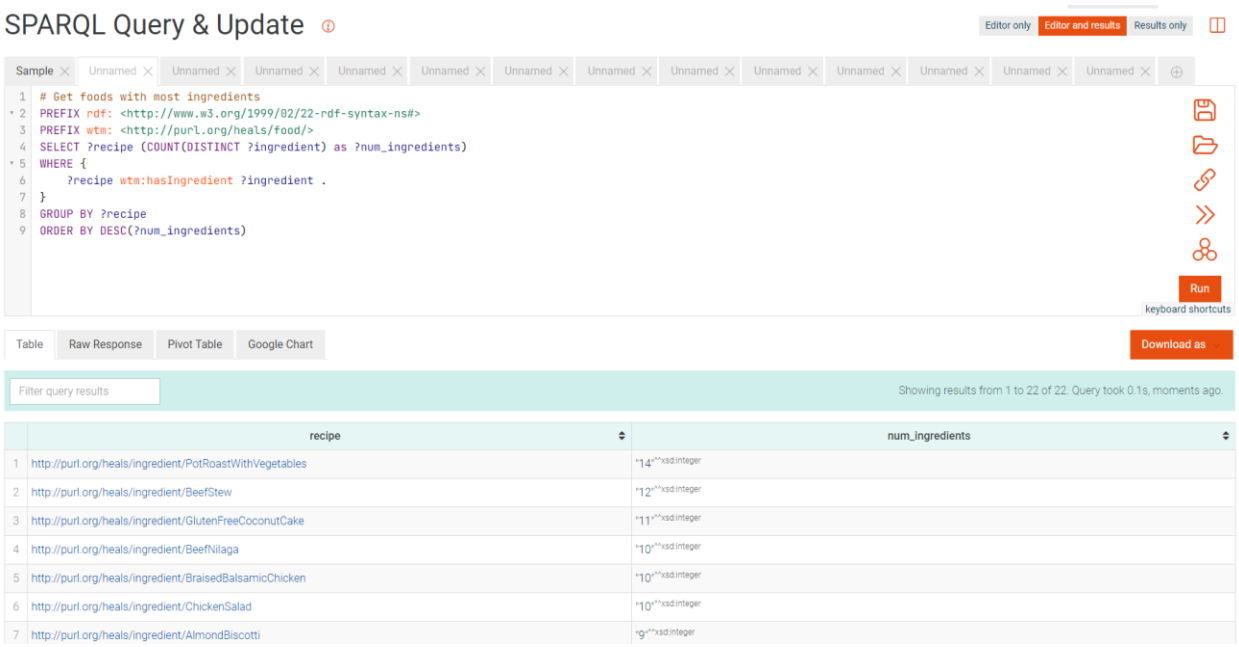
```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX ind: <http://purl.org/heals/ingredient/>
SELECT DISTINCT ?recipe
WHERE {
    {
        ?recipe wtm:hasIngredient ind:Banana .
    }
    MINUS
    {
        ?recipe wtm:hasIngredient ind:Walnut .
    }
}
```

Резултат от изпълнението:

- **Заявка 2:** Сортиране на храните по брой съставки

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?recipe (COUNT(DISTINCT ?ingredient) as ?num_ingredients)
WHERE {
  ?recipe wtm:hasIngredient ?ingredient .
}
GROUP BY ?recipe
ORDER BY DESC(?num_ingredients)
```

Резултат от изпълнението:

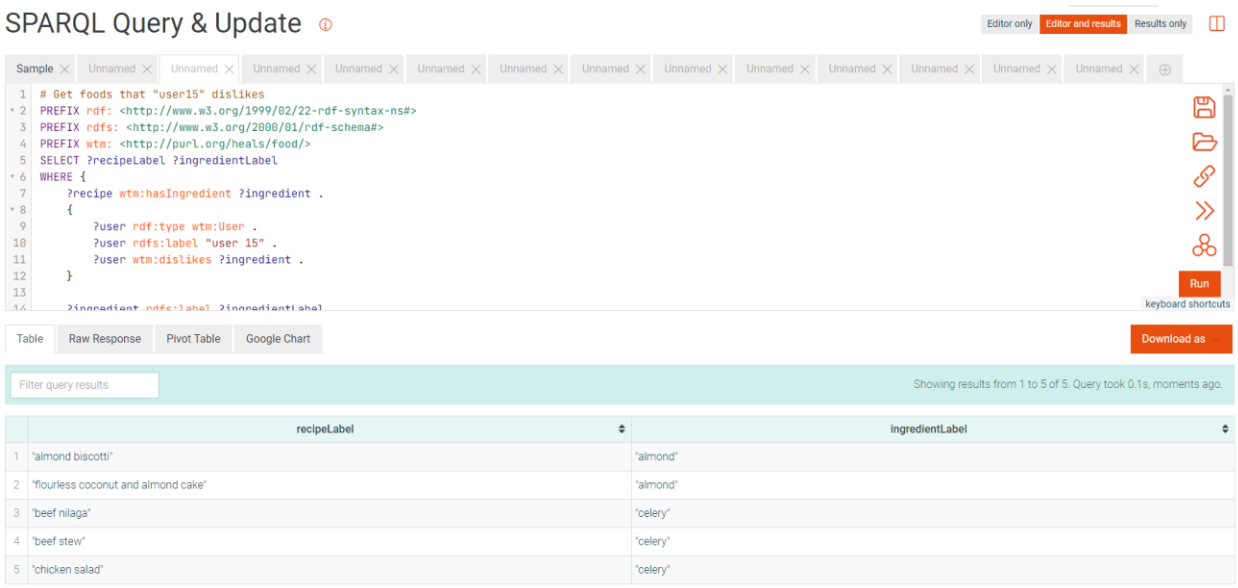


- **Заявка 3:** Извличане на храните, които потребителят „user15“ не харесва

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?recipeLabel ?ingredientLabel
WHERE {
    ?recipe wtm:hasIngredient ?ingredient .
    {
        ?user rdf:type wtm:User .
        ?user rdfs:label "user 15" .
        ?user wtm:dislikes ?ingredient .
    }

    ?ingredient rdfs:label ?ingredientLabel .
    ?recipe rdfs:label ?recipeLabel .
}
```

Резултат от изпълнението:

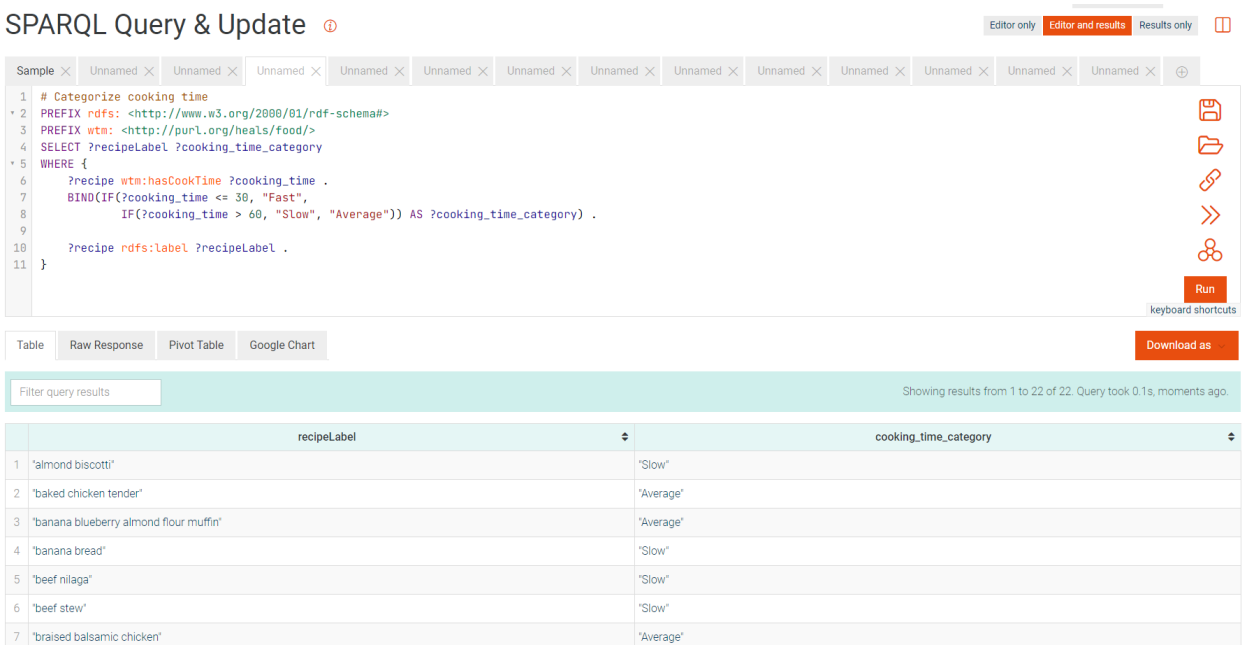


- **Заявка 4:** Категоризиране на храните спрямо времето им за приготвяне („Slow“ до 30 мин. вкл., „Medium“ между 30 и 60 мин. вкл. и „Fast“ над 60 мин.)

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?recipeLabel ?cooking_time_category
WHERE {
    ?recipe wtm:hasCookTime ?cooking_time .
    BIND(IF(?cooking_time <= 30, "Fast",
            IF(?cooking_time > 60, "Slow", "Average")) AS
?cooking_time_category) .

    ?recipe rdfs:label ?recipeLabel .
}
```

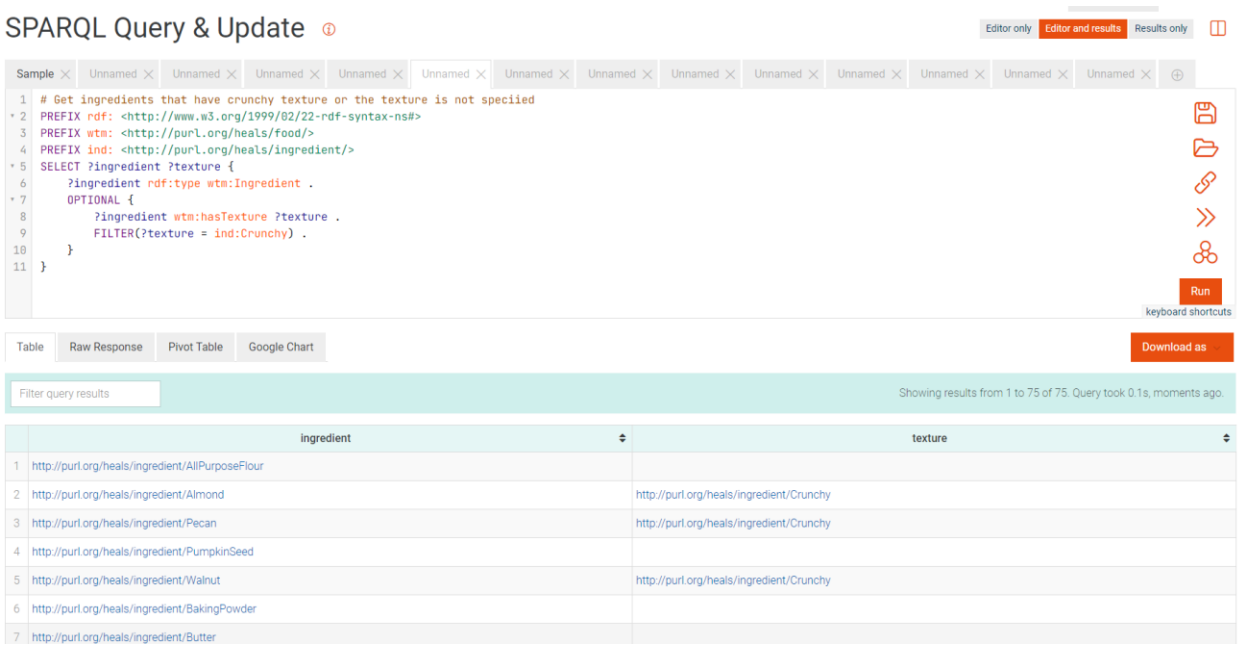
Резултат от изпълнението:



- **Заявка 5:** Извличане на съставките, които хрупкава текстура, като се включват и съставките без зададена структура в базата

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX ind: <http://purl.org/heals/ingredient/>
SELECT ?ingredient ?texture {
    ?ingredient rdf:type wtm:Ingredient .
    OPTIONAL {
      ?ingredient wtm:hasTexture ?texture .
        FILTER(?texture = ind:Crunchy) .
    }
}
```

Резултат от изпълнението:

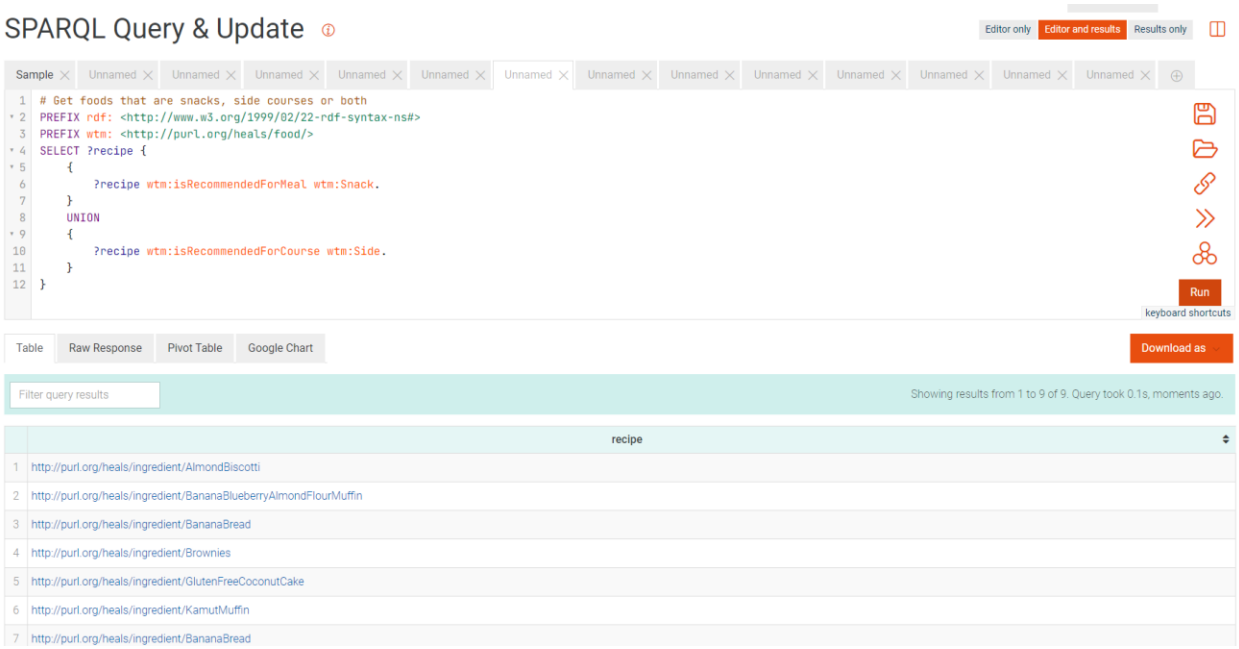


	ingredient	texture
1	http://purl.org/heals/ingredient/AllPurposeFlour	
2	http://purl.org/heals/ingredient/Almond	http://purl.org/heals/ingredient/Crunchy
3	http://purl.org/heals/ingredient/Pecan	http://purl.org/heals/ingredient/Crunchy
4	http://purl.org/heals/ingredient/PumpkinSeed	
5	http://purl.org/heals/ingredient/Walnut	http://purl.org/heals/ingredient/Crunchy
6	http://purl.org/heals/ingredient/BakingPowder	
7	http://purl.org/heals/ingredient/Butter	

- **Заявка 6:** Извличане на храните, които се препоръчват или за междинно ядене, или са гарнитура, или са и двете

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?recipe {
    {
        ?recipe wtm:isRecommendedForMeal wtm:Snack.
    }
    UNION
    {
        ?recipe wtm:isRecommendedForCourse wtm:Side.
    }
}
```

Резултат от изпълнението:

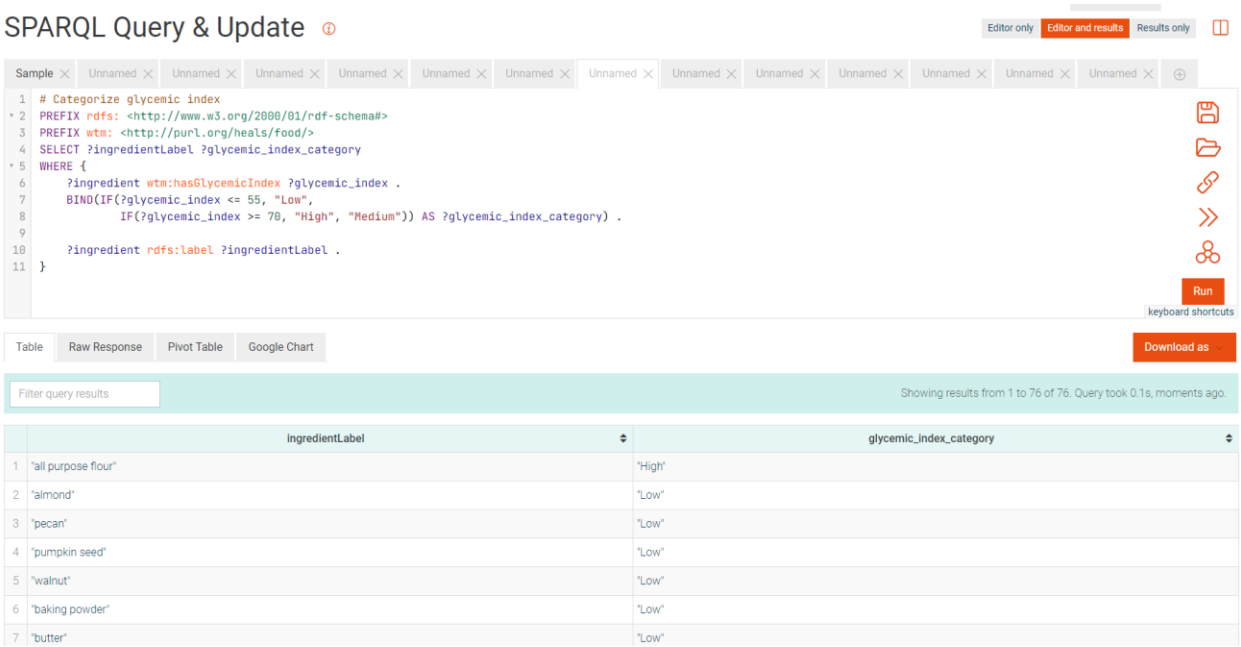


- **Заявка 7:** Категоризиране на храните спрямо гликемичния им индекс („Low“ до 55 вкл., „Medium“ между 55 и 70 и „High“ над 70 вкл.)

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?ingredientLabel ?glycemic_index_category
WHERE {
    ?ingredient wtm:hasGlycemicIndex ?glycemic_index .
    BIND(IF(?glycemic_index <= 55, "Low",
            IF(?glycemic_index >= 70, "High", "Medium")) AS
?glycemic_index_category) .

    ?ingredient rdfs:label ?ingredientLabel .
}
```

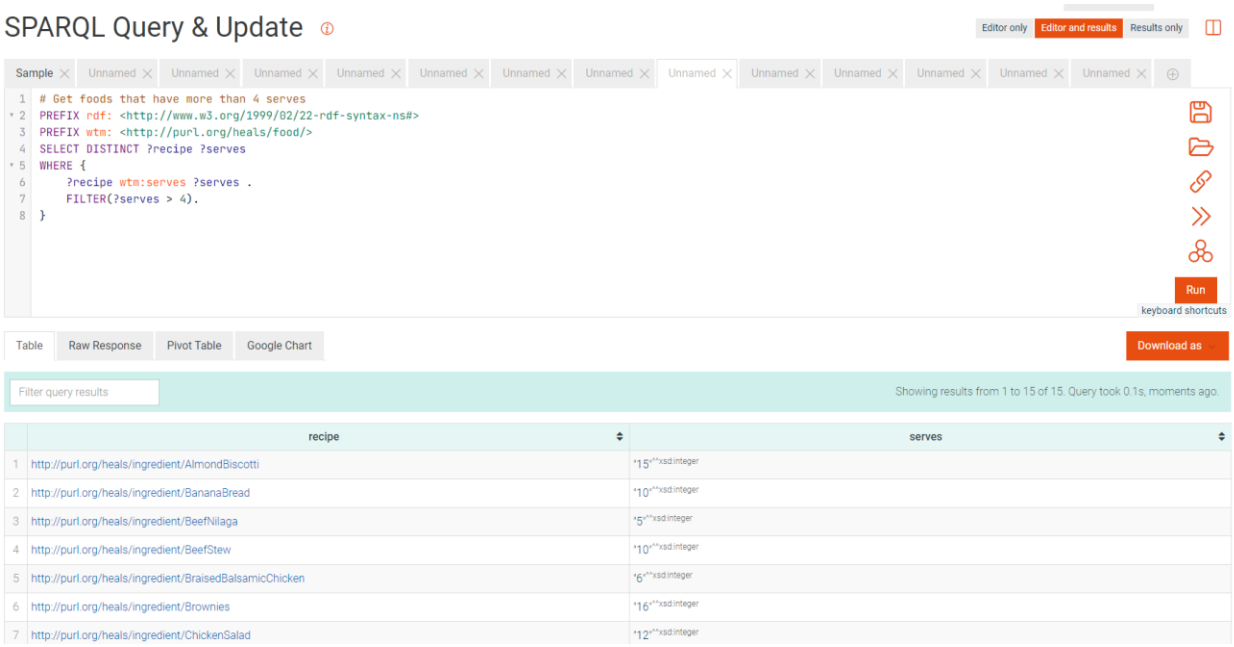
Резултат от изпълнението:



- **Заявка 8:** Извличане на храните, които имат повече от 4 порции

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT DISTINCT ?recipe ?serves
WHERE {
    ?recipe wtm:serves ?serves .
    FILTER(?serves > 4).
}
```

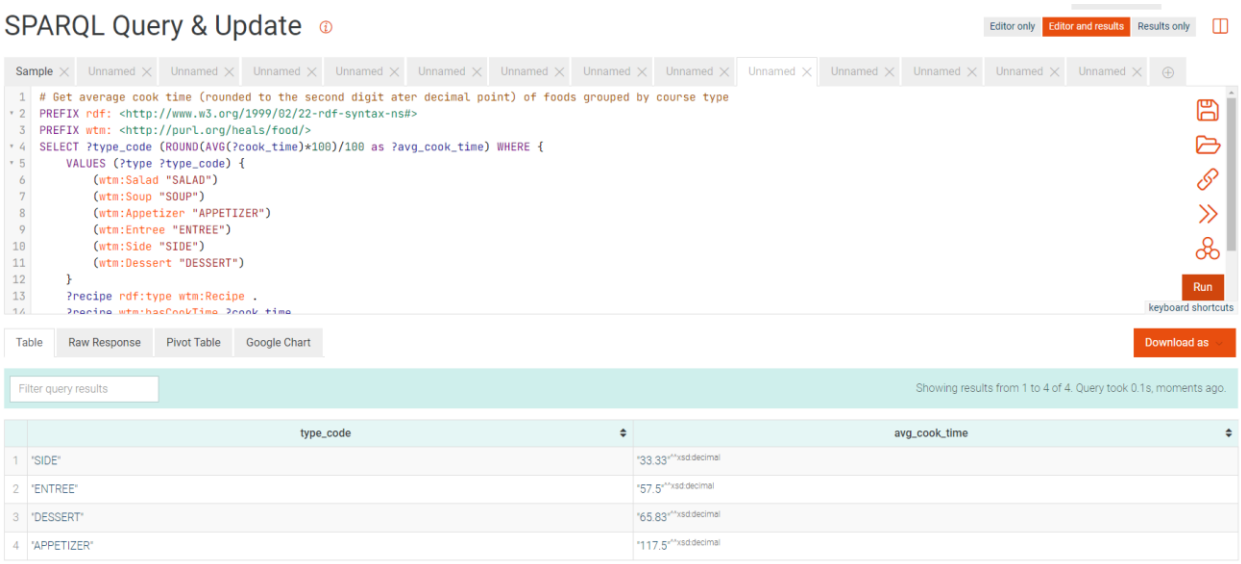
Резултат от изпълнението:



- **Заявка 9:** Извличане на средното време за приготвяне на храните, групирани по типа ястие. Резултатът е закръглен до втората цифра след десетичната запетая.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?type_code (ROUND(AVG(?cook_time)*100)/100 as ?avg_cook_time) WHERE {
  VALUES (?type ?type_code) {
        (wtm:Salad "SALAD")
        (wtm:Soup "SOUP")
        (wtm:Appetizer "APPETIZER")
        (wtm:Entree "ENTREE")
        (wtm:Side "SIDE")
        (wtm:Dessert "DESSERT")
    }
    ?recipe rdf:type wtm:Recipe .
    ?recipe wtm:hasCookTime ?cook_time .
    ?recipe wtm:isRecommendedForCourse ?type .
}
GROUP BY ?type_code
ORDER BY ?avg_cook_time
```

Резултат от изпълнението:



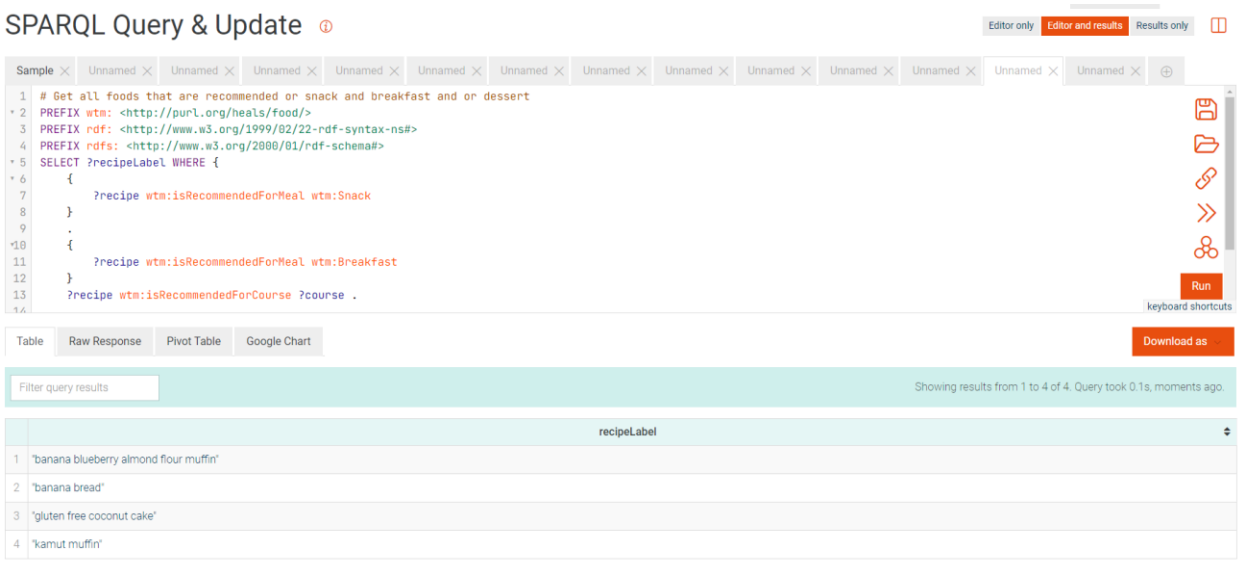
- **Заявка 10:** Извличане на храните, които се препоръчват и за междинно ядене, и за закуска, и за десерт

```
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?recipeLabel WHERE {
    {
        ?recipe wtm:isRecommendedForMeal wtm:Snack
    }
    .
    {
        ?recipe wtm:isRecommendedForMeal wtm:Breakfast
    }
    ?recipe wtm:isRecommendedForCourse ?course .

    FILTER(?course = wtm:Dessert)

    ?recipe rdfs:label ?recipeLabel .
}
```

Резултат от изпълнението:

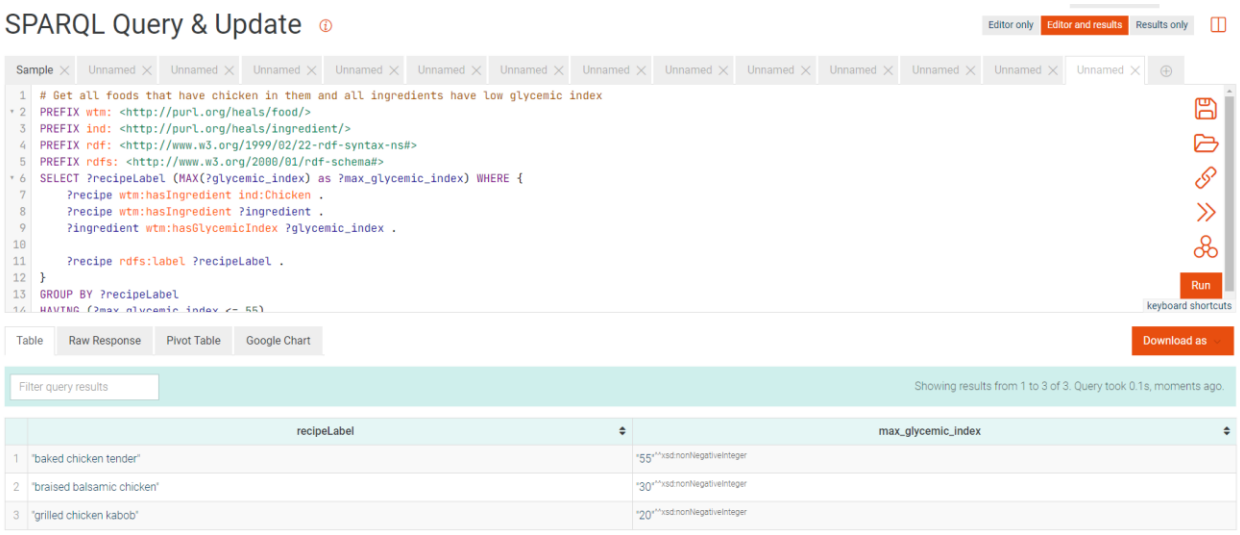


- **Заявка 11:** Извличане на храните, които имат пиле в тях и всичките им съставки имат нисък гликемичен индекс (<= 55)

```
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX ind: <http://purl.org/heals/ingredient/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?recipeLabel (MAX(?glycemic_index) as ?max_glycemic_index) WHERE {
    ?recipe wtm:hasIngredient ind:Chicken .
    ?recipe wtm:hasIngredient ?ingredient .
    ?ingredient wtm:hasGlycemicIndex ?glycemic_index .

    ?recipe rdfs:label ?recipeLabel .
}
GROUP BY ?recipeLabel
HAVING (?max_glycemic_index <= 55)
```

Резултат от изпълнението:

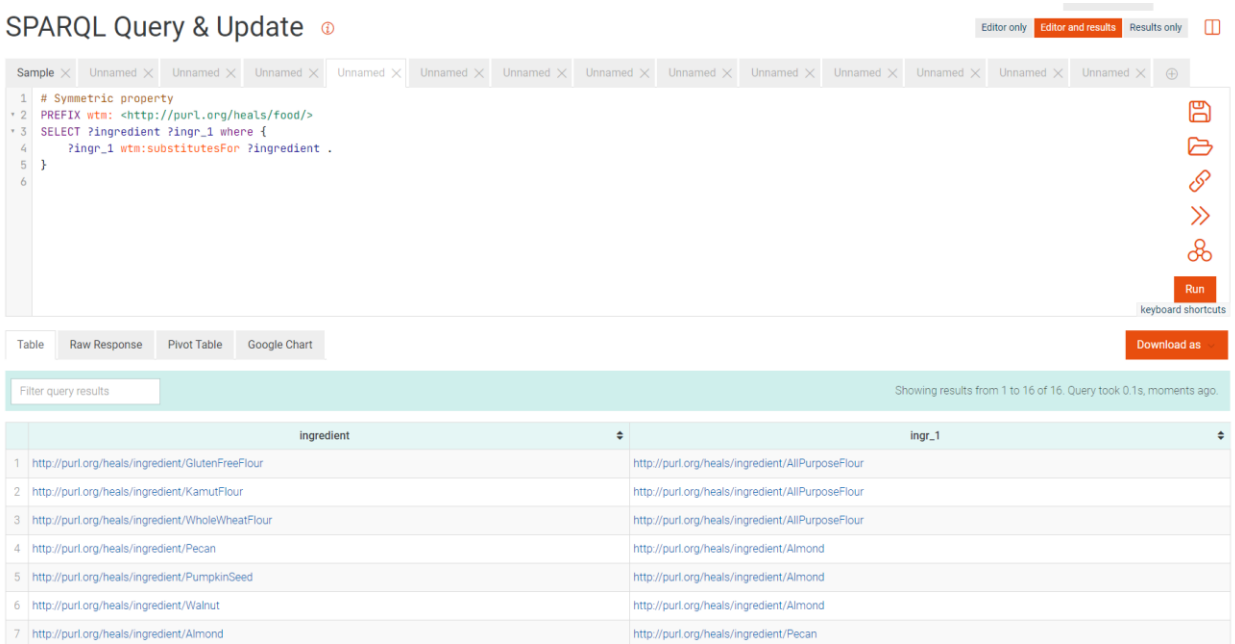


Заявки, които демонстрират начина на взимане на решения от базата знания:

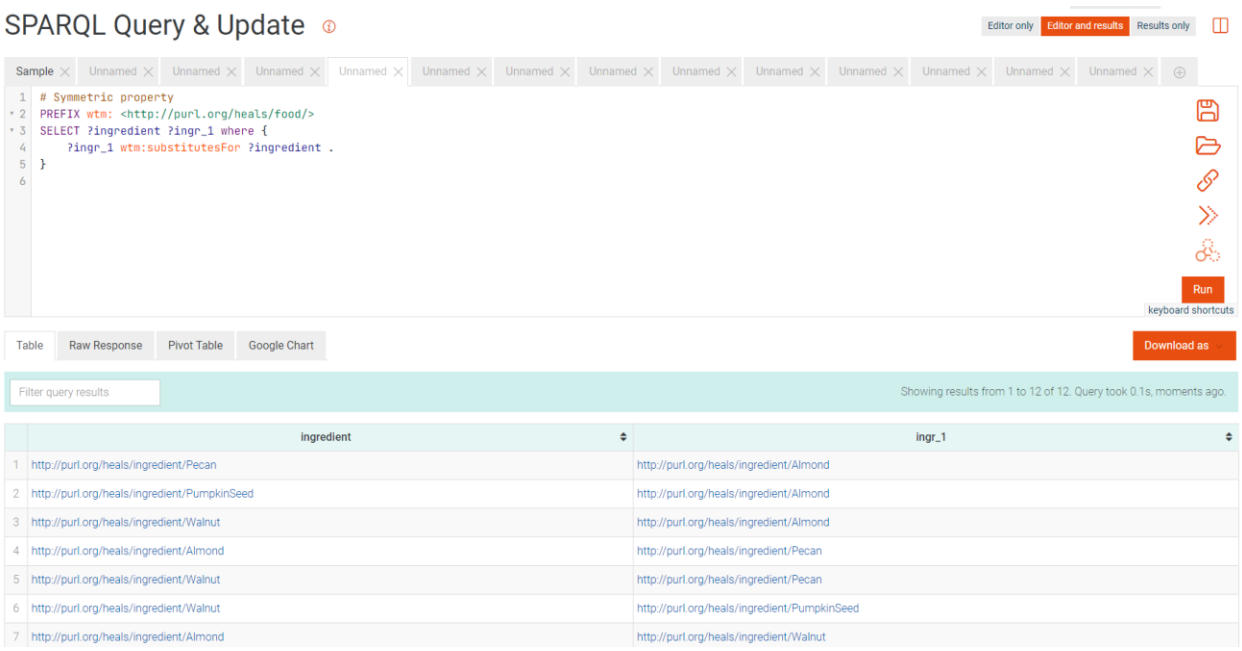
- **Заявка 12:** Извличане на съставките, които са заместители на други съставки (използвайки симетричната релация **wtm:substitutesFor**)

```
PREFIX wtm: <http://purl.org/heals/food/>
SELECT ?ingredient ?ingr_1 where {
  ?ingr_1 wtm:substitutesFor ?ingredient .
}
```

Резултат от изпълнението (с включен извод):



Резултат от изпълнението (с изключен извод):

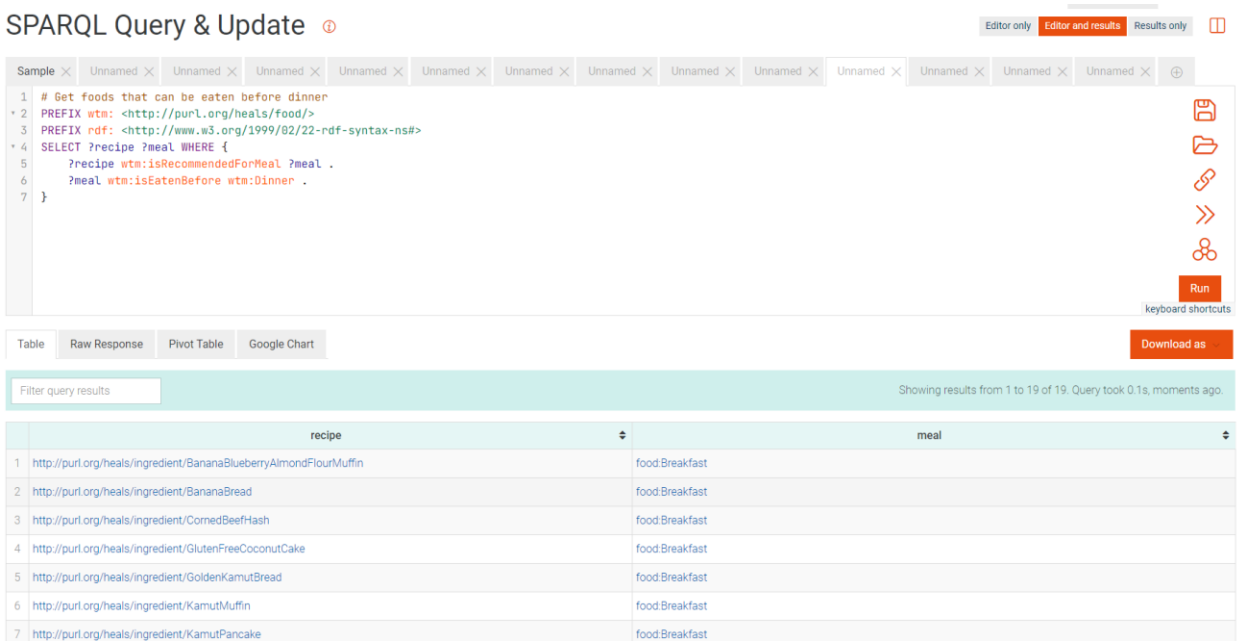


	ingredient	ingr_1
1	http://purl.org/heals/ingredient/Pecan	http://purl.org/heals/ingredient/Almond
2	http://purl.org/heals/ingredient/PumpkinSeed	http://purl.org/heals/ingredient/Almond
3	http://purl.org/heals/ingredient/Walnut	http://purl.org/heals/ingredient/Almond
4	http://purl.org/heals/ingredient/Almond	http://purl.org/heals/ingredient/Pecan
5	http://purl.org/heals/ingredient/Walnut	http://purl.org/heals/ingredient/Pecan
6	http://purl.org/heals/ingredient/Walnut	http://purl.org/heals/ingredient/PumpkinSeed
7	http://purl.org/heals/ingredient/Almond	http://purl.org/heals/ingredient/Walnut

- **Заявка 13:** Извличане на храните които могат да бъдат ядени преди вечеря (използвайки транзитивната релация **wtm:isEatenBefore**)

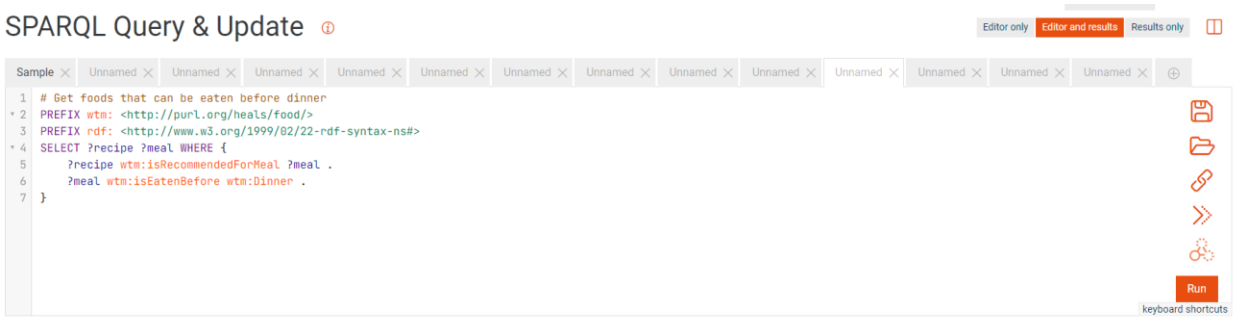
```
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?recipe ?meal WHERE {
    ?recipe wtm:isRecommendedForMeal ?meal .
    ?meal wtm:isEatenBefore wtm:Dinner .
}
```

Резултат от изпълнението (с включен извод):



	recipe	meal
1	http://purl.org/heals/ingredient/BananaBlueberryAlmondFlourMuffin	food:Breakfast
2	http://purl.org/heals/ingredient/BananaBread	food:Breakfast
3	http://purl.org/heals/ingredient/CornedBeefHash	food:Breakfast
4	http://purl.org/heals/ingredient/GlutenFreeCoconutCake	food:Breakfast
5	http://purl.org/heals/ingredient/GoldenKamutBread	food:Breakfast
6	http://purl.org/heals/ingredient/KamutMuffin	food:Breakfast
7	http://purl.org/heals/ingredient/KamutPancake	food:Breakfast

Резултат от изпълнението (с изключен извод):

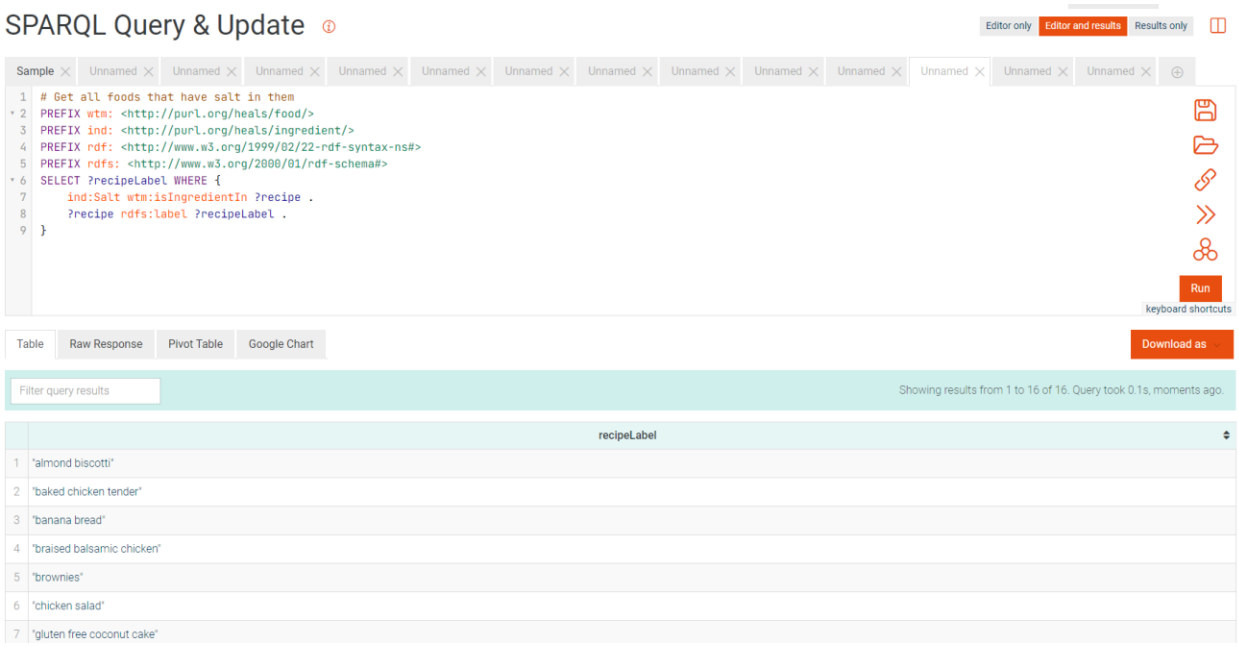


	recipe	meal
1	http://purl.org/heals/ingredient/BakedChickenTender	food:Lunch
2	http://purl.org/heals/ingredient/BeefNilaga	food:Lunch
3	http://purl.org/heals/ingredient/BeefStew	food:Lunch
4	http://purl.org/heals/ingredient/BraisedBalsamicChicken	food:Lunch
5	http://purl.org/heals/ingredient/ChickenSalad	food:Lunch
6	http://purl.org/heals/ingredient/GrilledChickenKabob	food:Lunch
7	http://purl.org/heals/ingredient/PotRoastWithVegetables	food:Lunch

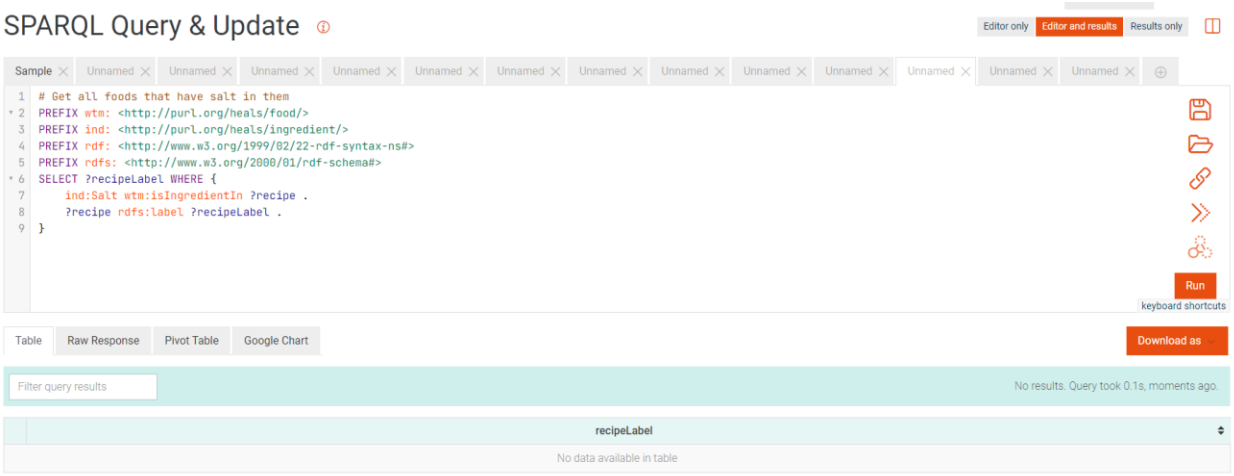
- **Заявка 14:** Извличане на храните които имат сол в тях (използвайки обратната на *wtm:hasIngredient* релация **wtm:isIngredientIn**)

```
PREFIX wtm: <http://purl.org/heals/food/>
PREFIX ind: <http://purl.org/heals/ingredient/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?recipeLabel WHERE {
    ind:Salt wtm:isIngredientIn ?recipe .
    ?recipe rdfs:label ?recipeLabel .
}
```

Резултат от изпълнението (с включен извод):



Резултат от изпълнението (с изключен извод):



Библиография

[1] Steven Haussmann, Oshani Seneviratne , Yu Chen, Yarden Ne'eman, James Codella, Ching-Hua Chen, Deborah L. McGuinness, and Mohammed J. Zaki, **FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation**, Rensselaer Polytechnic Institute, Troy NY, USA, IBM Research, USA, (https://www.researchgate.net/publication/336599164_FoodKG_A_Semantics-Driven_Knowledge_Graph_for_Food_Recommendation)

[2] Официална документация на FoodKG, (https://foodkg.github.io)

[3] Официална документация на GraphDB, (http://graphdb.ontotext.com/documentation/standard/reasoning.html)

[4] Официална документация на owlready2, (https://owlready2.readthedocs.io/en/v0.36/intro.html)