RECIPE ONTOLOGY TERM PAPER

SUBJECT:

ADVANCE WEB ENGINEERING

PREPARED BY:

WALEED AHMED

MSCS71F21S010

AROOBA SULTAN TIWANA

MSCS71F21S012

MSCS (1ST)

SESSION 2021-2023

SUBMITTED TO:

MR. SAAD RAZZAQ

(COURSE INSTRUCTOR: ADVANCE WEB ENGINEERING)

DATED:

MONDAY, AUGUST 29, 2022

DEPARTMENT OF CS & IT

UNIVERSITY OF SARGODHA

SARGODHA, 40100

Semantic Web: Recipe Ontology

AROOBA TIWANA
MSCS71F21S012

WALEED AHMED
MSCS71F21S010

ABSTRACT:

Food research has gained significant growth in the past few years. Formulating new recipes is more of an art than a science. Machine Learning and commercial AI have proved their significance in many fields from the last few decades, including bioinformatics, medical imaging, and data mining. In recent, machine-generated recipes are not out of the question, and one can develop computer programs that can float novel recipes rich in a taste. The major hurdle for researchers in this domain is the heterogeneity of data sources in syntax and structure and the lack of machine-readable ingredients and instructions. We proposed a standard recipe ontology that can model all kinds of recipes and integrate different websites to bridge the heterogeneity across multiple sources. We also idented the ingredient type, value, units, actions, state of ingredient, time, temperature, and tools from the ingredients and instructions that would be a major milestone towards recipes evolution. We acquired one million recipes data from 350 different recipe websites. The data is used to define the extent of ontology and mapped on ontology to check the validity of ontology. The ontology also provides the substitution of ingredients. We developed the ontology in OWL language using the protege/VOWL tool. The ontology consists of 168 classes, 48 object properties, and 21 data properties.

INTRODUCTION:

Recipe is a collection of ingredients and methods that define how to prepare food. Ingredient is basically a food and used to produce food. It is the combination of food and quantity The instruction or method is a list of steps that describe how to combine the ingredients. Recipe can have multiple stages, where each stage has its ingredients and instructions. Recipes are classified into different classes according to ingredients, cuisine, course, season, and occasions, etc. Millions of recipe books are available that help people in preparing their favorite dishes.

The food domain has become an interesting research area, where researchers try to explore new techniques and use the latest technology. The research within the domain of food technology focuses on food processing from raw material to end products and the development of innovative food structures to fulfill the demand of the consumer for tasty, safe, and sustainable foods. The lack of common representation of data from distributed and heterogeneous systems and databases are the major hurdles for researchers. Also, the information available on the internet is available in different formats such as text, pictures, models, UML, spreadsheet, and many others. It makes it challenging to find an existing relationship between data sets. In this vein, the ontologies were created to help researchers to homogenize large and complex datasets and to link data based on its semantics.

What is Ontology?

The ontology is defined as the "formal explicit specification of a shared conceptualization"[7]. The description of these terms is shown in 1.1. In computer science, an ontology is a data model that represents knowledge as a set of concepts within a domain and defines a relationship between these concepts. It is a form of knowledge management; it captures the knowledge within an organization as a model. This model can then be queried by users to answer the complex questions and display a relationship across the organization.

Figure 1.1:Ontology Definition

Ontology is the backbone of the semantic web. It is the way of effective communication and data sharing. It helps users to achieve desired outcomes in a shorter searching time by semantically linking the data from multiple sources. Ontology is used to solve the problems of interoperability between applications. They can apply in any domain of research where data access and interoperability are required [8]. The use of ontology makes data less ambiguous and more Find-able, Accessible, Inter-operable, and Reusable (FAIR).

There are four main components of ontology:

1. Concept
2. Relations
3. Axioms
4. Instances.

A concept is a conception of a domain or task, which is arranged in a hierarchy. Relations are the interaction between objects of the domain. Axioms are assertions that are always true for the domain of Interest. These are coherent descriptions between concepts, properties, relations via logical expressions. Instances represent specific elements.

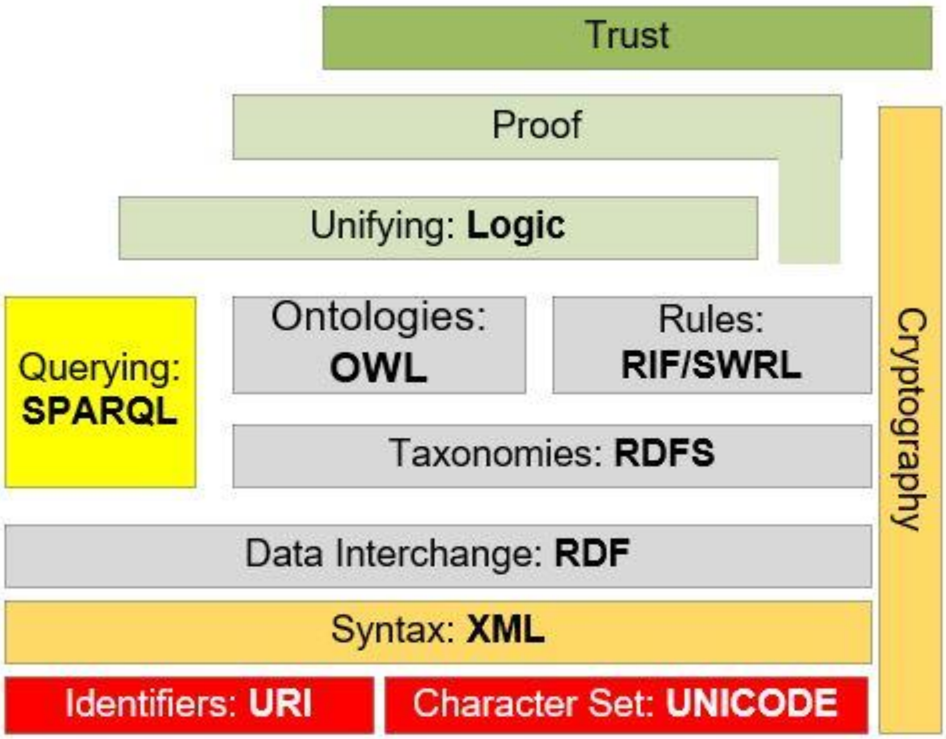

Figure 2:Semantic Web Stack

Reasoning:

Another feature of ontology is that it can be processed by a reasoner. Reasoner is software that is used to infer new facts from existing ones. It is used to reduce the redundancy in the knowledge base and helps in finding conflicts in the content. The two main services offered by the reasoner are:

Compute inferred class hierarchy: It is used to test whether a class is a subclass of another based on axioms defined in an ontology. Consistency Checking: It finds the consistency based on the description of a class.

Reasoner checks whether or not a class can have an instance. The ontology is deemed to be inconsistent if not possible to have any instance.

FOOD ONTOLOGIES:

BBC Food Ontology:

The BBC Food Ontology is a well-known food ontology that offers a small set of concepts related to recipes, ingredients, menus, and diets. BBC food ontology has been developed to publish data about recipes. It has a vast collection of terms related to recipes. The nutritional information about the recipes is not covered in this ontology.

FoodOn Ontology:

The FoodOn ontology is a more broadly scoped ontology that describes the commonly known foods that exist in nature as well as various food categorizations based on different standards. It is ideal for use in applications that involve food-related ontologies since the terms within it cover a broad range of domains in the _eld of food. It is developed using OWL. It consists of 31,394 classes, 430 individuals and 72 properties.

FOBI:

The FOBI (Food-Biomarker Ontology) describes food and its metabolites entities in hierarchical manners. FOBI is composed of two independent sub-ontologies that are interconnected. One is a 'Food ontology' based on FoodOn, and the other is a 'Biomarker' Ontology based on ChEBI. FOBI consists of 1334 terms, four properties, 13 top-level food classes, 11 top-level biomarker classes, and more than 13583 edges.

PROBLEM STATEMENT:

With the advent of machine learning in the food domain, the interest of researchers increasingly grows toward the automatic evolution of recipes. One can develop computer programs that can oat novel recipes rich in taste. The hurdle for researchers in this domain is the heterogeneity of sources in syntax and structure. Also, the lack of machine-readable ingredients and instructions is another problem in recipes evolution. Many recipe ontologies have been developed to provide a vast amount of recipes data, but all of them have some limitations. Such as, BBC food ontology only focuses on recipes and does not incorporate the dietary restrictions, and WhatToMake food ontology lacks the nutritional information and quantity of ingredients. Also, according to our best knowledge, no ontology works on recipes evolution.

CHALLENGES

We faced a few major and minor challenges in designing our ontology. These challenges are listed below.

Deciding classes and sub-classes: The first major challenge was finding classes and sub-classes for our ontology. For this, we got help from our 1 million recipe data set created by extracting the data from existing recipe websites. For example, we defined ingredient Type and quantity classes to map the ingredients used in a recipe.

Basic structure: Ontology structure defines the hierarchy of classes, sub-classes, constraints, and their relationships. To design the structure of our Recipe Ontology, we studied some existing recipe ontologies to get an idea of the structure.

Finding a tool: Many tools are there for ontology development. WebOnto, OntoEdit, Protege, WebODE, TODE etc. But protege is the best for ontology development.

Visualizing our ontology: Many tools are available used to visualize the classes and sub-classes with their attributes. BioOntoVis, OntoVis, OWLViz, OntoGraf, WebVOWL etc. But we needed to find a tool that could visualize classes and sub-classes with relations defined between them. Also, if there is a need to extend the ontology and requires changes, they should be made on the spot. So, we found VisualWebVOWL, an online tool that helped us a lot.

RESEARCH GAP AND MY CONTRIBUTION:

As we have already discussed that all the existing recipe ontologies developed for recipes have some limitations. Different ontologies lack different concepts or classes of recipes, such as dietary information, categories, nutritional information, and instructions. Also, there is no existing ontology that can model reviews, ratings, comment count, and dates information of recipes. To the best of our knowledge, no existing ontology identifies the actions and ingredients from the instructions. The primary contribution of this study is to develop a Recipe Ontology that would be able to map all kinds of recipes available on the internet and integrate data of websites in a structured form. We acquired a one million recipes dataset to develop and validate our ontology. It can help people to search their desired recipes. Another objective of this research is the identification of ingredient type, quantity, actions, state, time, and temperature from the ingredients and instructions for the evolution of the recipes. The ontology will also provide the substitution of ingredients.

DEVELOPMENT OF ONTOLOGY:

In ontology development, we developed an ontology capable of mapping all kinds of recipes available on the Internet. To that end, we designed a comprehensive ontology that includes all the possible classes, their relations, and properties needed for recipes. In the initial step, we used the recipes dataset to define classes, sub-classes, and relationships. We also reviewed existing foods and recipes ontology to understand their composition. Based on recipes data and reviewed data, the structure of the ontology is defined. A recipe is a broad term and can be viewed as an object with different attributes or properties. It is the central class from where the process starts. The basic information about the recipe includes its name, images of the dish, the URL of the recipe, author name, rating of the recipe, published date, modified date, created date, keywords or tags and the difficulty level to tell users whether it is easy to cook or not, etc. The recipes also contain classification information that helps to categorize it. Recipes are classified according to course, meal type, or the cuisine of a recipe. A recipe could belong to a class of snacks, side dishes, salad, main course, and many others. Recipe ingredients and instructions are the most important parts of the recipe. The ingredients are a combination of food and quantity. And the instructions are the list of steps to be performed on ingredients to make food or a specific dish. These actions involve series of steps, called instruction or process. Besides ingredients and instructions, recipes also contain time and tools needed to perform steps. The ontology structure has been finalized using these concepts. Basic concepts and relationships are shown in figure below:

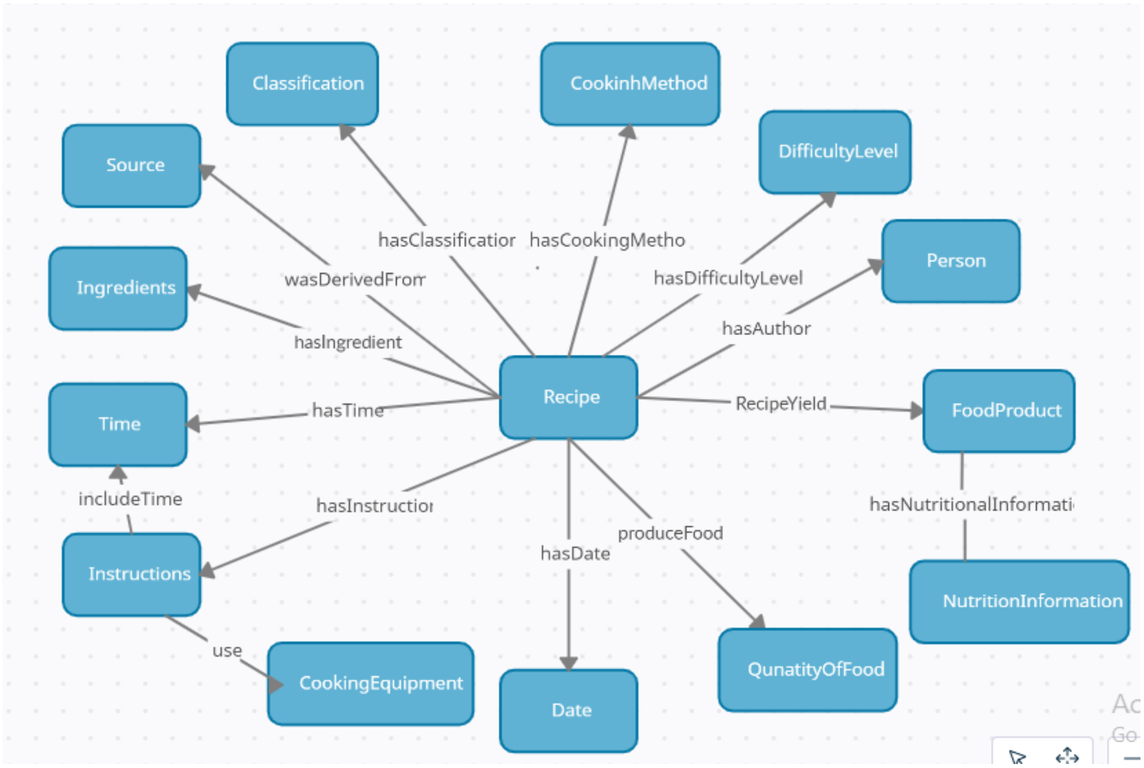

Figure 3: Basic Concepts and Relationships for Ontology

For the development of the ontology, we have studied different ontology languages, for example, RDFs, OWL, and OIL. But OWL is more expressive than other languages. Therefore, we decided to work on OWL. OWL is the extension of RDF and defines the classes, properties, and individuals in a more precise way. Unlike the other object-oriented languages, the OWL focuses on defining the relationship between classes based on semantic. There are two main properties of OWL that might be defined in an ontology: the object property that links two individuals and the data property that links individuals to the data value. Each property has domain and range. OWL provides the facility to create an association between classes and properties using property restrictions. Property restrictions are the constraints that all the individuals of class must meet.

There are three types of property restriction: Quantifier Restrictions, Cardinality Restrictions, and hasValue Restrictions. The quantifier restrictions are further categorized into universal and existential restrictions. The existential restriction describes at least one relation and is denoted by 'some' keyword, the universal restrictions define the class of individual that only have relationship to a member of specific a class and it is denoted by 'only' keyword, the Cardinality Restrictions describes the at least, at most, exactly, min, and max relationships, and the Value restriction gives specific value to individuals. These types of restrictions can be applied on both the object properties and the data properties. We defined the restriction properties for classes and inferred the membership of resources based on their relationship to other resources using the reasoner. Reasoner helps in finding the superclass and subclass relationship automatically. OWL also provides the annotation property, which is used to add descriptions or comments. Recipe Ontology (RO) was developed using the Protege software. It is a famous ontology development tool that provides a graphical visualization of the ontology.

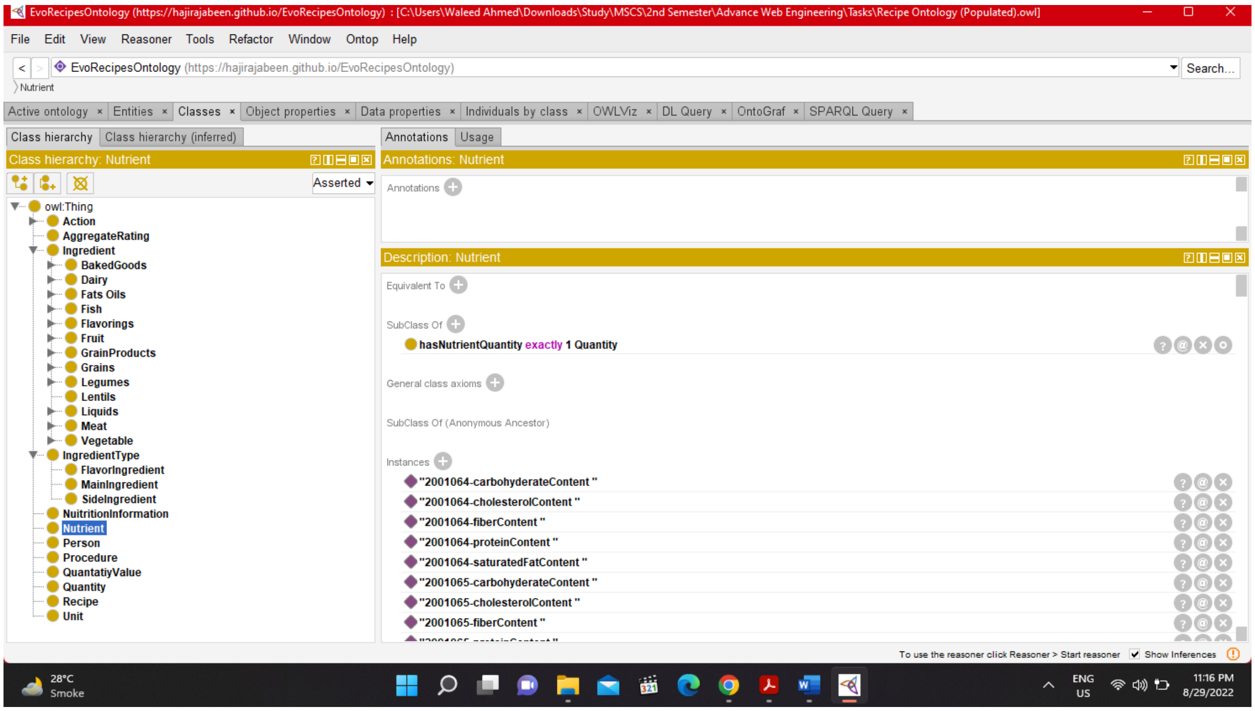

Figure 4: Recipe Ontology Classes

ONTOLOGY POPULATION

In third phase we worked on ontology population. The dataset of one million recipes was used to add the individuals of the classes. For example, we added all the possible types of nutrition like, calories, protein, fat, iron, etc. The ontology has been populated to enhance its vocabulary set.

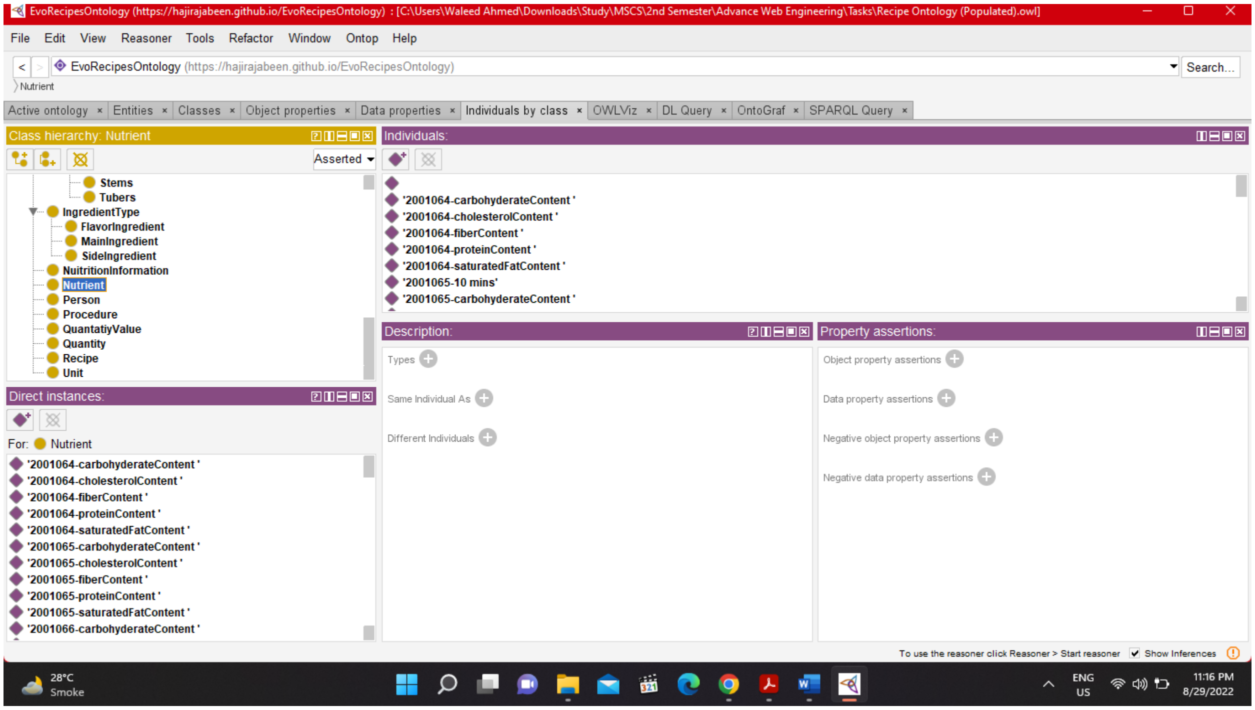

Figure 5: Populated Recipe Ontology Instances

COMPETENCY QUESTIONS:

To measure the completeness and accuracy of our ontology, we designed multiple competency questions. Here we will discuss if our ontology capable of answering all these questions.

1) What can I substitute for 1 tbsp all purposed flour?

As we have defined 'hasSubstitution' property in our ontology, that has ingredient class as domain and range as well. Each ingredient is divided into two parts type and quantity. So, our proposed ontology can provide the substitution of ingredients with quantity.

2) What can I use instead of cocoa powder?

The ingredientType class has a relationship with ingredient class, so when we add the 'hasSubstitution' property with the 'IngredientType' class, the ontology will infer it as an instance of ingredient class. Therefore, ontology can also substitute only ingredients without quantity.

3) American recipes for breakfast.

The ontology consists of "cuisine" and "MealType" classes, and the Recipe class has a relationship with both of them. Therefore, the ontology will provide all those recipes having both American cuisine and meal-type breakfast.

4) Thai recipes suitable for diabetic patients.

The ontology consists of the "SuitableForDiet" class to cover the dietary information and the 'RecipeCuisine' class for the cuisine of recipes. Both classes have a relation with the Recipe class.

Therefore, in the result of this query, the ontology will provide only those recipes that have both values in a recipe, Thai for cuisine and diabetic for SuitableForDiet.

5) Dinner recipes I can cook in less than 15 minutes.

Dinner is the instance of the "MealType" class, and the less than 15 minutes is the instance of the "RecipeType" class. The Recipe class has relation with both classes. Therefore, the ontology will be able to answer such queries. It will provide all those recipes that have cuisine dinner and can cook in less than 15 minutes.

6) What can we cook for breakfast with 2 eggs, cheese, and tomato for 4 people?

The ontology has "MealType", "Ingredient", and "QuantityOfFood" classes. The "Recipe" class has a relation "ofFoodQuantity" with the "QuantityOfFood," which provides the yield of the recipe. The "ingredient" class provides ingredient type and quantity, and the "MealType" class provides a type of meal. Therefore, ontology can address such kinds of queries as well.

VALIDATION (REASONER AND DL QUERY):

We validated some of these queries using the protege DL query tab. The validation of others has been done manually by checking whether the class was available and if they have proper relations. We have observed that our proposed ontology can answer all these queries as we have covered all the possible concepts of recipes and relationships between them.

COMPETENCY QUESTIONS FOR FUNCTIONAL REQUIREMENTS:

Q.1 Core-Recipe Related Competency Questions

Q.1.1 I am feeling sick and low energy. Let me know which soup recipes can be made in 30 minutes (i.e quickly or in less time)?

Q.1.2 What are diabetic friendly recipes?

Q.1.3 which are high protein vegetarian recipes?

Q.1.4 Which are the appropriate recipes for eggeterians?

Q.1.5 Which Italian recipes can be made easily?

Q.1.6 Which breakfast recipes are rated more than 1000 times(i.e mostly rated)?

Q.1.7 Which appetizers can be made using oven?

Q2 Ingredients Related Competency Questions

Q.2.1 Which recipes contain vegetables and exclude red meat & seafood (suitable for gout patients)?

Q.2.2 What are non milk (gluten free) recipes?

Q.2.3 What are wheat free (non celiac) recipes?

Q.2.4 Which deserts uses sugar in low amount (for pre-diabetic persons)?

Q.2.5 Which recipes contain leafy greens (suitable for reducing belly fat)?

Q.2.6 Which rice recipes have less number of spices?

Q.2.7 Which recipes use fish as the main ingredient?

Q.2.8 I have chicken, yogurt, spices, & olive oil at hand. Which recipes can be made using these available ingredients?

Q.2.9 which ingredients are needed for pulao?

Q3 Actions Related Competency Questions

Q.3.1 What are the non-fried recipes?

Q.3.2 Which are boiled9 recipes?

0.3.3 which fried recipes can I make using the fryer?

Q.4 Nutrition's Related Competency Questions

Q.4.1 Which recipes have low cholesterol (suitable for persons that have bad lipid profile)?

Q.4.2 What are most suitable iron rich recipes (for anemic persons)?

Q.4.3 Which recipes are rich in vitamin c[3] (for immunity boosting)?

Q.4.4 What are protein rich recipes?

Q.4.5 What are fiber rich recipes?

Q.4.6 What are low calories recipes?

Q.5 Procedures Related Competency Questions

Q.5.1 Which recipes contain baked chicken?

Q.5.2 Which recipes contain boiled noodles?

Q.5.3 What recipes contain fried chicken?

CONCLUSION:

In this research work, we proposed a recipe ontology that can model all kinds of recipes and integrate data of different websites to bridge the homogeneity across websites. The ontology also provides the substitution of ingredients. The proposed ontology can also identify the ingredients type, quantity, action, time, temperature, and state from ingredients and instructions. We acquired one million recipes data from 350 different websites. The dataset has been used for ontology development and validation. The proposed ontology consists of 168 classes, 48 object properties, 21 data properties, and 598 instances. We mapped the information of recipes presented in the Recipe dataset on ontology elements to check the validity of the ontology. We concluded that the proposed ontology could map all types of recipes. As discussed in the result section, the first two objectives which are the mapping of all kinds of data and ingredients substitution have been achieved impressively. For the recipe evolution, we identified the ingredient type, action, quantity, time, and temperature for each ingredient and instruction. But the generation of new recipes requires further work.

THANK YOU