Designing an Ontology for Managing the Diets of Hypertensive Individuals

A thesis submitted to the College of Communication and Information of Kent State University in partial fulfillment of the requirements for the Master of Library and Information Science and Master of Science dual degree program

> By Julaine Clunis January, 2016

Thesis Written By

Julaine Clunis

BSc., Northern Caribbean University, 2005

M.L.I.S., Kent State University, 2016

M.S., Kent State University, 2016

Approved By

Marcia Lei Zeng, Advisor

Jeffrey W. Fruit, Director, School of Library and Information Science

Amy Reynolds, Dean, College of Communication and Information

Table of Contents

List o	f Figures	v
List o	f Tables	vii
List o	f Acronyms and Abbreviations	viii
Gloss	ary of Terms	X
Te	ms Relating to Health and Medicine	. xii
Ackn	owledgements	xiii
Chap	ter I. Introduction	1
1.1	Problem Statement	2
1.2	Background: Medication and Food Interactions	2
	.2.1 Medical Nutrition Therapy	2
	.2.2 Hypertension	3
	.2.3 Medication and Food Interactions	4
	.2.4 Recipes	6
1.3	Information Seeking in Web-based Environments	7
	.3.1 Example Case Story	8
1.3	2 Search Engine Limitations	. 10
1.4	The Linked Data and the Semantic Web Potential	. 11
	.4.1 The Semantic Web	. 12
	.4.2 Ontologies and OWL	. 13
1.5	Objectives	. 14
Chap	ter II. Review of Literature and Related Applications	16
2.1	Drug and Nutrition Related Databases and Controlled Vocabularies	. 16
2.2	Domain Specific Ontologies	. 18
ź	.2.1 Ontologies for Managing Disease	. 21
Chap	ter III. Methodology	23
3.1	Task I. Data Collection	. 23
-	.1.1 Information Retrieval	. 24
-	.1.2 Web Scraping Limitations	. 26
	.1.3 Recipe Web Scraping	. 26
-	.1.4 Identifying Correct Pages	. 27

3.1.5 Identifying Correct Tags	
3.1.6 Content Tags	
3.2 Task II. Data Storage	
3.3 Task III. Designing the Ontology	
3.4 Task IV. Testing	
Chapter IV. Research Findings and Results	43
4.1 Overview	
4.2 Ontology Class Concepts	
4.2.1 Food Data Model	
4.2.2 Drug data model	
4.2.3 Person data model	
4.2.4 Recipe Data Model	51
4.3 Ontology Properties	
4.3.1 Object Properties	
4.3.2 Data Properties	55
4.4 Reasoners	56
Chapter V. Summary and Conclusion	58
5.1 Overview	
5.2 Discussion of Research Findings	58
5.3 Challenges and Limitations	
5.3.1 Data Retrieval Challenges and Potential Solutions	
5.3.2 Limitations of Study	
5.4 Future Studies	66
5.5 Conclusion	
Appendix	68
References	

List of Figure	es	Page
Figure 1.	Pinterest user content type interactions	1
Figure 2.	Explanation of basic components of recipes	6
Figure 3.	Example recipe with annotations	7
Figure 4.	Recipes selected for example case	9
Figure 5.	Search engine results demonstrating search engine limitations	10
Figure 6.	Semantic Web Technology Stack	12
Figure 7.	Example food ontology	14
Figure 8.	Locating correct tags using the Firefox Inspector	28
Figure 9.	Example of a triple	30
Figure 10.	RDF mappings for USDA	31
Figure 11.	Basic concept of domains and relationships for ontology	34
Figure 12.	Protégé Software showing ontology and classes	36
Figure 13.	Concept Map of Diet and BP Management Ontology	44
Figure 14.	Property restriction shown in the Protégé Ontology Editor	46
Figure 15.	'Food' class data model	47
Figure 16.	'Drug' data model	48
Figure 17.	'Person' data model	49
Figure 18.	Property Restrictions to show BP Level	50
Figure 19.	'Recipe' Data Model	51
Figure 20.	Food Preparation Methods	52
Figure 21.	Nut Allergy property restriction	52
Figure 22.	Object property with domain, range and inverse property set	53
Figure 23.	Restrictions on the range of values accepted by the data property	56
Figure 24.	Protégé showing reasoner inferencing recipes triggering shellfish allergies	59
Figure 25.	Protégé showing reasoner inferencing cholesterol interactions	60
Figure 26.	Protégé showing assertions for substituting ingredients	61
Figure 27.	Cooking subject terms	64
Figure 28.	Example recipe 1	68
Figure 29.	Example recipe 2	69
Figure 30.	Example recipe 3	70

Figure 31.	Subset of USDA food items and nutrients used in ontology	71
Figure 32.	USDA Abbreviated Excel file with all nutrients	71
Figure 33.	Excel file with extracted recipe data	72

List of Tables

Pages

Table 1.	Drug Interaction Categories.	.5
Table 2.	Web scraping techniques and descriptions	.25
Table 3.	Tags containing required information	.29
Table 4.	Example of hypertensive proto-persona	.40
Table 5.	Example of pre-hypertensive proto-persona	41
Table 6.	Example of normal proto-persona	42
Table 7.	Blood Pressure Levels	.50
Table 8.	List of Object Properties	54
Table 9.	List of Data Properties	.55

List of Acronyms and Abbreviations

ACE	 Angiotensin-converting enzymes
ARBs	– Angiotensin II receptor blocker
ASP	– Active Server Pages
BBC	– British Broadcasting Corporation
BP	– Blood Pressure
CSV	- Comma Separated Values
DASH	- Dietary Approaches to Stop Hypertension
FAO	- Food and Agriculture Organization
FDA	- Food and Drug Administration
FOAF	– Friend of a Friend
HBP	– High Blood Pressure
HTML	– Hypertext Markup Language
HTN	– Hypertension
HTTP	– Hypertext Transfer Protocol
ICD	- International Classification of Diseases
JSP	– Java Server Pages
KOS	- Knowledge Organization Systems
LCSH	- Library of Congress Subject Headings
MeSH	– Medical Subject Headings
MNT	– Medical Nutrition Therapy
NDC	– National Drug Code
NIH	– National Institutes of Health
OWL	– Web Ontology Language
OWL-DL	- Web Ontology Language (Description Logics)
PHP	– PHP: Hypertext Preprocessor
RDF	 Resource Description Framework
RXCUI	 – RXNorm concept unique identifier
SKOS-XL	- Simple Knowledge Organization System (eXtension for Labels)
SW	– Semantic Web
USDA	- United States Department of Agriculture

- URI Uniform Resource Identifier
- W3C World Wide Web Consortium
- WHO World Health Organization
- XML eXtensible Markup Language

Glossary of Terms

Definitions provided in this section are intended to help readers with the specific terminology used in this thesis. In addition, some of these have more general meanings or other contexts; the definitions given here are specific to the context of this study.

CARDINALITY CONSTRAINTS

- the number of instances of one entity that can or must be associated with another.

DATA MODEL

 a *data model* documents and organizes *data*, how it is stored and accessed, and the relationships among different types of *data*.

DATABASES

- a structured collection of data arranged for ease and speed of search and retrieval.

DATASETS – a collection of related sets of information that is composed of separate elements but can be manipulated as a unit.

DISJOINTNESS

- the state where an element cannot be an instance of two different classes

DOMAIN – a realm of knowledge about a specific unit of analysis for the construction of a KOS.

HYPERTEXT WEB

-a collection of documents (or "nodes") containing cross-references or "links" which, with the aid of an interactive browser program, allow the reader to move easily from one document to another.

KNOWLEDGE BASE

- the collection of knowledge expressed using some formal knowledge representation language and the underlying set of facts, assumptions, and rules that a computer system has available to solve a problem

KNOWLEDGE ORGANIZATION SYSTEMS

- the tools that present the organized interpretation of knowledge structures such as authority lists, classification systems, thesauri, topic maps, ontologies etc.

ONTOLOGY – a specification of a conceptualization

- **OWL** a family of knowledge representation languages for authoring ontologies
- **PERSONA** a knowledge management tool used to provide a prototypical approximation of a target user.
- **PROTÉGÉ** an ontology editor and framework for building intelligent systems.

PROTO-PERSONA

- a type of persona created in situations where there is not enough money or time to create a research based persona, typically based on assumptions about the target user.

- **RECIPE** a set of instructions for preparing a particular dish, including a list of the ingredients required.
- SCHEMA the structure of a knowledge organization system described with some formal language and refers to how the data is organized within that system.

SEARCH ENGINE

 computer program that searches databases and Internet sites for the documents containing keywords specified by a user

SEMANTIC WEB

- an extension of the current *web* in which information is given well-*defined meaning*, better enabling computers and people to work in cooperation.

SEMANTIC ANALYSIS

- a method for sharing and consolidating information in natural language format with the use of software to process that information, making it easier to find.

STRUCTURED DATA

- refers to information with a high degree of organization
- **TAGGING** nonhierarchical terms or keywords assigned to a piece of information

THESAURUS

-a structured type of controlled vocabulary that provides information about each term and its relationships to other terms within the same thesaurus

UNSTRUCTURED DATA

 information that either does not have a pre-defined data model or is not organized in a pre-defined manner.

DOMAIN KNOWLEDGE

 a valid knowledge used to refer to an area of human endeavor, or other specialized discipline.

WEB SCRAPING

- software technique for extracting information from websites.

Terms Relating to Health and Medicine

ANGIOTENSIN II

- a natural substance in the body that affects the cardiovascular system for example by narrowing blood vessels

HYPERKALEMIA

– medical term describing potassium level in the blood that is higher than normal.

HYPERTENSION

- abnormally high blood pressure

HYPOTENSIVE

- relating to or suffering from low blood pressure

ORAL CLEARANCE

- relationship between oral dose rate of drugs and the average state concentration

PHARMACODYNAMIC

 the biochemical and physiological effects of drugs on the body or on microorganisms or parasites within or on the body and the mechanisms of drug action and the relationship between drug concentration and effect.

PHARMACOKINETIC

- the time course of drug absorption, distribution, metabolism, and excretion.

RENIN ANGIOTENSIN SYSTEM

- a hormone system that regulates blood pressure and fluid balance

Acknowledgements

I would first like to thank God for sustaining me throughout this program and this particular project. Second, I would like to express my gratitude and appreciation to my mother, Cherine Clunis and also my host families here in the United States who have supported me, Pastor Ken & Cindy Ferguson and Mr. & Mrs. David and Christina Humble.

Finally, I would like to express sincere admiration and gratitude to Dr. Marcia Zeng for her unfailing support and encouragement and to Dr. Meehan for the pep talks. A special thank you also to Dr. Wei Fan for his help and to Dr. Karen Gracy for consenting to be a part of the thesis committee.

"This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health."

Chapter I. Introduction

Increasingly, food research takes place online, whether it's browsing for recipes on websites or using location services on mobile devices to find a restaurant in the area in real-time. In a recent survey of Canadian Internet users, over half of those surveyed said they searched the Internet for specific recipes. For women, the number was even higher (at 64%) (EMarketer.com, 2013). Another compelling example of Internet users searching for food related information comes from the popular website Pinterest, which has grown from 700K when it first came online, to almost 20M unique visitors in the last year – or about half the number of Twitter's unique visitors; among Pinterest users about 57% used the site to interact with food content (Caine, 2012).



Figure 1. Pinterest user content type interactions Source: (Caine, 2012), <u>http://blog.compete.com</u> (Accessed 2015-07-30)

To put this into context, consider that statistics show 87% of U.S. adults use the Internet. Seventy-two percent of those users have looked online for health information. Sixty percent of adults say they track their health, diet or exercise routine, and 33% track health indicators or symptoms such as blood pressure, blood sugar, headaches or sleep patterns. Around 21% of people use some form of technology to keep track of their health and 51% search the Internet

specifically for information related to nutrition and diet (Pew Research Center, 2013). The food we eat has an impact on our social lives but more importantly, it plays a huge role in our health. These statistics are a clear indication that a large number of people care about how the food they eat every day is impacting their health.

1.1 Problem Statement

Even as people spend more time searching for health information, more people are also suffering from chronic diseases such as heart disease, diabetes and hypertension. The problem to be addressed by this study is that of considering how semantic technologies can be applied to the management of chronic diseases and specifically hypertension. This research attempts to create an ontology that describes the domain and acts as a knowledge base to enable hypertensive individuals or health professionals monitoring those in this risk group to (1) identify and avoid foods that may worsen or trigger their conditions and (2) select condition-appropriate meals. The focus will be specifically on selection of recipes from the Internet and highlighting the risks involved with the chosen recipe, whether it be with a) the required ingredients, b) cooking methods, c) the nutrients in the ingredients or some other aggravating factor.

The researcher presents here an introduction to the issues under consideration, including a definition and explanation of key terms such as recipe, medical nutrition therapy, and hypertension. Furthermore, she discusses the issues surrounding food and health information research through search engines. Last, she introduces basic concepts of the Semantic Web and suggests its potential for solving these problems.

1.2 Background: Medication and Food Interactions

1.2.1 Medical Nutrition Therapy

What we put into our bodies plays a huge role in the state of our health and can influence our risk for many health conditions such as allergic reactions, digestive diseases such as celiac disease and irritable bowel syndrome, heart disease, hypertension, diabetes, and even cancer.

Medical nutrition therapy (MNT) is an approach to treating medical conditions like HTN using a diet devised specifically for the particular nutrition/metabolic condition and monitored by a registered dietitian or professional nutritionist (U.S. Department of Health and Human Services, 2001). Nutritionists are often tasked with educating patients about different foods and their properties, often recommending substitutes or the complete removal of some foods or

elements from the diet. Several factors besides the nutrition profile of foods may influence their suitability for an individual, such as age, gender, and general health condition. A study of nutrition can help us understand how and why certain nutrients affect us in the way that they do. A nutritionist is trained to understand these relationships between food and our health and as such can provide advice and solutions regarding the negative impacts of a particular diet.

An understanding of nutrition is no longer limited to just these health professionals. Nutrition is taught in schools, food products have nutrition labels often regulated by laws, and many databases with nutrition information exist. Some of these can be found on the web and dietitians, nutritionists as well as other non-professional persons are able to do research on various ingredients. Recipe websites and food packages have now started to let people know how much of different ingredients and properties are in the foods, but they each go about it in different ways. If an individual knows what and how much to eat as well as what not to eat, they can practice nutrition therapy on their own. Supported by a licensed health professional, these individuals could improve the quality of their lives and reduce the risks to themselves by eliminating unhealthy ingredients from their diets.

As people become more involved with managing their own health, they very often search for tools to help them with this task of finding appropriate foods to eat. Patients with hypertension may find the search and retrieval process to be a challenge due to the vast amount of information on the web. They often become confused or overwhelmed by the numerous recipes and products from which they must choose, while simultaneously being either unaware or unsure of any hidden, potentially harmful ingredients (food additives; colorings, stabilizers, and preservatives) as well as more healthy options they could choose.

1.2.2 Hypertension

High blood pressure (HBP), also called "hypertension" (HTN), is a medical condition that occurs when the force of the blood passing through the arteries is too strong and stays this way for prolonged periods. Uncontrolled HTN can lead to more serious conditions such as kidney failure, heart attack, stroke, and death (PubMed Health, 2015). Blood pressure is determined by the amount of blood your heart pumps and the amount of resistance to blood flow in your arteries; the more blood being pumped through the heart and the narrower the arteries, the higher

the blood pressure will be. Symptoms do not always present themselves to the patient, leading to unnoticed damage to the heart and blood vessels. Unchecked, HTN can result in more serious conditions such as heart disease, strokes, kidney failures, aneurisms, vision loss, trouble with memory and understanding, and metabolic syndrome, among other things.

There are two types of HTN: primary HTN which develops slowly over the years, and secondary HTN, which is often caused by some other underlying condition or cause such as obstructive sleep apnea, kidney or thyroid problems, certain drugs (prescribed or illegal), as well as certain birth defects. There are also certain risk factors such as age, race, genetics, weight, activity level, the use of tobacco, sodium and potassium intake, vitamin D deficiency, stress, and alcohol intake, which make HTN more likely to present in certain individuals.

Doctors may prescribe medication (which may have contraindication with certain foods) to the patients. Common medicines for treating hypertension include thiazide diuretics, betablockers, angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARBs), calcium channel blockers, and renin inhibitors (Mayoclinic.org, 2015). HTN can be also be managed through a combination of drugs and lifestyle changes, the most effective management technique being that of a change in the diet. Increased blood pressure is often associated with weight gain, which in turn causes disrupted sleep. Losing weight will help to lower blood pressure. Furthermore, certain minerals such as potassium and sodium have a direct effect on blood pressure and flow.

Managing the diet by paying attention to the amounts of certain nutrients present in the food, together with monitoring the amount of calories being consumed in an effort to lower body weight, are the preferred methods for management of hypertension. Some individuals have found that through lifestyle changes alone they have eliminated the need for prescription medications.

1.2.3 Medication and Food Interactions

What individuals eat and drink has the potential to affect the medicines that they take. Food at its most basic levels is made up of chemical compounds in much the same way as medicines. Some food and drugs, when taken simultaneously with medicines, change the body's ability to avail itself of the particular food or drug and sometimes causes serious side effects. Food/drug interactions can, a) cause a medicine to not work the way it should, b) cause a side effect from a medicine to be either worse or better, or c) cause a new side effect. Clinically significant drug interactions, which pose potential harm to the patient, may result from changes in pharmaceutical, pharmacokinetic, or pharmacodynamic properties. A drug interaction occurs when the food an individual eats affects the activity of a drug in such a way that the effects are increased or decreased, or they produce a new effect that neither produces on its own (Bushra, Aslam, & Khan, 2011).

Patients who are placed on antihypertensive drugs benefit from sodium restricted diets. Rich protein foods appear to be beneficial and a change in their diets from high carbohydrate/low protein to low carbohydrate/high protein can result in better oral clearance. Conversely, orange juice may inhibit the absorption of some beta-blockers, while grapefruit juice interacts with most medicines because it modifies the liver's ability to work the drug through the entire system. Licorice should also be avoided by people taking some medicines prescribed for HTN because of its potential for leading to sodium retention and potassium depletion (Bushra, Aslam, & Khan, 2011). Finally, potassium rich foods taken together with some medicines can lead to other serious medical conditions. These examples show how food interacts with health in very significant ways. Food interactions can be categorized into a range of levels.

Drug/Food Interaction Categories		
Contraindicated	Never use this combination of drugs/foods	
	because of high risk for dangerous interaction	
Major	Potential for serious interaction; Avoid high	
	intake of this food while taking this medication	
Moderate	Potential for significant interaction – generally	
	avoid moderate to high intake of substance	
	that interacts	
Minor	Interaction is unlikely, minor, or non-	
	significant, monitor for side effects	

Table 1. Drug Interaction Categories

Source: generated based on information from http://www.drugs.com/drug_interactions.html

There are also drug-induced nutritional deficiencies to take into consideration when formulating a plan for management of the disease through diet.

1.2.4 Recipes

Individuals with hypertension must manage their diets by carefully monitoring the type and amounts of foods they eat as well as paying attention to the nutritional content and preparation methods of those foods. This combination of factors is equivalent to what is known as a recipe.



Figure 2. Explanation of basic components of recipes

A recipe can be defined as a set of instructions for making food from various ingredients ("recipe", 2015). This means that in order for something to qualify as a recipe it must include these two basic components: a list of ingredients and instructions for their use. By this definition, this recipe on the left (Figure 2) for Mars Bar Cake does not qualify as a recipe based on the definition because it does not organize the ingredients into a list. Nor does it place the instructions into any recognizable grouping that would qualify as a set. To qualify according to the definition we would need to convert the text as shown on the right of Figure 2.

Recipes often include other information such as the quantities of the ingredients mentioned in the list; this is helpful for the proportion of the recipe, the taste, and the texture. It also allows one to compute the nutritional information for a particular recipe. Recipes often also include cooking and preparation times, instruments/utensils needed for cooking, as well as the

cooking method that must be utilized to prepare the dish. They also sometimes include information that help them to be placed into categories. These other pieces of information are helpful in judgment calls about the appropriateness of a particular recipe for someone trying to manage HTN.



Figure 3. Example recipe with annotations

Source: Annotated based on a recipe found from www.allrecipes.com

1.3 Information Seeking in Web-based Environments

When people make a query through a search engine they usually are shown a large number of results, which may or may not result in useful or relevant information and may leave them feeling more frustrated. Recipe searches in particular return thousands of hits owing to the large number of individuals who collect and post recipes online as a hobby as well as the countless dedicated recipe websites available on the Internet. Some recipe websites give users the option to search for recipes based on ingredients they have available and some even allow users to select recipes based on certain criteria for e.g. low calorie, gluten free, dairy allergies, et cetera. A system that can provide food recommendations about what to eat and what to avoid may help decrease the health risks and effects of having HTN by helping individuals understand exactly what nutrients are present in the foods they consume and how it could impact them. Moreover, the search results should clearly show the source of the data in order that patients may be able to trust it. Being able to convert this wealth of information into structured, usable data is crucial.

1.3.1 Example Case Story

To highlight the issues described above, consider the case of an individual with a specific health condition who is looking for a way to take the guesswork out of what recipes they can use safely. Although there are general guidelines about appropriate foods to be found, individuals have personal preferences in terms of what they like to eat or even what they feel like eating on any given day. Having a condition that limits one's choices makes this even more of a challenge.

Jane is a 49-year-old woman who has recently been diagnosed with HTN after blood pressure noted in doctor's appointments elevated to 186/110 mmHg on three occasions. She is 5'4" and her weight has been fluctuating between 280 and 300 lbs. Mary was initially given a prescription for 10 mg of *amlodipine besylate* daily. However, her BP continued to fluctuate and showed no apparent signs of improvement. That caused her doctor to change her prescription to 5 mg of amlodipine and 10 mg of *Lisinopril* daily. Her doctor has recommended that she follow the lower-sodium Dietary Approaches to Stop Hypertension (DASH) diet, an eating plan that emphasizes vegetables, fruit and low-fat dairy foods — and moderate amounts of whole grains, fish, poultry and nuts.) The DASH diet has been shown to lower blood pressure in almost all subgroups, whether those were defined by race, sex, age, body mass index, education, income, physical activity level, alcohol intake or hypertension. On this diet, Mary's daily sodium intake is 1500 mg. Mary's obesity is also a contributing factor to her condition, so she has a daily calorie limit of 1450 calories.

After a conversation with a friend in the medical profession, Mary realized some of her prescriptions had interactions with food. She then did research online and found out that *Lisinopril* had a moderate food interaction, which requires patients to avoid moderate to high dietary intake of potassium because of its potential to cause hyperkalemia in patients who use it. She would also need to pay particular attention to the potassium content of salt substitutes (Drugs.com, 2015a). In addition, she found that *amlodipine* also has minor food interactions with the consumption of grapefruit juice, which increases plasma concentrations of *amlodipine* and

inhibits the metabolism of CYP450 3A4 by certain compounds of the grapefruit (Drugs.com, 2015b). Although Mary doesn't quite understand all the medical jargon, she does realize that she should probably avoid grapefruit and foods high in potassium. Mary is making a special dinner to celebrate her husband's birthday and really wants to have an extra special meal but not aggravate her medical condition. She has some black beans and avocado on hand in the pantry and she knows for sure she wants to make some kind of crispy chicken dish because it's a favorite of his. She does a Google search for a recipe containing both avocado and black beans a nd finds this salad recipe and also some kind of crispy chicken dish. These both look good to her: the salad because it seems like a nice low calorie option with items she has on hand while at the same time being hearty, and the chicken because it is a baked dish, which she believes is a much healthier choice than a dish that requires frying as a cooking method.

On the surface both dishes look like good options, but there are problems that the average individual might miss. First, the fat and cholesterol content for this meal are high owing to the use of avocadoes and mayonnaise. Second, Mary can only have 1500mg of sodium per day but the chicken dish alone has 967 mg of sodium per serving. Furthermore, if Mary should have one serving of both of these meals together she will be consuming about 1093 calories, about three quarters her calorie limit for the day in one meal alone.



Figure 4. Recipes selected for example case

Source: Recipes from <u>www.allrecipes.com</u> and <u>www.food.com</u>

The biggest issue, and the one that Mary may not notice by just looking at the nutritional information, is that of the interactions with her medication. Mary needs to pay special attention to her potassium consumption and should carefully monitor dishes with high potassium content. The salad is a nice low calorie option perfectly in alignment with the DASH diet she has been placed on. However, both black beans and avocado are high potassium ingredients. In combination, the levels are significantly higher. Should Mary have the salad?

1.3.2 Search Engine Limitations

From our example case we gain an understanding of the myriad factors that can affect the recipe that Mary ultimately chooses to use or modify before use. Search engines, while able to aggregate millions of answers very quickly, do not present the answers in a way that is needed for her specific case. In the next example as shown in Figure 5, there is another case where the search engine does not specifically answer the user's question, "*Does eating canned food cause hypertension?*" First, the results are unstructured, so it is difficult to know which answer contains the correct information and which answers are completely irrelevant. Furthermore, the search engine retrieved over 3 million results. This is more than any one person will ever or would even attempt to wade through. When searching for critical information such as those relating to health and nutrition we often require more precise results. Search engines in general are keyword based and are unable to answer questions that need to be parsed for understanding, making it difficult to get the exact information needed.



Figure 5. Search engine results demonstrating search engine limitations Source: <u>www.google.com</u>

All patients can find a lot of confusing, misleading, or incorrect information on the web about what they can or cannot consume. As demonstrated by the results shown in the above search, there is a need for people to get more specific answers about whether or not a specific food would be suitable for them. Furthermore, for most people learning how to eat a healthy diet is not easy and many seek help from various sources in accomplishing this task. An intelligent agent that can automatically give people guidance for how to choose food that would aid their health and reduce health risks would be a great help. Searching through a collection of web pages is not yet an intelligent process that yields the kind of answers for which an individual may be searching. A good solution needs to consider the user's preferences, through some kind of personalization, resulting in recommendations based on the user's specific needs.

1.4 The Linked Data and the Semantic Web Potential

A major portion of web content was designed for humans to read, leaving computers with an inability to process semantics beyond basic parsing of layout, headers, and links in pages. Although there are many food and health websites in the current Web environment, the structure of their content is not uniform across all resources, thus creating difficulties for computers to do much more with the information than simple display of it to users who access it.

The hypertext Web, encoded using Hypertext Markup Language (HTML), defines how web documents are rendered when users access the page through the browser. Although these documents are linked, the links contain no information about the nature of the link or of any relationship between documents. The data in these documents are often either unstructured or semi-structured, and the information contained within the pages are limited to headings, paragraphs, lists, and other information about the structure of the document which makes for limitations in machine reading and understanding or interpretation of this information.

The idea of the Semantic Web is that of i) adding in meaning to documents such that better resource discovery is enabled, and ii) creating a common framework such that data can be shared and reused. Semantic Web technology can add layers of meaning to food and health information on the web, such that users may interact with the content in more intuitive and meaningful ways.

1.4.1 The Semantic Web

The Semantic Web, also referred to as Web 3.0, extends the current web to allow computers and people to work better together through the power of hypertext links that connect multiple items to each other. To accomplish this, computers should be able to access ontologies which contain machine readable semantic annotations that they can use to conduct automated reasoning and logic tasks (Berners-Lee, Hendler, & Lassila, 2001). Giving the web the ability to use logic, i.e., rules to make inferences, choose a course of action, or answer questions, is one of the major goals of the Semantic Web community.

Ontological languages powerful enough to express this logic, such as Resource Description Framework (RDF) and Web Ontology Language (OWL), allow individuals to annotate the Web using their own tags and to add whatever structure they deem appropriate to documents. In addition, OWL is able to represent concepts in ontological form and includes in the described concept a subset of instance data from the domain in question. These ontologies form a knowledge base which return results in a manner compatible with human communication but in machine-understandable format. In this way we see that the Semantic Web is a method for sharing and consolidating information in natural language format with the use of software to process that information, making it easier to find.

Semantic Web technologies can be used in various applications, including data integration, resource discovery and classification, and cataloging, as well as by intelligent software agents, and in content rating, among other things (W3.org, 2015b)



Figure 6. Semantic Web Technology Stack Source: (Parapontis, 2012)

The World Wide Web (W3) consortium maintains the standard for the semantic web framework shown in Figure 6 above. It was originally conceived by Berners-Lee and contains layers made up of identifiers, character sets, syntax, data interchange, taxonomies, ontologies, rules, queries, logic, proof, trust, and cryptography, which support user interfaces and applications.

1.4.2 Ontologies and OWL

An ontology defines a common vocabulary for researchers who need to share information in a domain. It has been defined as a formal and explicit specification of shared conceptualization (Gruber, 2009). It includes machine-interpretable definitions of basic concepts in the domain and relations among them. Ontologies are often developed for the purpose of sharing common understanding of the structure of information among people or software agents, enabling reuse of domain knowledge, making domain assumptions explicit, separating domain knowledge from the operational knowledge, and for the analysis of domain knowledge (Noy & McGuinness, 2001).

The Web Ontology Language (OWL) is used by applications that need to process the content of information instead of merely offering a representation of that information. Able to represent machine interpretable content on the web through an explicit representation of terms in a vocabulary and relationships between those terms, OWL is better suited for expressing semantic information than previous languages such as XML and RDF variants. Although RDF provided a data model for objects and relations between them, OWL offers the ability to describe classes and properties including such things as disjoint relations, cardinality, equality, better property typing capabilities, enumerated classes, and property characteristics. (W3.org, 2015a)

Ontologies are useful for various purposes including: providing a controlled vocabulary; for customization and personalization of search possibilities; as a structure that can be used for extracting document content; for word sense disambiguation; and, for semantic annotation of textual documents. Ontologies can describe both concrete and abstract objects, the set of which, described by a set of attributes, is referred to as a class. Attributes are the features and characteristics of an object; for example, the class *Person* may have the attributes of *age* and *name*. Relationships show how objects are related to each other; for example, the classes of *Person* and *Food* may be related through a relation such as *hasPreferenceFor*. Figure 7 below shows an example ontology of food.



Figure 7. Example food ontology

1.5 Objectives

The study attempts to resolve some of the issues discussed by answering the following research questions:

- 1. Can semantic technologies offer a solution for managing chronic diseases?
- 2. How can an ontology act as mechanism to provide guidance on safe foods and recipes, and answer specific questions that will improve health outcomes?

An ontology for recipes and the relationships between them, and the related concepts of ingredients, disease, and medicines, was developed with the aim of demonstrating how an ontology can be applied to existing datasets such as the USDA nutrient database, and the FDA drug database; combining and extending that knowledge base with additional domain concepts which seek to specifically aid individuals who suffer from and are seeking to manage hypertension through diet. Specifically, through the ontology creation process, the aim is to:

- define and model relationships between the different domains of interest.
- aggregate data from different sources to provide recommendations and answers to definite questions that cannot simply be plugged into a search engine
- design an ontology for recipes and the relationships between them and the related concepts of ingredients, disease and medicines.

- demonstrate how an ontology can be applied to existing datasets such as the USDA nutrient database, and the FDA drug database.
- present a case for utilizing ontologies to act as the knowledge base for intelligent systems which act as health aids.

It should be noted that this ontology does not simply describe recipes but also defines and models relationships that represent how specific recipes interact with a person with hypertension who is taking specific medications and has specific diet restrictions in place. The ontology aggregates data from different sources to provide recommendations and answers to definite questions that cannot simply be plugged into a search engine. Some examples of possible queries might include:

- Chicken dishes that do not use frying as the cooking method.
- Desserts with no added sugar and less than 200 calories.
- Pasta dishes with less than 200 mg of sodium.
- Vegetable dish which limits potassium.

The ontology could serve as the knowledge base of an app or website recommendation system which would take user input and run queries against the knowledge in the dataset, thereby aiding individuals to manage their hypertension.

This study describes the concepts, entities, and relationships specific to these particular concepts utilizing the Semantic Web and other tools to aid in the management of hypertension by leveraging the knowledge of users and their preferences, the nutritional information of different foods, ingredients in recipes, medication, and the disease itself as well as the relationships that can be created between these different data in such a way as to alert a user to what items are safe, what may be uncertain, and what should be completely avoided. Moreover, it will provide recommendations of recipes consistent with their needs.

Chapter II. Review of Literature and Related Applications

2.1 Drug and Nutrition Related Databases and Controlled Vocabularies

Information on food and medications is not restricted to labels found on the back of these items. There exist a large array of systems, which were developed for holding relevant information. Health professionals are able to access these sources to look up data necessary to their practice. This information is not always accessible or understandable by the non-health professional, however. In order to make the information available and accessible to them through the creation of systems that they can easily use, it is necessary to extract and present this information in a reliable and understandable format. As we design a solution, we consider a number of these systems, which have potential impact as we define the semantics of the ontology.

Some of the major sources of knowledge and concepts in the nutrition domain where nutritionists and private users can look up interesting facts about ingredients, nutrients, special diets, recommended daily values, and medications, among others, come from databases. These databases provide information related to this study that can inform design of the ontology structure.

- Drugs@FDA stores all the FDA approved drug products, both prescription and over the counter, as well as therapeutic biologicals, currently approved for sale in the United States. This entire database is available for download and use by interested individuals. Results of a search in the database include the names of drugs, their active ingredients, dosage forms and routes of administration and other information related to marketing and related companies. It provides tables of grouped therapeutic equivalents, over the counter drugs with similar ingredients, and links to documents and web pages with approval history, drug safety, and patient information (U. S. Food and Drug Administration, 2015a).
- The USDA Nutrient Database for Standard Reference is published by the United States Department of Agriculture and contains nutrient information for about 8,618 different foods. The database is searchable and provides full nutrient information for the foods stored within it, including calories, nitrogen-protein conversions, scientific names, and LanguaL codes. Users may run searches and download results as PDF or CSV files (USDA, 2011).
- *EuroFIR* is an international non-profit association whose purpose is to develop, publish, and exploit food composition. The EuroFIR maintains both databases and thesauri that draw

information from compiler organizations across the world; data includes nutrients, bioactives, and food allergens. This data guides the key individuals in the food industry in the production of healthier foods and also provides information to nutrition professionals, researchers, software developers, and consumers (Eurofir.org, 2015).

Health organizations and governments have created a variety of knowledge organization systems (KOS), covering a wide range of areas related to diet and prescribed drugs. Related to this study these KOS provide not only the terminologies and standardized coding systems, but also the ontological classes and properties to be considered.

- Agrovoc is a thesaurus or controlled vocabulary created by the Food and Agriculture Organization (FAO) of the United Nations (UN), which covers the interest areas of food and nutrition, as well as fisheries, agriculture, and environment. Currently Agrovoc is a SKOS-XL concept scheme and a linked open data set aligned with more than 10 other knowledge organization systems. Agrovoc may be downloaded in RDF core or LOD format for use in various applications (Datahub.io, 2015).
- MeSH Medical Subject Headings is the National Library of Medicine's controlled vocabulary. It includes subject headings for all medical conditions, with a set of terms naming descriptors in a hierarchical structure, permitting search. It contains over 27,455 descriptors and 220,000 entry terms that make it easier to identify synonymous terms. Headings range from broader terms to more narrow terms (National Institutes of Health, 2015a).
- The International Classification of Diseases (ICD) is the standard diagnostic tool for epidemiology, health management, and clinical purposes. It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases, and provides a general picture of the health situation of countries and populations (World Health Organization, 2015). Used by physicians, nurses, researchers, health information managers and coders, insurers, and other organizations to classify diseases and other health problems recorded on health records, the code set allows for 14,400 different codes and permits tracking new diagnoses. The core classification of ICD-10 is a three character code, which is a mandatory level of coding for international reporting; four character subcategories are often used in work-related settings, expanding the code set to 16,000 codes (Wikipedia, 2015a).

- The *National Drug Code (NDC)* is a unique product identifier for drugs in the United States intended for human use. All drugs commercially distributed must be identified and reported using the NDC. The National Drug Code is a unique 10-digit, 3-segment numeric identifier. The first segment, the labeler code, represents any firm that manufactures, repackages, or distributes the drug; the second segment, the product code, identifies the specific strength, dosage form, and formulation of the drug; the third segment, the package code, identifies package forms and sizes (U.S. Food and Drug Administration, 2015b).
- *RxNorm* is a standardized nomenclature for clinical drugs and drug delivery devices and a tool for supporting semantic interoperation between drug terminologies and pharmacy knowledge base systems. Produced by the National Library of Medicine, it is used by hospitals, pharmacies, and other organizations to record and process drug information. As these systems often use different systems to record and process their information, RxNorm provides normalized names and unique identifiers for medicines and drugs. RxNorm creates concepts that represent the drug names, attaches unique identifiers to the concepts, and then creates relationships between these concepts to link them together. It uses term types to indicate drugs at different levels of specificity (National Institutes of Health, 2015b).

2.2 Domain Specific Ontologies

Ontologies are a uniquely adapted way to build individual diet patterns and provide nutritional facts about a healthy diet. Ontologies are a formal explicit description of i) concepts in a domain of discourse (classes or concepts), ii) properties of each concept describing its various features and attributes (slots, or roles, or properties), and, iii) any restrictions on the properties. The combination of an ontology with the specific individual instances of the classes result in the creation of a knowledge base.

A relatively new technology, ontologies are fast becoming a highly preferred method to connect different types of information systems through the Semantic Web, though they suffer from the lack of a standard format that must be used for their development. The nutrition data found in the various databases, including those already mentioned, have been used in various ways to create food-related ontologies that provide functionality beyond those available in the databases. One example of such an ontology is the *Wine Ontology*, which describes the wine domain; it conceptualizes two main types of wines—red and white—as classes, and adds subclasses of each class that further represent the classifications that exist in that domain. This

study highlights the fact that ontologies should represent the classifications present in the real world in their design and structure (Noy & McGuinness, 2001).

Another wine ontology study focused on various features of the wine domain such as the fermentation process, characteristics, region, maturity of wines and classifications dependent on regions rather than on whether the wine was red or white (Graca, Pinto, Loureiro, Monteiro & Anunciacao, 2005). The researchers first considered reusing the previous wine ontology but found that rather than red and white, wines were classified according to three main dimensions, i) maceration, ii) grape maturity state, and iii) fermentation process. Each dimension was further classified and expanded before comparing and analyzing the results with the already existing wine ontology, allowing them to determine whether the ontology could be reused in whole, in part, or needed to be replaced by building a new one. In this instance they found that the existing concepts and relations in the ontology did not match what really existed in the domain and in literature and so needed to be redone. The aim of this ontology design process was to enable the selection of a particular wine, given a specific desirable set of features. This ontology design process demonstrated how other domains affect the domain of interest. For example, although the main focus was on wine, factors such as the region where the wine is produced could influence characteristics such as its taste and quality. This offers important guidelines for our research as we consider the relationship between diet and health, and on how medications can have interaction with food, which in turn affects health.

The *Food Product Ontology* describes the food product domain to help manufacturers, retailers, governments and institutions publish data in ways that maximize reuse. The designers of this ontology made the decision to reuse parts of a previously built ontology. They extended the *GoodRelations* ontology, a standardized ontology for product, price, store, and company data. It is used by many web pages to structure their data through the conceptualization of just four entities: agent, object, promise, and location, represented by the classes Business Entity, Offering, Product, or Location and Service (Hepp, 2008). The *Food Product Ontology* extends *GoodRelations* by adding such classes as food and ingredients with subclasses of food additives, e-additives, energy, carbohydrate, and other nutrient contents of the specific foods. They also further created and defined relationships between the existing and added classes (Kolchin & Zamula, 2013). Focused mainly on food products published on the web, this study shows how it

is possible to extend and refine a previously published ontology so that it can answer more questions users may have about objects in a particular domain.

Other general food ontologies include the *BBC Food Ontology* and the *BioNutrition Ontology*. The *BBC Food Ontology* was created for publishing data about recipes including the foods they are made from and the foods they create, as well as diets, menus, seasonal influences, courses, and the occasions they suit. This ontology is applicable to a variety of use cases across the web where recipe data are published (BBC, 2014). More focused on the medical field, the *Bionutrition Ontology* relates concepts and terminologies used for human nutrition in a clinical and biomedical setting. It consists of 100 classes, the major ones being Diet Plan, Protocol Development, and Research Nutrition Assessment Procedure, broken down further into subclasses representing concepts from nutrient compositions to specific diet plans for patients (National Center for Biomedical Ontology, 2015).

Some ontologies represent recipes, rather than focusing on specific products or food in general. Sam et al. developed a content ontology design pattern to allow integration of information from different recipe websites. Focus was placed on modeling concepts in the domain, such as recipe. Beginning with some natural language queries, they developed graph structures as a basis for the pattern and as a means of checking the validity of the queries. The developed model could be populated through the indentation and tagging of structures in XML documents to extract recipe information as semantic annotations (Sam, Krisnadhi, Wang, Gallagher, & Hitzler, 2014). Another design used an ontology to conceptualize the cooking domain; the model for this ontology was based on acquiring knowledge of the domain from cookbooks and led to the realization of 1,151 classes, 92 properties (52 of which establish relations), and 311 instances (individuals). The authors highlight the conceptualization process, breaking it into 1) identification of concepts and their properties, 2) classification of groups of concepts in classification trees, 3) description of properties, 4) identification of instances, and 5) description of instances. Relations between classes and decisions about whether a concept should be modeled as a class or an instance are important at this stage of the process (Ribeiro, Batista, Pardal, Mamede, & Pinto, 2006). A more complicated ontology representing recipes focuses on adapting recipes based on machine learning techniques and fuzzy logic to match and replace similar ingredients. This ontology holds 570 recipe ingredients and stores recipes in a XML database with 298 instances (DeMiguel, Plaza, & Diaz-Agudo, 2008).

More specific applications tend to focus on food recommendation and safe food consumption systems. For example, the ontology used as the framework for the *FOODS expert system* contains specifications of ingredients, substances, nutrition facts, recommended daily intakes for different regions, dishes, and menus to help find appropriate dishes based on user preferences. Using a bottom-up and top-down approach, the researchers first formulated questions they wished the ontology to answer, which in turn determined the scope of the ontology and the types of data to be stored (Snae & Bruckner, 2008).

In another case, the ontology-driven mobile safe food consumption system, *Food-Wiki*, was built with the stated purpose of helping individuals in risk groups avoid foods by helping them automate the decision-making process of what and how much to eat, as well as what to avoid, based on the nutrient profile and especially the presence or absence of potentially dangerous additives that have been added to many processed foods. This ontology used four classes that represented the concepts of diseases, person, ingredient, and product; it also contains 58 subclasses, 15 object type properties, and 12 data type properties. There were also 1,530 instances (individuals) and 210 semantic rules (Çelik, 2015). Another recommendation system, developed by Suksom, Buranarach, Thein, Supnithi, and Netisopakul (2010), helps individuals choose the foods they want to eat daily based on their nutritional requirements. This system combines two ontologies, one comprised of food and nutrition concepts and another based on a user's personal profile, diet goals and favorites. The system runs inferences between the two ontologies to generate rule-based food recommendations. This ontology consisted of 710 classes and 94 properties, modeled based on the concepts of person and food.

2.2.1 Ontologies for Managing Disease

Ontologies are also leveraged for the management of diseases. One study proposed an *automated ontology construction mechanism* that would more intuitively generate concepts, properties, relations, and restrictions; it was modeled against a nutrition composition dataset, which was broken down into 18 major food categories, and further classified into nutrient categories. Crafted around the diet requirements of diabetes patients, nutrients were assigned levels and rank to indicate the suitability of the food for the person. (Li & Ko, 2007). Finally, the *Food Ontology for Diabetes Control* used Protégé Desktop software (Stanford Center for Biomedical Informatics Research(BMIR)) to model concepts in the Eurocode2 food coding system that informs the class hierarchy and representing the types of food and their nutritional

content, and recommended daily intake. This ontology was later translated into OWL-DL where cardinalities, constraints, disjointness, and functional properties were defined and allowed to interface with the diabetes and product ontology through the designed interface; it contained 177 classes, 53 properties, and 632 instances (Cantais, Dominguez, Gigante, Laera, & Tamma, 2005).

Although there are several ontologies here that describe the domain of food, some were too specific, focusing on specific issues like wine or nutrients, some were too focused in that they were designed for specific applications, and others were not broad or deep enough. None of the ontologies researched included all the domains of interest to this study. One of the steps in the ontology building process is to consider reusing similar domains of interest. This is recommended as a method to increase interoperability between similar ontologies. Although none of the ontologies can be reused in the exact form in which they currently exist, from our review of the currently existing ontologies, there are possibilities for reuse or mapping to some classes.

This work adapts some elements from the previous mentioned ontologies but differs in that it focuses only on developing an ontology which can recommend and choose recipes appropriate for individuals with hypertension using the DASH diet and possibly taking medications to manage the disease. The ontology will be developed for the specific purpose of giving this author experience in the design and building of these types of knowledge organization systems. This ontology is intended for individuals and their family members who are trying to manage hypertension or those who wish to eat in a way that prevents hypertension. Instead of simply representing food concepts, it represents recipes and tries to help the user make appropriate choices. The study only considers the disease hypertension, including some medications prescribed for its management. It uses its own ontology and does not consider any specific applications, although the potential for the development of a recommendation system in the future is acknowledged.

Chapter III. Methodology

In this section the researcher describes the methodology that has guided this project. This study began with a pilot that saw the collection of recipes, semantic structuring of recipe and nutrition database information, followed by a process that enabled the development of an ontology for helping hypertensive individuals select appropriate recipes. This was followed up with the design of proto-personas that were used to test the proposed ontology. This study has been a combination of research and also a trial implementation of the methods learned.

3.1 Task I. Data Collection

Because a large number of individuals who search for recipes use either a dedicated food website or a favorite recipe blog to look for recipes, the pilot study began with the collection of recipes from the web.

There are countless recipes all over the Internet. Because documents on the web come from a variety of sources without any set standards to which they conform, differences will occur among the dataset. Some examples of different recipe pages can be found in the Appendix with descriptions. Since no absolute standard exists for how recipes are presented, we determined that we would select recipes that fit the following three criteria:

- 1. The recipes should match the definition presented here and include ingredients and their quantities and directions.
- 2. The recipe could be taken from any recipe website or blog.
- 3. The recipes should be somewhat structured, with the use of tags.

The first criterion was necessary for data quality purposes to ensure that all the recipes would have a similar composition in terms of the data that could be obtained from them and that an item was truly a recipe and not some other similar form of data lacking in the necessary features to actually make a complete recipe. For the second criterion, efforts were directed at food blog sites, selecting two on which to test our extraction method. The blogs that were chosen all used some form of tagging method, which was the third criterion, to structure and markup the recipe data; this stipulation was to ensure that it would be possible to identify the data that was needed.
With the increase in recipe websites, it became more important that search engines could crawl, index, and understand the data. Because the data is designed primarily for human consumption, however, machines often lack an understanding about the meaning of the information that is available. Many programming languages have been used to publish data to the web, including HTML, JSP, ASP, Javascript, PHP and others. Today, content management systems such as Drupal are being used and these also offer functionalities that make use of Linked Data technologies. This growth in the quantity of data has given rise to a need for the information to be stored, usually in large databases stored on servers. The programs that run on the servers will generate the web pages that we see.

3.1.1 Information Retrieval

Information retrieval can be defined as the tracing and recovery of specific information from stored data. In essence, we want to recover specific information from the vast amounts of data stored, and present only this to the user. By simply logging on to a website a user can use keyword searching or browsing methods to find and view documents, which they must themselves analyze to find the desired information. But owing to the non-structured format of the data on the web, this can be difficult. The Semantic Web as previously described creates a potential for machines to understand the semantics of the web.

One way of getting information from the web is through web scraping (web harvesting or web data extraction), a process whereby data is automatically collected from the World Wide Web using computer software techniques to extract pre-specified information. This simulates the process of exploring a recipe website document through the implementation of HTTP protocols or by embedding a browser. The result is a transformation of unstructured data usually encoded in HTML format into structured data that can be stored and analyzed in a database or spreadsheet. Transformation of this data into a machine-understandable format then makes it easier to query and identify specific results. Furthermore, it is an opportunity to take data from disparate sources, merge it with what is already known, and provide unique and purposefully fashioned information to the user.

Various web scraping solutions exist, ranging from those that require human effort to fully automated systems capable of scraping entire websites.

TECHNIQUES	DESCRIP TION
Human Copy & Paste	Manual human copy and paste of text, useful especially
	when sites block machine automation.
Text grepping	Information extraction using Linux's grep command or
	regular expression matching capabilities provided by
	programming languages.
HTTP programming	Uses socket programming to retrieve web pages by posting
	HTTP requests to the remote web server.
HTML parsers	Use of a wrapper to detect similarly encoded data, extract
	its content, and translate it into relational form.
DOM parsing	Embedding a browser to retrieve dynamic content and
	parse the pages into DOM trees.
Web Scraping Software	Software tools that attempt to recognize the structure of a
	page automatically with functions to store the scraped data
	into a database.
Vertical Aggregation Platforms	Platforms that create and utilize bots with no human
	involvement by establishing a knowledge base to harvest
	specific verticals.
Semantic Annotation Recognizing	Recognition and targeting of the metadata or semantic
	markup of web pages. In one case, scraping is done from
	annotations embedded in the page, e.g. microdata,
	sometimes considered a special kind of DOM parsing and
	in the second case, scraping of data from annotations
	organized in a semantic layer.
Computer vision web page	Using machine learning and computer vision to attempt to
analyzers	identify and extract from web pages by interpreting pages
	visually as humans do.

Table 2. Web scraping techniques and descriptions

Source: (Wikipedia, 2015b)

For this project, we use a form of Semantic Annotation Recognizing to harvest the recipe data. This information extraction technology allows recognition of named entities within text. The metadata can be of various kinds, ranging from the author of a recipe document, for example, to the amounts or kind of ingredients the recipe contains. Semantic Annotation is used to better describe knowledge contained in web documents and assign entities in the text link to their semantic descriptions.

The original plan involved collecting and storing recipes to Pinterest boards then manually saving, evaluating, and creating a database that could store them. The database structure could then be used to generate XML schema. However, this process was determined to be infeasible due to the sheer volume of content as well as the time it would take to manually store each recipe as well as to identify and separate the parts of the data. This led to a realization that machine-readable data, where the structure of the data related to the information contained within it, would be ideal. Machine-readable formats include CSV, RDF, JSON, and XML. Based on the end product needed, the decision was made to use web-scraping methods to extract structured recipe content which could be used.

3.1.2 Web Scraping Limitations

There are some circumstances under which web-scraping techniques will not work. These include cases where there is badly formed HTML code that does not contain any structural information; systems which prevent automatic access like the CAPTCHA authentication system; systems that are session-based and which use browser cookies; when bulk access has been blocked by the server administrator; and, in cases where it is prohibited.

3.1.3 Recipe Web Scraping

Most webpages are authored in a markup language of some form. Traditional markup tags present content, while additional metadata tags give information about the content. More recently, there have been new forms of markup that attempt to present the semantics of content. There are many tools available to help with web scraping and choosing the right one is largely a matter of personal preference as well as intent. A web scraper is simply a piece of code written in a language such as Ruby, Python or PHP. Scraping was accomplished in two steps:

- Step 1) Finding the right pages to focus on
- Step 2) Identifying the correct elements within these pages that contain the desired information.

In order to perform these two tasks one first needs to understand the structure of the website and database. For website display, HTTP and HTML are the technologies used most often, with HTTP being the way the browser communicates with the server and requests resources, and HTML being the language in which the website is composed.

3.1.4 Identifying Correct Pages

HTML pages are structured into a hierarchy of boxes defined by HTML tags. These tags have a multitude of tasks, from producing tables, images, or links to acting as unique identifiers, and they usually belong to groups referred to as a class. These classes are what allow for identifying and capturing individual elements in a web page. An exploration of the structure of the websites led to the discovery that the pages containing links to all the recipes were usually to be found under the category section, e.g.:

→*examplerecipes.com/recipes/ingredient/categoryname*

Once the correct pages were identified, it was possible to harvest all the uniform resource locators (URLs) for all the recipes on the site.

After the structure of the page was known, a web scraper which relies on Python was used. The script calls on the *urllib2* and the *BeautifulSoup* libraries to help parse web pages. The urllib2 module provides an updated API for using Internet resources identified by URLs and designed to be extended by individual applications to support new protocols or add variations to existing protocols (such as handling HTTP basic authentication). Beautiful Soup is a Python library used for pulling data out of HTML and XML files. The *soup.find_all* code will locate the href tags and pass them to the print function. These results are then pasted into an Excel file and cleaned up by removing all links that do not point to a recipe. The resulting CSV file was then used as the source of the recipe pages that needed to be scraped in the next stage of the scraping process.

3.1.5 Identifying Correct Tags

The inspector was again used with actual recipes to identify the tags that mark the information to be extracted, regardless of how the page looks. The *div* tag is commonly used to mark a large amount of content sections, with individual elements indicated using the *class* tag.



Figure 8. Locating correct tags using the Firefox Inspector

Table 3 gives an example of tags and the way information might look inside of them. Passed to the Python script that will first read the addresses from the CSV file created with all the URLs, the tags are used to first identify then extract the required information which is then dumped into a new CSV file.

Field	Tags – Example 1		
Recipe Name	<h2 class="fn">Pesto Lasagna Roll Ups</h2>		
Prep Time	<pre>20 minutes</pre>		
Cook Time	<pre>45 minutes</pre>		
Total Time	<pre>65 minutes</pre>		
Yield/Servings	Serves 12		
Ingredients	<pre><div class="ingredient"></div></pre>		
Method/Directions	<pre><div class="instructions"></div></pre>		
Field	Tags - Example 2		
Recipe Name	h2>Honey Mustard Chicken Recipe		
Prep Time	<span <="" class="preptime" itemprop="prepTime" th="">		

3.1.6 Content Tags

	<pre>content="PT00H10M">10 minutes<span class="value-</pre></th></tr><tr><th></th><th>title" title="PT00H10M"></pre>
Cook Time	<pre>span class="cooktime" itemprop="cookTime"</pre>
	content="PT00H45M">45 minutes <span class="value-</th></tr><tr><th></th><th>title" title="PT00H45M">
Total Time	Not present
Yield/Servings	<pre>span class="yield" itemprop="recipeYield">Serves</pre>
	4-6
Ingredients	<div class="ingredient"></div>
Marthaul/Discutions	

Table 3. Tags containing required information.

3.2 Task II. Data Storage

The next steps include modeling the recipes, the nutrition databases, and the drug databases for building the ontology. Ontologies rely on XML, XML Schema, RDF, RDF Schema, and OWL, where data can be defined and linked using RDF and OWL so that there is more effective discovery, automation, integration, and reuse across different applications. When a large dataset is published on the web it is usually in the form of tabular data, as is the case with the USDA food database. This database contains over 8,000 food entries with associated nutrition information for all possible nutrients. In the Semantic Web, RDF and linked data are advocated as a standard publication format for data integration and visualization. There are a number of tools available that aim to facilitate the process of "lifting of tabular data to reach semantically structured and interlinked data" (Ermilov, Auer, & Stadler, 2013). Because tabular data does not preserve domain semantics and structure, it is difficult to interpret, integrate, and visualize the data. This project used the Open Refine tool to semantically structure the information contained within the tabular data obtained from both the web scraping exercise and from the USDA dataset. This process is not a completely automatic process, as some editing was performed on the resulting RDF structure.

While the USDA dataset did not require any cleaning, there were a few things that needed to be done to standardize the recipe information so that it could be accepted into an ontology. Clean up tasks included: i) normalizing naming of recipes; ii) removing characters that do not convert well such as "", ~, and foreign characters; and, iii) trimming leading and trailing

spaces. Some textual data was also removed from rows that were holding specifically numeric data. The next step after cleaning the data was to create the structure of the RDF file to which we wanted to map the tabular data using the Protégé Editor program. RDF syntax is made up of triples:

- subject a consistent addressable URI for the resource
- predicate a URI that describes the relationship between the resource and the value of the triple
- object the value of the statement made in the triple, a literal value or another URI where the value of resource is another resource

In the following example, a triple for a recipe that has the title African Chicken Peanut Stew could look something like:



Figure 9. Example of a triple

The first requirement was to create a small ontology that that would match the structure of the data contained in the CSV file as well as data and object properties, and then to add the resulting file to the ontology as a prefix. Editing RDF mappings inside *OpenRefine* required uploading the RDF schema file to which properties and data will be mapped; setting the URI for addressing each resource; and, setting up prefixes for the data using a controlled vocabulary as well as a prefix for the vocabulary of the project. Finally all that remained was to set up mappings between the data contained in each column in the tabular dataset and the predicates from the ontology, as shown in Figure 10.

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node. Base URI: http://w3id.org/bpontology/ edit RDF Skeleton RDF Preview Available Prefixes: rdfs foaf owl xsd rdf foods +add prefix *manage prefixes foodgrpcd cell Shrt_Desc URI \square × >-foods:foodgroupcode> ×owl:NamedIndividual Shrt_Desc cell × ≫foods:fooditemname→ add rdf:type Energ_Kcal cell × ≫foods:energy_kcal→ × »-foods:protein a→ Protein_(g) cell × ≫foods:totalfat_g→ □ Lipid_Tot_(g) cell

Carbohydrt_(g) cell

Potassium (mg) cell

Fiber_TD_(g) cell
 Sugar_Tot_(g) cell

 \times >-foods:carbohydrate_g \rightarrow

× »-foods:potassium ma→

× »-foods:fiber_g→

× ≫foods:sugar_g→

Figure 10. RDF mappings for USDA

Add another root node

This same data was converted by OpenRefine and represented in RDF, as in this code sample for the food item salted butter.

```
<NamedIndividual rdf:about="&foods;butter_with_salt">
        <rdf:type rdf:resource="&foods;FoodItems"/>
        <foods:fiber_g rdf:datatype="&xsd;decimal">0</foods:fiber_g>
        <foods:sugar_g rdf:datatype="&xsd;decimal">0.06</foods:sugar_g>
        <foods:carbohydrate_g rdf:datatype="&xsd;decimal">0.06</foods:carbohydrate_g>
        <foods:protein g rdf:datatype="&xsd;decimal">0.85</foods:protein g>
        <foods:foodgroupcode rdf:datatype="&xsd;string">01</foods:foodgroupcode>
        <foods:ndb_no rdf:datatype="&xsd;string">01001</foods:ndb_no>
        <foods:gramwtdescrip1 rdf:datatype="&xsd;string">1 pat, (1&quot; sq, 1/3&quot;
high)</foods:gramwtdescrip1>
        <foods:monoFat g rdf:datatype="&xsd;decimal">21.021</foods:monoFat g>
        <foods:cholesterol mg rdf:datatype="&xsd;integer">215</foods:cholesterol mg>
        <foods:potassium_mg rdf:datatype="&xsd;integer">24</foods:potassium_mg>
        <foods:polyFat g rdf:datatype="&xsd;decimal">3.043</foods:polyFat g>
        <foods:saturatedFat g rdf:datatype="&xsd;decimal">51.368</foods:saturatedFat g>
        <foods:sodium mg rdf:datatype="&xsd;integer">643</foods:sodium mg>
        <foods:energy kcal rdf:datatype="&xsd;integer">717</foods:energy kcal>
        <foods:totalfat_g rdf:datatype="&xsd;decimal">81.11</foods:totalfat_g>
        <foods:fooditemname rdf:datatype="&xsd;string">butter with salt</foods:fooditemname>
        <foods:foodgramweight1>5</foods:foodgramweight1>
        <foods:foodgramweight2>14.2</foods:foodgramweight2>
        <foods:gramwtdescrip2>1 tbsp</foods:gramwtdescrip2>
    </NamedIndividual>
```

Save

The exported RDF output was then included in the ontology via an import statement.

3.3 Task III. Designing the Ontology

An ontology provides a precise vocabulary with which knowledge can be represented. This vocabulary allows us to specify which entities will be represented, how they can be grouped, and what relationships connect them together (Segaran, Taylor, & Evans, 2009).

The Ontology 101 guide presents a series of steps for building ontologies which suggests the following steps: a) determine the domain and scope of the ontology; b) consider reusing existing ontologies; c) enumerate important terms in the ontology; d) define classes and the class hierarchy; e) define the properties of the classes and slots; f) define the facets of the slots; and, g) create slots (Noy & McGuinness, 2001). Presented here is a description of the ontology development process:

Determine the domain and scope of the ontology: Already determined is the domain and scope in this document when we discussed: (1) the domains that the ontology would cover, (2) what we would use the ontology for, and (3) who would use the ontology. The following step in this first section involves identifying (4) the types of questions for which the information in the ontology should provide answers. To accomplish this, we created a list of competency questions that a knowledge base dependent on the ontology should be able to answer. These questions were used to determine whether or not the ontology has enough information and detail to answer the questions and was taken into consideration during crafting of the proto-personas (discussed further in section 3.4) to help inform their design and also was enhanced after that design process was complete.

Competency questions for the ontology proposed by this study:

- Will a recipe aggravate a pre-existing allergic condition?
- Does this recipe have high levels of nutrients in its ingredients that are known to cause interactions?
- If user is limited to 1500 mg of sodium daily, is the sodium content in that recipes ingredients too high?
- Are there any interactions between ingredients in a recipe and medication being taken?

• What is an alternative ingredient that could be used in place of an offending ingredient in a recipe?

Consider reusing existing ontologies: Although there exist some ontologies that describe various aspects of the domains in question, as previously discussed in the references, none address the issue in its entirety, therefore they were not included in the ontology. For example, although the food ontology for diabetes control (Cantais et al, 2005) focused on similar terms, it was modeled to match the Eurocode2 database instead of the USDA database. That ontology also does not concern itself with how much of the ingredient is used or make any calculations, furthermore, it is not publicly available for use. The automated ontology construction mechanism (Li & Ko, 2007) had a different focus and did not align well with our goals. The Linked Open Vocabularies (LOV) food ontology has the closest alignment with our purposes in terms of its description of the food concept. Instead of importing this ontology in its entirety, mappings were made using the *isdefinedby* annotation property to concepts that were similar. The major issues here come from subtle differences in interpretation of concepts. In future, perhaps the ontology modeled here might be redesigned to allow for a full import of the LOV food ontology with some modifications. The ontology created in this project has also reused some concepts from schema.org and FOAF, it borrows terms from Dublin Core, RXNorm, and SKOS, and uses the LCSH vocabulary to standardized terms related to recipe preparation methods.

Enumerate important terms in the ontology: Identifying the key terms that are used in the ontology provide a basis for development of a hierarchy. This informs the process of identifying what terms will eventually become a class, an object, or a data property. Key terms used in this ontology include names of food items as retrieved from the USDA database. There are also nutritional terms like proteins, fats, sodium, and so on, as well as names of recipes, cooking methods, diseases, and drug names and types.

Define classes and the class hierarchy: The class hierarchy is an attempt to organize the key terms and reflects the datasets as well as the domains in question.

Define the properties of classes and properties: Defining classes and properties required an examination of the terms and decisions on whether that resource was storing data values, describing relationships, or holding individuals. In order to preserve the integrity of the data

from the databases, there are many data type properties in the ontology. However object properties and property restrictions are created that make the ontology expressible.

Define the facets of the slots/properties: Cardinality constraints, and value restrictions have been defined in the ontology.

Create instances/individuals: Food types, recipes, drug items and classes have been modeled as individuals in the ontology.

A basic model of the concepts is shown in the diagram below.



Figure 11. Basic concept of domains and relationships for ontology

Ontologies are often expressed as RDF so that they are machine interpretable. The Semantic Web has standards such as the Web Ontology Language that extends RDF with more fine-grained concepts, allowing for the precise definition of ontological concepts such as classes, properties and individuals. OWL is similar to other object-oriented syntaxes in that it also defines classes and properties, however, it differs in that its focus is on defining semantics, i.e., relationships between entities. RDF, RDFS, and OWL offer methods for creating associations between classes and properties through reasoning rules using constraints or through inference (Hooland & Verbogh, 2014).

For this reason ontologies are more property oriented rather than object oriented. In the object-oriented model, the properties that belong to an entity are defined by class membership, whereas in a semantic model class membership is determined by the properties, which define that class (Segaran, Taylor, & Evans, 2009). A reasoner can then infer the membership of a resource to a class based on its relationship to other resources that share those properties. The *domain* distinguishes the relationship between classes and properties, and the *range* is the value that is expected or the data type of a property. OWL has two main categories of properties that might be defined in an ontology:

- Object properties that link individuals to individuals.
- Datatype properties that link individuals to data values.

OWL also allows data to be described with the annotation property, which is used to record information associated with some portion of the construct (W3.org, 2004). The Ontology was built using Stanford's Protégé Desktop Software (See Figure 12), a popular ontology that allows a graphical visualization of the ontology and imported some classes and properties from other Linked Vocabularies (such as schema:Recipe and foaf:Person. The OWL file was then validated with the use of the Pellet and FACT++ reasoners. Finally some editing was done by hand with the help of the Sublime Text editor.



Figure 12. Protégé Software showing ontology and classes

The ontology makes use of schema.org and FOAF to create an upper model by using classes from these vocabularies. This helps to create mappings between the ontology and others vocabularies on the Semantic Web. Once classes were established, properties from the tabular data sets that were converted to RDF/XML were added to the ontology and aligned with the classes which are their domain. For example the class Recipe and its properties that uses the format Class (*property*) as follows: -

Recipe(name, prepTime, cookTime, recipeYield, recipeInstructions, ingredientList)

A more detailed description and look at the model will be presented in the section Research Findings and Results.

3.4 Task IV. Testing

In an effort to improve the usability of the ontology, and to test and validate functioning, proto-personas were created. A review and analysis of the different domain concepts was undertaken, and the results were then used to create an outline of individual users that may be used for planning, design, and development. The resulting outline or skeleton was a bulleted list

of characteristic data ranges for each kind of user (Adlin & Pruitt, 2010). A persona is a representation of a user, based on user research and representing user goals, needs and interests. Persona creation is a process where data is summarized, clustered and analyzed in an effort to discover themes.

Personas are of three types: data-driven, institutional, and procedural. They are also described as having three layers of detail: requirements, relationships, and humanization (Brown, 2011). Traditional personas are heavily researched so that valid representations of the target audience can be created, which is usually an expensive and time-consuming process. One of the characteristics of a good persona is that it can help provide an understanding of a user's context, behaviors, attitudes, needs, challenges/pain points, as well as goals and motivation. Because of the limitations of this study, the proto-persona approach will substitute for the traditional type of persona development. Proto-personas are a provisional persona that is not a scientifically proven model of the target user. These differ in that they are not at first a result of user research but are created instead from brainstorming sessions that try to capture the creator's ideas about who is using the product and what motivates them to do so. After being created, these proto-personas can be then be tested against real users to validate their accuracy, and improved as any flaws are found. The design of the proto-persona is guided by a document with four quadrants: i) a sketch of the individual, ii) a name, iii) some basic demographics, behaviors and beliefs of the persona, and iv) needs and goals (Gothelf, 2012). The proto-persona is heavily based on the designer's beliefs about who the typical user might be, which makes this approach more feasible for smallscale research projects such as this study.

To add validation to the assumptions made about the user, the proto-persona creation process was preceded by an in-depth look at the domain in question. Focus was placed on three types of individuals: 1) the hypertensive individual on medication and with other risk factors; 2) the pre-hypertensive individual, not currently on medication; and, 3) the normal individual that wants to eat healthily to prevent HTN. An attempt was made to answer questions such as:

- Who is a typical hypertensive individual?
- Are there some characteristics that hypertensive individuals have in common?
- What kind of demographic information is necessary for the ontology to model the person?

- What kind of demographic information is usually taken from a patient whose complaint is hypertension when they go in to see the doctor?
- What kind of lifestyle and food choices does the individual make?
- Why would this person use a service like this tool?
- What goals would they most likely have for themselves?

Research shows that certain conditions, behavior, family history and other characteristics can affect the risk of HTN; these include conditions such as diabetes and prehypertension. Patients with prehypertension are at increased risk for progression to hypertension; those in the 130/80 to 139/89 mm Hg BP range are at twice the risk to develop hypertension as those with lower values (Chobanian et al, 2003). This leads to the belief that an individual who is pre-hypertensive will definitely be interested in lowering their risk through modification of their diets. Other risk factors mentioned include the kind of diet a person has; for example, ingesting too much sodium, not eating enough potassium, being physically inactive, obese, and using tobacco or alcohol. Furthermore, genetics and family history also indicate risk, especially when combined with lifestyle risk factors. Finally, age, sex and ethnicity also play a role in susceptibility to hypertension (Centers for Disease Control and Prevention (CDC), 2014).

Decisions regarding the types of information to be placed in the demographics quadrant came from an examination of patient information and medical history forms. Consideration of user likes and dislikes, and everyday eating and exercise habits should be undertaken. Furthermore, in some cases the user is not necessarily the individual using this application but a loved one, a client, or some other person with which a relationship exists and whose preferences and needs must also be considered. The most common types of information collected and those that are relevant to this tool are replicated in the proto-persona examples given. These factors all influence the brainstorming session and must be taken into consideration when designing a proto-persona, as they are likely to be elements that influence their choices and current health circumstance.

Brown's three layers of details – requirements, relationships and humanization – are reflected in each of the four mentioned quadrants. The requirements layer is implemented in the demographics and needs and goals quadrants, while relationships and humanization are reflected in the area of the quadrant that speaks to behaviors and beliefs. From this process, proto-personas are created using the prescribed format that circumscribe the questions above and generally

provide insight as to a) who the user is, b) the status of their health, c) their personal food preferences, d) what medicines they are taking, and e) the diet they have been prescribed by their health professional. An itemized list was created in each quadrant for each user and a consideration of the likes and dislikes of the personas, their allergies, and current medications, as well as goals, will be used to inform additional competency questions and refinements that can then be made to test and modify the ontology to ensure that it can answer their questions and similar user needs.

Proto-persona 1 - Hypertensive	Behaviors and Beliefs	
Mary	 Behaviors and Beliefs Cooks for her family everyday Likes to browse recipe blogs for new recipes Enjoys fried foods, cheese, and chips Does not take medicines at set times Forgets to check blood pressure regularly Exercises once or twice per week Believes that once the food product is whole grain, has diet, or low fat/fat free on it, it is healthy Does not like visiting the doctor Does not like taking medicines, sometimes skips 	
Demographics	Needs and Goals	
Age: 45-50	• Wants to reduce Blood Pressure without	
Sex: Female	medication	
<i>Weight</i> : 315 pounds	 Needs help choosing healthier recipes 	
Height: 5ft 4"	• Must limit potassium consumption to 3000	
Blood Pressure: >150/98	mg per day	
<i>Family History</i> : Hypertension,	• Needs to lose >100 pounds	
Waist 48"	• Wants to find recipes with healthier cooking	
Allergies: Shellfish	 Wants to reduce sodium consumption 	
Medication: Valsartan, Amlodipine	- wants to reduce sodium consumption	
Lifestyle/Psychological: Non-smoker, no		
alcohol use		

Table 4. Example of proto-persona showing a hypertensive individual

Proto-persona 2 – Prehypertensive	Behaviors and Beliefs		
Tom	 Software engineer Spends a lot of time sitting at desk Enjoys beef and pork, especially deli meats Eats out more than 3 times per week Does not exercise Dislikes fish 		
Demographics	Needs and Goals		
<i>Age:</i> 39			
Sex: Male	• Reduce BP by avoiding high sodium foods		
Weight: 180 pounds	 Avoid high cholesterol foods 		
Height: 6ft 1"	• Substitute unhealthy food ingredients with		
<i>Blood Pressure</i> : >130/84	healthier options		
Family History: Heart Disease	Stop consumption of alcohol		
Waist: 36"	• Prepare more meals at home		
Allergies:	-		
<i>Medication</i> : N/A			
Lifestyle/Psychological: Non-smoker,			
occasional alcohol use			

Table 5. Protopersona representing pre-hypertensive individual

Proto-persona 3 – Normal	Behaviors and Beliefs		
Pippa	 Exercises 5x per week Cooks everyday Recently married a vegetarian Housewife Reads a lot of food and health materials 		
Demographics	Needs and Goals		
<i>Age:</i> 28			
Sex: Female	• Wants to eat mostly vegetable based recipes		
Weight: 130 pounds	• Reduce fat and sodium intake		
Height: 5ft 4"	• Needs to know when dairy products are in		
<i>Blood Pressure</i> : <120/80	ingredients		
Family History: Hypertension, Diabetes	• Find out substitutes for dairy items		
Waist: 22"			
Allergies: Dairy			
Medication: N/A			
Lifestyle/Psychological: Non-smoker, no			
alcohol use			

 Table 6. Protopersona representing normal individual

Chapter IV. Research Findings and Results

4.1 Overview

This study presents an ontology as an example of semantic technology that aims to represent the domains of nutrition and health and how these can positively or negatively affect chronic illnesses, particularly hypertension. It demonstrates the use of ontologies to help individuals manage hypertension by a mechanism of highlighting risk factors in foods and recipes. In the beginning stages the study made a review of reference materials and other similar existing projects. Data collection efforts culminated with the conversion of the resulting CSV datasets into RDF with the use of a schema model that was built to accept data for both recipes and the USDA food database. Attention also was given to crafting the relations between the domains of interest. Finally, through small research efforts and brainstorming sessions, protopersonas (simple representations of the user) were created to represent the types of individuals who might use a system informed by this ontology. The resulting knowledge from this process is used to guide the kind of information that the ontology provides and what queries might be run against it. The study also demonstrates that a fairly standard procedure can be used to extract recipes from the web and further, that tabular data once transformed to a flat XML structure can be imported into the ontological model.

4.2 Ontology Class Concepts

Through the process of creating the *Diet and BP Management* ontology, determination of the domain led to the development of four top-level classes (i.e., classes that do not have a superclass except Thing): Food, Drug, Person, and Recipe. They are represented in the concept map presented in Figure 13. All other classes in the ontology are subclasses of these four concepts. The classes make use of terms from other vocabularies such as schema.org (prefix: schema), FOAF (prefix: foaf), Dublin Core (prefix: dc), as well as a few terms from SKOS, Library of Congress, and RXNorm. There are also terms that are unique to this ontology (prefix: dbp). Classes were added for concepts that were relevant to the domain and where properties and relationships were assigned. In total, there were 75 classes, 22 object properties, and 33 data properties. 3 classes and 11 properties were borrowed from schema.org and FOAF.



Some classes were also added because they enhanced the expressivity of the ontology, and were defined property restrictions. Property restrictions are a type of class that describes a class of all individuals that satisfy the restriction. Restrictions are usually of two kinds: value constraints (constraints on the range of the property) and cardinality constraints (constraints on the number of values a property can take). OWL restrictions usually have the form and are a subclass of owl:Class:

```
<owl:Restriction>
```

```
<owl:onProperty rdf:resource="(some property)" />
(precisely one value or cardinality constraint, see below)
</owl:Restriction>
```

Such restrictions can be applied to both object and datatype properties (W3C, 2004). The researcher will use data models and short descriptions to demonstrate each of the top-level classes.

4.2.1 Food Data Model

The "**FoodConcept**" class represents an abstract concept of food that involves food items and food groups. The FoodItems class has several subclasses that make up the 'USDAFoodGroups' class, and all inherit their properties from 'FoodItems'. Via these properties drug-food interactions and allergies can be associated with particular foods and nutrients. A data property is associated with each of the nutrients used. Constraints on the values associated with these classes are added by placing restrictions on the class. For example, to be considered part of the 'DairyandEggProducts' class, only food items with the *foodgroupcode* data type property with a value of 01 as expressed in the RDF code shown here. In Protégé (Figure.14) we simply define the class to be equivalent to the value of a datatype property. The reasoner will then infer what individuals belong to that class.



Figure 14. Property restriction shown in the Protégé Ontology Editor

A subclass called 'NutrientWarnings' was also defined to identify the food items that most often have interactions with drugs as well as those that hypertensive individuals are most often advised to avoid. There are several potential nutrition warnings that could be made for 'FoodItems' An example of this is with the class 'HighPotassium,' which has restrictions placed on it to display only food items that contain more than 200 mg of potassium. A model of this top-level concept is presented and uses dark orange boxes to show inferred classes based on property restrictions.



Figure 15. 'Food' class data model

4.2.2 Drug data model

Representing the Drug Concept in the ontology was a less formal process. First we used Drugs.com and Medline Plus to find information on the drugs commonly used to treat hypertension. We created a spreadsheet with information about these drugs, including the brand name, generic name, and the drug class type. Further research using RXNorm gave information such as RxNorm concept unique identifier (RXCUI) numbers, alternative drug forms, and

formulations, as well as a URI for the drug. The researcher did not include all of the drugs found in the database but randomly selected a subset of drug classes and some common medications prescribed that belong to that class, with the purpose of including enough examples of drugs to demonstrate that relationships between the medicines and recipes or food items is possible.





Drug/Food Interactions is an interesting concept to model. An examination of the subject suggests that interactions occur based on the body systems with which the drug interacts. For example, angiotensin receptor blockers (ARBs) are a class of drugs that affect the body's renin angiotensin system and specifically, angiotensin II. Drugs in this class selectively inhibit angiotensin II, which has several effects on the body by antagonizing angiotensin II-induced vasoconstriction, aldosterone release, catecholamine release, arginine vasopressin release, water intake, and hypertrophic response (Barreras & Gurk-Turner, 2003). Up to 10% of people who take these drugs have induced hyperkalemia and therapy usually includes assessing whether the patient is getting excess potassium from the diet, supplements, or some other factor (Raebel, 2012). The scope of this ontology, however, does not include the body or its internal systems and

mechanisms of action. For this reason, this researcher has simply chosen to model some classes of drugs, and then add information about their potential interactions via relationships between the drugs and foods. Furthermore, interactions are usually of four types—minor, moderate, major and contraindicated—as was discussed above. They are represented in the ontology as individuals, and with the use of reasoning about their relationships to drug items, inform the type of interactions medicines can have.

4.2.3 Person data model

The person class represents two kinds of individuals. The first is the author of the recipes. Person has as its properties *foaf:familyName, foaf:givenName* and *foaf:age*, which all subclasses inherit. The patient subclass represents the individual about whom inferences relating to their medical condition will be made. In the model we also see the 'BloodPressure' class inferred from the systolic and diastolic pressure readings of the patient.



Figure 17. 'Person' data model

An individual's blood pressure can be in one of four ranges: hypotensive, normal, prehypertensive, and hypertensive. Property restrictions are defined to represent the information in Table 7.

Classification	Systolic BP	Diastolic BP
Normal Blood Pressure	90 - 119	<i>and</i> 60 – 79
Prehypertension	120 – 139	or 80 – 89
Stage 1 Hypertension	140 – 159	or 90 – 99
Stage 2 Hypertension	≥ 160	$or \ge 100$
Isolated Systolic Hypertension	≥ 140	or < 90

Table 7. Blood Pressure Levels

Source: (Med-Health.net, 2015)

Restrictions and inferences are shown below in figure 18.

Annotations: Hypertensive		
Annotations 🕀		
rdfs:comment		080
Patient blood pressure readings >140/90		
Description: Hypertensive		
Equivalent To 🛨		· · · · · · · · · · · · · · · · · · ·
(hasDiastolicReading some xsd:integer[> 90]) and (hasSystelicReading some xsd:integer[> 140])	Hypertensive	?@XO
Sub Class Of	(Individual inferred)	
BloodPressure		?@XO
Patient		? @
General class axioms 🛨		
Sub Class Of (Aponymous Appestor)		
Instances 🕂		
◆ Mary		?@×
0		
Target for Key 🛨		
Equivalent To 🕂		
(hasDiastolicReading some xsd:integer[< 60])	Hypotensive	?@ X0
and (hasSystolicReading some xsd:integer[< 90])	,	
Equivalent To 🛨		
(hasDiastolicReading some xsd:integer[>= 60, < 90])	Normal	
and (hasSystolicReading some xsd:integer[>= 90 , < 130])		
Equivalent To		
(hasDiastolicReading some vsd:integer[>= 80 <= 90])	Prehypertensive	0000
and (hasSystolicReading some xsd:integer[>= 00, <= 50])	· · · · · · · · · · · · · · · · · · ·	0000
and (hass second during some Astanikeger [r = 150 / 3 = 140])		

Figure 18. Property Restrictions to show BP Level

4.2.4 Recipe Data Model

The 'Recipes' class is modeled to represent the data that was captured through web scraping. Recipes have several data properties associated with them but function in the ontology based on relations between them and other classes. All recipes have a cardinality constraint placed on them in that they must use a minimum of one food item in their creation. The major relationship between recipes and the other classes that has a bearing on an individual's health is the ingredients that they use. This relationship will allow the reasoner to make several inferences about the recipe, including whether it might trigger an allergy, has too much salt, is low calorie, or high cholesterol, etc.



Figure 19. 'Recipe' Data Model

Recipes have preparation methods stored in the ontology as individuals. This will allow for searching of recipes that, for example, do not involve frying, or recipes that do not require cooking.



Figure 20. Food Preparation Methods

The subclass 'MealType' represents the different types of categories recipes may fall under; for example, lunch or dinner. Added to 'MealType' is a special 'AllergyInducing' meal type. This class flags recipes that include items that may cause allergies such as shellfish, dairy, and nuts through the *usesIngredients* property and certain food categories.



Figure 21. Nut Allergy property restriction

4.3 Ontology Properties

The Diet and Disease Management ontology uses three types of properties to describe resources. If an entity can be identified with a URI then it is considered to be a resource. (Segaran, Taylor, & Evans, 2009). Object properties describe relationships, data properties store values, while annotation properties are used for documentation purposes and are not used by the ontology for reasoning. A resource is identified by the properties that are used to describe it. Properties usually have a domain and a range to indicate how they are used. A description regarding the number of properties has already been given.

4.3.1 Object Properties

Object properties, as previously mentioned, describe relationships between resources. For example, the *usesIngredient* property describes the relationship between 'Recipe' (domain) and 'FoodItems' (range). Some properties also have inverse properties, which create two way relationships in the ontology. Not only can one find out what recipes use certain food items, but a closer look at individual food items can reveal the recipes in which they are used.



Figure 22. Object property with domain, range and inverse property set

Object Property Domain Range isAllergic Patient Food *isPrescribedFor* Drugltem Patient *isTakingMedicine* Patient Drugltem mealTypeIncludes MealType Recipe isMealType Recipe MealType *dietAppropriateFoods* **DietaryNeeds** Food suitableForDiet Food DietaryNeeds RecipeAuthor Creator Recipe dislikedFoods Person Food hasDrugItems **MedClass** Drugltem hasInteraction *interactsWithDrugClass* FoodItems Drug hasInteractionLevel Drug Drug hasInteractionWith FoodItems Drug MedClass hasMedClassType Drugltem hasSubstitute FoodItems FoodItems ingredientUsedIn FoodItems Recipe *isCreatedBy* Recipe RecipeAuthor mayCause Medical Condition Drug mayTreat Drug Medical_Condition usesCookingMethod PreparationMethod Recipe usesIngredient Recipe FoodItems

Table 8 provides a list of all object properties with their associated domain and range specified.

Table 8. List of Object Properties

4.3.2 Data Properties

Data properties are used to store literal values; for example, the nutrient values of a certain food item. Following are a list of classes and their associated data properties. Properties exist as in the case of *foaf:familyName^*xsd:string*; where *foaf* is the ontology prefix, *familyName* is a data property and *xsd:string* is the type of data value assigned to the property.

Class	Property Name	Label	Value Expected
			(Range)
FoodItem	foodgroupcodevalue	Food Group Code	string
	ndb_no	Nutrient Database Number	string
	carbohydrate_g	Carbohydrates in grams	decimal
	sugars_g	Sugars in grams	decimal
	saturatedfat_g	Saturated Fats in grams	decimal
	monofat_g	Monounsaturated Fat in	decimal
		grams	
	polyfat_g	Polyunsaturated Fat in	decimal
		grams	
	potassium_mg	Potassium in milligrams	integer
	sodium_mg	Sodium in milligrams	integer
	energy_kcal	Energy in Kilocalories	integer
	Fooditemname	Food Item Name	string
	cholesterol_mg	Cholesterol in milligrams	integer
	gramweightdescrip1	Gram Weight Description	string
	gramweightdescrip2	Gram Weight Description	string
	foodgramweight1	Weight of Food in Grams	literal
	foodgramweight2	Weight of Food in Grams	literal
DrugItem	Medbrandname	Drug's Brand Name	string
	Medgenericname	Drug's Generic Name	string
	Cui	Concept Unique Identifier	string
	Rxcui	RXNorm Concept Unique	string
		Identifier	
Person	foaf:familyname	Family Name	string

	foaf:givenname	Given Name	string
	foaf:age	Age	integer
	diastolicReading	Diastolic Reading	integer
	systolicReading	Systolic Reading	integer
Recipe	schema:url	URL of recipe item	anyURI
	schema:name	Name or Title of recipe	string
	schema:recipeYield	Number of servings	integer
	schema:recipeInstructions	Recipe Instructions	string
	schema:prepTime	Preparation Time	integer
	schema:cookTime	Cook Time	integer
	schema:totalTime	Total Time	integer
	schema:ingredientList	List of Ingredients	string

Table 9. Data Properties for Blood Pressure ontology

Some data properties have value constraints applied so that they do not accept values outside the range that is set, as in the example below where constraints are set for systolic and diastolic blood pressure readings.



Figure 23. Restrictions on the range of values accepted by the data property

4.4 Reasoners

Reasoners infer logical consequences from a set of explicitly asserted facts or axioms and typically provide automated support for reasoning tasks such as classification, debugging, and querying. Knowledge in an ontology is not always explicit and reasoners provide the ability to deduct from the implicit knowledge the correct results (Abburu, 2012). Ontologies may have inconsistencies when applied, which indicates that there is an error somewhere in the ontology or

a conflict that prevents reasoning. If not caught, the ontology will have faulty semantic reasoning. Using a reasoner will greatly reduce the number of such conflicts in the ontology as well as ensuring that the ontology is consistent. There are a number of popular reasoners including FaCT++, Pellet, Hermit, RACER, and CEL that are commonly used to obtain facts from the ontology. This project mainly uses Pellet for reasoning with occasional support and checking from FACT++.

Pellet, like other reasoners, provides automated support for reasoning tasks such as classification, debugging, and querying. Pellet at its core is a description logic (DL) reasoner extended to OWL 2 and is optimized to reason in ontologies that include individuals. Pellet's built in query engine tries to efficiently answer ABOX queries expressed in SPARQL or RDQL through query simplification and reordering techniques. It also has a datatype oracle which it uses to check the consistency of built-in or derived XML Schema based datatypes such as those mentioned above, e.g., xsd:integer. Pellet supports user-created types based on numeric date or time types and further provides justification for understanding its reasoning output through axiom pinpointing. It allows the combination of DL with rules so that it is able to reason over OWL datatypes and SWRL rules as well as including an import function that allows for importing and reasoning over other ontologies without loss of context (Sirin, Parsia, Grau, Kalyanpur & Katz, 2007).

Studies have shown that tableau-based reasoners such as Pellet, when placed in situations with limited memory and small heap space, often fail due to memory exhaustion during the classification process. Pellet also performs more slowly in various timed performance tests compared to other reasoners. However, in its favor, it is readily available as a Protégé plugin and has an open source license. Furthermore, Pellet checks all the boxes of reasoning characteristics such as soundness, completeness, rule support, justifications, and ABOX reasoning tasks as well as having SROIQ(D) expressivity (Dentler, Cornet, ten Teije & de Keizer, 2011). The main reason for its use in this study is because it does incremental classification, in that if changes are made to the ontology it does not attempt to reclassify it in its entirety but only reclassifies those sections that are affected by the change, which was helpful with a very slow system.

Reasoners are implemented within Protégé by adding its plugin to the program files. This makes the reasoner available in the application and able to perform the reasoning tasks described above. Results from reasoning will be shown in the next chapter in a discussion of answers to the

competency questions.

Chapter V. Summary and Conclusion

5.1 Overview

This study developed a method for utilizing semantic technologies to manage chronic diseases. Using the case of hypertension, the researcher created an ontology that aims to show how the interaction among foods, recipes, and drugs impacts the management of the disease. More specifically, the study aims to use these relationships to generate specific recommendations or warnings for hypertensive individuals. A number of information resources and datasets were used to describe the domain and model the ontology. After creating schema structures, recipe and food data were converted from tabular form into RDF and added to the ontology. The project then focused on creating classes, relationships, and properties that the reasoner used to get answers as detailed by the competency questions. The Pellet reasoner was used to check the consistency of the ontology and to show the inferred reasoning.

5.2 Discussion of Research Findings

The research questions addressed by the study are whether semantic technologies offer a solution for managing chronic diseases and how an ontology can act as mechanism to provide guidance on safe foods and recipes. This was accomplished first, through a review of relevant research, and the study and creation of concept maps that attempted to describe the relevant domains. Second, web harvesting methods were used to aggregate data from different sources to provide recommendations and answers for the competency questions. This data was converted to RDF and imported into the ontology. Third, an ontology was created using the Protégé Desktop software for recipes and to establish relationships between them and the related concepts of ingredients, diseases, and medicines. In section four, the researcher presented a description of the class models. The data from the datasets were applied within those classes as instances. Assertions, constraints and relationships are used by the reasoner to make logical conclusions and to perform inferencing. Rules and queries can also be applied in the ontology to answer additional competency questions.

Based on the results shown here, we have clear evidence that ontologies can provide nutritional guidance for persons with chronic illnesses. Health, disease, and drug concepts have been described in classes and subclasses and through relationships and as demonstrated below, aid an individual using the system in obtaining answers to the kind of questions that would have returned ambiguous answers on the web.

Consider how the ontology addresses the competency questions created to test the ontology:

• Will a recipe aggravate a pre-existing allergic condition?

The ontology has property restrictions that identifies recipes containing foods that may trigger allergic reactions. These include shellfish, nuts, and dairy products.



Figure 24. Protégé showing reasoner inferencing recipes triggering shellfish allergies

• Does this recipe have too high levels of nutrients known to cause interactions in its ingredients?

Although the ontology does not yet allow calculation of total nutrient values, it gives information to the user about recipes that use ingredients with nutrient values that are high or that may cause interactions. For example, the ontology shows that the drugs in the class of ARBs have moderate interactions with high potassium containing foods. Since the individual can see how much of the nutrient is present, they can make a choice about whether or not the total amount consumed will be injurious.

• *Are there any interactions between ingredients in a recipe and medication being taken?* The ontology highlights all ingredients that have interactions with medications, whether those are interactions with cholesterol levels, potassium, grapefruit or otherwise.


Figure 25. Protégé showing reasoner inferencing cholesterol interactions

• If a user is limited to 1500 mg of sodium daily, is the sodium content in that recipe's ingredients too high?

High sodium ingredients are restricted to a class of similar individuals. Any recipe that uses ingredients with a sodium content higher than 450 mg will be flagged. The user can then decide if they may use that recipe after considering how much sodium they have consumed up to that point.

• What is an alternative ingredient that could be used in place of an offending ingredient in a recipe?

The ontology defines an object property that allows ingredients to be listed as substitutes for other ingredients. For example, olive oil is a substitute for salted butter in the ontology.



What is an alternative ingredient that could be used in place of an offending ingredient in a recipe?

Figure 26. Protégé showing assertions for substituting ingredients

From the examples and demonstration of concepts in the ontology editor shown above, there is clear demonstration of how the ontology answers directly through assertion or by inferencing answers to the competency questions. Other similar questions that may also be answered via the ontology:

- What are recipes that use low fat ingredients?
- What are recipes that use mostly vegetables or that don't include fish?
- What are recipes that don't use frying as a cooking method?
- What recipes will help me reduce cholesterol intake?
- What recipes will help me increase protein intake?
- What recipes will help me avoid hidden alcohol?

All of these questions can be formulated with description logics, semantic web rules, assertions, logical assumptions, and SPARQL since all of these concepts are described and modeled within the ontology.

5.3 Challenges and Limitations

5.3.1 Data Retrieval Challenges and Potential Solutions

When considering recipes, there are some issues that have to be worked through. For example, some but not all recipes have information about nutritional value; those that do usually represent it as nutrient per serving. Although some recipe websites contain nutritional information with each posted recipe, most food blog sites did not.

Of more concern, however, is that the recipes do not all represent similar information in a uniform manner. Consider the example of time: some recipes represent time in minutes only, hours only, a combination of hours and minutes, or even days. Sometimes, time is written in a variety of ways, for example:

"Bake for 3 hours and 20 minutes."

"Bake for 3 h and 20 m"

"Bake for 3:20 h"

"Bake for three hours and twenty minutes"

These differences necessitate manually going through the collected data and normalizing it. This requires deciding on a standard format that will be used, and using that format in all cases. For this study the researcher chose to represent time by specifically using the full term, such as 15 minutes, 2 hours, or 1 day, in all cases.

Another example is that of temperature for baking and cooking, which can be written either in Celsius or Fahrenheit. Which scale you use depends on what part of the world you live in. In the ontology, however, temperature is always represented in Fahrenheit. The final example is that ingredient and their quantities do not follow any set rules for how they are written. Some recipes use numbers such as '1', '2' or '3' while others use letters such as 'one', 'two' or 'three'; fractions get represented as '½', '0.5' or 'half' and other expressions such as 'pinch', 'dash', and 'to taste' also occur in the datasets. To complicate matters, the USDA also uses several sequences of measurements for individual food items. Since the focus is not so much on calculation of caloric and nutrient values but on interactions between them, this data was left as is, and imported as a string.

Ingredients are also described in different ways. First, recipes appear in a string of text that includes some quantity, unit of measurement as well as the ingredient item. Separating this information is not an easy undertaking. Aside from manually separating this information,

perhaps the use of natural language processing may help to highlight and isolate the actual ingredients, such as salt or oil. Second, some recipes may be more specific than others in describing types, parts, or varieties of a food. A recipe that uses chicken, for example, may say '2 lbs. of *chicken*' while another similar recipe might expressly state that chicken thighs are to be used. A recipe might also call for butter while not specifying whether this should be salted or unsalted butter. A recipe could require shredded cheese to be layered on the top before baking, but not tell the user what kind of cheese to use. This has an impact not only on the flavor profile of the resulting recipe but also on the nutritional content of the meal. Assigning a default ingredient solved this issue. For example, if a recipe calls for butter but does not specify the type of butter, however that recipe also includes salt, one might choose to use unsalted butter as the default instead. Another thing that can be done is to specify substitute food items. These established substitutions serve a dual purpose; first, they can indicate that a recipe could use either item, and second, they provide information to a user about a possible alternative food that may be used in place of an offending one.

Finally, cooking methods have some impact on the healthfulness of a meal for someone with hypertension, as some methods decrease the amount of fats that are needed in the preparation of the meal. The challenge with this is that cooking methods are all embedded within the instructions and there was no easy way to extract them. The researcher manually added these using standard cooking terms from the Library of Congress Subject Headings (LCSH), as seen in Figure 27.



Figure 27. Cooking subject terms Source:id.loc.gov/authorities/subjects.html

Issues such as those presented above are often encountered when working with large datasets and are a problem common to anyone dealing in big data. Creating and using controlled vocabularies, in combination with other knowledge organization systems, is one way of addressing this issue. Creating a standard list of terms is helpful for normalizing data, as instead of going through and making changes to every different item, relationships can be created through application of exact match, close match, same as, broader, and narrower term concepts that link the URLs of the resources to the URL of the resource which is the standard. This way, even if time is represented as hour, h, hh or hr, the system can know it is the same as the standard term hour. Finally, methods such as NLP and automated semantic entity extraction could be helpful to clean and normalize the data as well as enhance linked data potential through reconciliation of terms to services such as Freebase.

5.3.2 Limitations of Study

The ontology design was informed in large part from the researcher's efforts at understanding the domains in question. In particular, the proto-personas were designed after looking through research papers and patient health websites, and having informal talks with a nurse and a nutritionist. Although protopersona design is deliberately meant to be a simple and informal process that takes the place of actual research-driven personas, perhaps a more formal conversation with a nutritionist or with individuals who suffer with the condition may have helped to identify other needs and requirements.

Another limitation in the design and testing of the ontology was related to processor/computer capabilities, as the reasoner failed to reason over a large dataset in a timely fashion. For this reason, the researcher chose about 10% of the food database items and recipes. These 10% included all the food categories in the USDA database with the exception of restaurant foods, candies, snack foods, and some native food items that are not commonly used to prepare recipes. In the case of meat products, only raw items - a maximum of two types of similar items example, a lean and full fat version were selected. In most cases the raw, fresh or uncooked version of the ingredient was chosen and in the case of vegetables and fruit, only raw, frozen and basic canned versions were included. Finally, ingredients that were not mentioned at all in the recipes and that were also imported or rare ingredients not commonly used in cooking applications in the U.S. were also removed. There do not seem to be major detrimental effects from not using the complete datasets. Food ingredients and recipe properties are the same throughout so major results would not be impacted. The one relationship that is concerning is that substitute ingredients are limited as not all ingredients are included.

The ontology uses many data property fields with values assigned. Currently the ontology will inform users about individual items in the recipes and any interactions detected. However, the intention was to use these fields to perform calculations about the nutrient content of entire recipes. One of the issues that prevented this was the fact that the dataset contains several sequences of measurement units, making attempts to calculate calories or approximate daily intakes unachievable at this juncture. Another issue is the researcher's unfamiliarity with semantic web rule languages that might make these calculations possible. However possibility exists that with better modeling and a better understanding of rule languages, this functionality could be implemented.

Moreover, the ontology does not yet include all nutrients available in the USDA records. This decision was taken because of the sheer number of nutrients included, and the aforementioned processor limitations. Because of the scope of the project, the focus was placed only on those nutrients that directly impact hypertensive health, have common drug/food interactions, and are most commonly included in the nutrition information available with recipes and food products. Including more nutrients would allow us to expand the ontology to include more diseases and interactions and is something to consider for future study.

Finally, although a wealth of information about ontologies exists on the web, ontologies devoted to just food and recipes and focused on health are few and for the purposes of this study even more limited because we could only consider those available in the English language. Although there are various nutrition and medication databases available, this study uses only the data contained in the USDA National Nutrient database for food items and the RXNorm and Drugs.com for information on drugs. It should also be noted that this study makes no attempt to categorize all available medications but focuses only on a subset of those prescribed for hypertension in the United States.

5.4 Future Studies

As identified in the discussion of limitations of the study, there is great potential for using ontologies as a knowledge base. Based on observations and experiences during the process, the researcher believes that a more formal research process with nutritionists as well as target individuals would yield a better understanding of their needs. Second, there is room for expansion and better modeling of the ontology to yield even more answers about the domain.

Another hope is to employ the use of rule languages to enhance the ontology's ability to answer questions about recipes based solely on the information about the food items used. Focus could be placed on the issue mentioned with the several sequences of measurements found in the food database and in the recipe structure. The researcher believes that if this is done, it will be possible to calculate the exact nutrition information for any recipe. Including more nutrients and drugs in the ontology will expand the knowledge base and provide further knowledge that may be able to answer questions about other disease and drug/food interactions. Finally, the ontology in its current state is not very user friendly, requiring knowledge of the software and of OWL to create answers. Future studies could include the design of a web application or app that will be intuitive and easy for users to learn, and which will communicate with the ontology via an API. Another application envisioned would have the ontology being integrated with a food diary application where users would be able to track nutrient intake and impact in a much more detailed fashion that is currently implemented by similar diary apps available now. This researcher hopes to continue working on this project and turn it into an actual functioning project that can benefit the target audience.

5.5 Conclusion

The findings and discussion presented here support the idea of semantic technologies as a method for the management of chronic diseases through the use of ontologies and other KOS modeled and applied to the needs of the target user. Through modeling of the target domains into hierarchies that represent individual medical and nutrition-related concepts, creation of relationships between these concepts, linking to external resources, aggregation of the necessary data, and the creation of rules and assertions, as well as reasoning and inferencing, the ontology is able to answer questions that will offer guidance to users and improve health outcomes.

Appendix



Pan-fried London Broil Steak Recipe

Cook time: 25 minutes Yield: Serves 4

We recommend using a well-seasoned cast-iron frying pan for this recipe, which can take high heat and are relatively stick-free. If you do not have a cast iron pan, you can use a thick-bottomed frying pan. If using stainless steel, heat a little canola oil or olive oil in the pan first, before adding the steak.

INGREDIENTS

- 2 lb top round cut of steak
- Kosher salt
- Dry mustard
- Pepper
- Butter, softened to room temperature

METHOD

1 Remove **steak** from refrigerator 2 hours before cooking to bring to room temperature (only do this with whole cuts of meat, never with ground meat.) Cut away any tough connective tissue on the surface of the steak. Use a meat pounder to even out the thickness of the steak if necessary. Lightly sprinkle with **kosher salt** on both sides.



2 Heat a large, cast iron skillet to medium high heat. Pat the steaks dry with paper towels. Rub a little dry mustard into both sides of the steak. Sprinkle both sides again with salt, and with a little black pepper. Rub butter over both sides of the steak.



3 Place the steak in the hot pan. Let cook for 2-3 minutes on each side (without moving), check before flipping to make sure it has nicely browned.



Figure 29. Example recipe 2

Print 🖶

340 18

TOTAL TIME 1hr

PREP 20 MINS COOK 40 MINS



This is for serious garlic lovers. The chicken bakes out so moist and tender! Grated cheese may be sprinkled on top of the breast the last 5 minutes of cooking if desired but that is optional. Double all ingredients for more than four

More

Skip to Next Recipe 🕹

INGREDIENTS Nutrition DIR Preh SERVINGS 4 Butt 1/3 cup butter, melted brea 2 tablespoons minced garlic (can In a use more or less) garli In ar 2 teaspoons garlic powder, 1/2 cdivided (garlic lovers can use chee more) blac 1/2 teaspoon seasoning salt (or can Dip c use white salt) Place minu 3/4 cup seasoned dry bread crumb take (seasoned or plain) will 1/2 cup finely grated cheddar Top with shred cheese 1/4 cup freshly grated parmesan cheese 1/2 teaspoon ground black pepper (or to taste) 4 boneless skinless chicken shredded cheddar cheese (optional and use any amount desired, or can use shredded mozzarella cheese)

ECTION	e	
eat oven t	Nutrition Info Serving Size: 1 (185 g)	C
sts use a li	Servings Per Recipe: 4	
owl, comł c, 1 teaspo	AMOUNT PER SERVING	% DAILY VALUE
other bow	Calories 452.1	
up finely g	Calories from Fat 235	52%
se 1 teasp	Total Fat 26.1 g	40%
pepper.	Saturated Fat 14.8 g	73%
hickon in k	Cholesterol 136.7 mg	45%
nicken in t	Sodium 852.6 mg	35%
in prepar	Total Carbohydrate 18	.6g 6%
tes or unti	Dietary Fiber 1.4 g	5%
more time roduce an	Sugars 1.5 g	6%
rounde un	Protein 34.9 g	69%

minutes of cooking (this is only optional.

Up Next



Foolproof Weeknight Chicken (2:04)

_	A	B	C	D	E	F	G	H		J	K	and the subscription of the	M	N	0	P	Q	R
1	NDB_No	Shrt_Desc	Energ_Kcal	Protein_(g)	Lipid_Tot_	Carbohydrt	Fiber_TD_(Sugar_Tot_	Potassium	Sodium_(m F	A_Sat_(g)	FA_Mono_	FA_Poly_(g	Cholestrl_(GmWt_1	GmWt_Desc1	GmWt_2	GmWt_Des F
2	01001	BUTTER,WI	717	0.85	81.11	0.06	0	0.06	24	643	51.368	21.021	3.043	215	5	1 pat, (1" sq, 1/3" hij	14.2	1 tbsp
3	01002	BUTTER, WI	717	0.85	81.11	0.06	0	0.06	26	659	50.489	23.426	3.012	219	3	1 pat, (1" sq, 1/3" hij	9.4	1 tbsp
-4	01003	BUTTER OII	876	0.28	99.48	0	0	0	5	2	61.924	28.732	3.694	256	12	1 tbsp	205	1 cup
5	01004	CHEESE,BLU	353	21.4	28.74	2.34	0	0.5	256	1146	18.669	7.778	0.8	75	28	1 oz	17	1 cubic incl
6	01005	CHEESE, BR	371	23.24	29.68	2.79	0	0.51	136	560	18.764	8.598	0.784	94	132	1 cup, diced	113	1 cup, shre
- 7	01006	CHEESE, BR	334	20.75	27.68	0.45	0	0.45	152	629	17.41	8.013	0.826	100	28	1 oz	144	1 cup, slice
8	01007	CHEESE,CA	300	19.8	24.26	0.46	0	0.46	187	842	15.259	7.023	0.724	72	28	1 oz	246	1 cup
9	01008	CHEESE,CA	376	25.18	29.2	3.06	0		93	690	18.584	8.275	0.83	93	28	1 oz		
10	01009	CHEESE,CH	406	24.04	33.82	1.33	0	0.28	76	644	19.368	8.428	1.433	102	132	1 cup, diced	244	1 cup, melt
11	01010	CHEESE,CH	387	23.37	30.6	4.78	0		95	700	19.475	8.671	0.87	103	28	1 oz		
12	01011	CHEESE,CO	394	23.76	32.11	2.57	0	0.52	127	604	20.218	9.28	0.953	95	132	1 cup, diced	113	1 cup, shre
13	01012	CHEESE,CO	98	11.12	4.3	3.38	0	2.67	104	364	1.718	0.778	0.123	17	113	4 oz	210	1 cup, large
14	01013	CHEESE,CO	97	10.69	3.85	4.61	0.2	2.38	90	344	2.311	1.036	0.124	13	113	4 oz	226	1 cup, (not
15	01014	CHEESE,CO	72	10.34	0.29	6.66	0	1.85	137	372	0.169	0.079	0.003	7	145	1 cup, (not packed)	113	4 oz
16	01015	CHEESE,CO	81	10.45	2.27	4.76	0	4	125	308	1.235	0.516	0.083	12	113	4 oz	226	1 cup, (not
17	01016	CHEESE,CO	72	12.39	1.02	2.72	0	2.72	86	406	0.645	0.291	0.031	4	113	4 oz	226	1 cup, (not
18	01017	CHEESE,CRI	342	5.93	34.24	4.07	0	3.21	138	365	19.292	8.62	1.437	110	14	1 tbsp	232	1 cup
19	01018	CHEESE,ED.	357	24.99	27.8	1.43	0	1.43	188	812	17.572	8.125	0.665	89	28	1 oz	198	1 package,
20	01019	CHEESE, FET	264	14.21	21.28	4.09	0	4.09	62	917	14.946	4.623	0.591	89	150	1 cup, crumbled	28.35	1 oz
21	01020	CHEESE,FO	389	25.6	31.14	1.55	0	1.55	64	800	19.196	8.687	1.654	116	132	1 cup, diced	108	1 cup, shre
22	01021	CHEESE,GJE	466	9.65	29.51	42.65	0		1409	600	19.16	7.879	0.938	94	28	1 oz	227	1 package,
23	01022	CHEESE,GO	356	24.94	27.44	2.22	0	2.22	121	819	17.614	7.747	0.657	114	28	1 oz	198	1 package,
24	01023	CHEESE,GR	413	29.81	32.34	0.36	0	0.36	81	714	18.913	10.043	1.733	110	28	1 oz	28	1 slice, (1 c
25	01024	CHEESE,LIN	327	20.05	27.25	0.49	0	0.49	128	800	16.746	8.606	0.495	90	134	1 cup	28.35	1 oz
26	01025	CHEESE,MC	373	24.48	30.28	0.68	0	0.5	81	600	19.066	8.751	0.899	89	132	1 cup, diced	113	1 cup, shre
27	01026	CHEESE,MC	300	22.17	22.35	2.19	0	1.03	76	627	13.152	6.573	0.765	79	112	1 cup, shredded	28.35	1 oz
28	01027	CHEESE,MC	318	21.6	24.64	2.47	0	1.01	75	710	15.561	7.027	0.778	89	28	1 oz	18	1 cubic incl
29	01028	CHEESE,MC	254	24.26	15.92	2.77	0	1.13	84	619	10.114	4.51	0.472	64	28	1 oz		
30	01029	CHEESE,MC	301	24.58	19.72	6.36	0	2.24	131	682	11.473	5.104	0.861	65	132	1 cup, diced	113	1 cup, shre
31	01030	CHEESE, ML	368	23.41	30.04	1.12	0	1.12	134	628	19.113	8.711	0.661	96	132	1 cup, diced	113	1 cup, shre
32	01031	CHEESE,NE	253	9.15	22.78	3.59	0	3.19	152	334	12.79	5.784	0.97	74	28	1 oz	85	1 package,
33	01032	CHEESE,PA	420	28.42	27.84	13.91	0	0.07	180	1804	15.371	7.13	1.386	86	100	1 cup	5	1 tbsp
34	01033	CHEESE,PA	392	35.75	25.83	3.22	0	0.8	92	1376	16.41	7.515	0.569	68	28	1 oz	10.3	1 cubic incl
35	01034	CHEESE,PO	352	23.78	28.2	0.57	0	0.57	136	534	16.691	9.338	0.729	123	132	1 cup, diced	113	1 cup, shre

Figure 31. Subset of USDA food items and nutrients used in ontology

	di	MI				4.4.7.5	6.1.1.1.6	1			1.1						ė		e 1 - 1 - 1		with the fill	nii - 0 - 1 -			0 BC I - E					
NDB_NO	Shrt_Desc	Water_(g) E	nerg_kcal P	rotein_(g)	Lipid_Tot_	Ash_(g)	Carbonydri F	iber_TD_(sugar_tot_c	alcium_(n iro	n_(mg)	Magnesiun P	nosphoru	Potassium_3	odium_(m z	inc_(mg)	copper_m	Manganese	selenium_(vit_C_(mg)	Thiamin_(r	Kibotlavin_	Niacin_(mg P	anto_Acic V	/it_B6_(m) FO	sate_lot H	OIIC_ACID_FOO	d_Holat Ho	Xate_DFE C	noiin
01001	BUTTER,WI	15.87	717	0.85	81.11	2.11	0.06	0	0.06	24	0.02	2	24	24	643	0.09	0	0	1	0	0.005	0.034	0.042	0.11	0.003	3	0	3	3	
01002	BUTTER,WI	15.87	717	0.85	81.11	2.11	0.06	0	0.06	24	0.16	2	23	26	659	0.05	0.016	0.004	1	0	0.005	0.034	0.042	0.11	0.003	3	0	3	3	_
01003	BUTTER OII	0.24	876	0.28	99.48	0	0	0	0	4	0	0	3	5	2	0.01	0.001	0	0	0	0.001	0.005	0.003	0.01	0.001	0	0	0	0	
01004	CHEESE, BL	42.41	353	21.4	28.74	5.11	2.34	0	0.5	528	0.31	23	387	256	1146	2.66	0.04	0.009	14.5	0	0.029	0.382	1.016	1.729	0.166	36	0	36	36	
01005	CHEESE, BR	41.11	371	23.24	29.68	3.18	2.79	0	0.51	674	0.43	24	451	136	560	2.6	0.024	0.012	14.5	0	0.014	0.351	0.118	0.288	0.065	20	0	20	20	
01006	CHEESE, BR	48.42	334	20.75	27.68	2.7	0.45	0	0.45	184	0.5	20	188	152	629	2.38	0.019	0.034	14.5	0	0.07	0.52	0.38	0.69	0.235	65	0	65	65	
01007	CHEESE,CA	51.8	300	19.8	24.26	3.68	0.46	0	0.46	388	0.33	20	347	187	842	2.38	0.021	0.038	14.5	0	0.028	0.488	0.63	1.364	0.227	62	0	62	62	
01008	CHEESE,CA	39.28	376	25.18	29.2	3.28	3.06	0		673	0.64	22	490	93	690	2.94	0.024	0.021	14.5	0	0.031	0.45	0.18	0.19	0.074	18	0	18	18	
01009	CHEESE,CH	37.1	406	24.04	33.82	3.71	1.33	0	0.28	675	0.16	27	473	76	644	3.43	0.056	0.033	28.3	0	0.027	0.434	0.039	0.481	0.049	26	0	26	26	
01010	CHEESE,CH	37.65	387	23.37	30.6	3.6	4.78	0		643	0.21	21	464	95	700	2.79	0.042	0.012	14.5	0	0.046	0.293	0.08	0.413	0.074	18	0	18	18	
01011	CHEESE,CO	38.2	394	23.76	32.11	3.36	2.57	0	0.52	685	0.76	26	457	127	604	3.07	0.042	0.012	14.5	0	0.015	0.375	0.093	0.21	0.079	18	0	18	18	
01012	CHEESE,CO	79.79	98	11.12	4.3	1.41	3.38	0	2.67	83	0.07	8	159	104	364	0.4	0.029	0.002	9.7	0	0.027	0.163	0.099	0.557	0.046	12	0	12	12	
01013	CHEESE,CO	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38	53	0.16	7	113	90	344	0.33	0.04	0.003	7.7	1.4	0.033	0.142	0.15	0.181	0.068	11	0	11	11	
01014	CHEESE,CO	81.01	72	10.34	0.29	1.71	6.66	0	1.85	86	0.15	11	190	137	372	0.47	0.03	0.022	9.4	0	0.023	0.226	0.144	0.446	0.016	9	0	9	9	
01015	CHEESE,CO	81.24	81	10.45	2.27	1.27	4.76	0	4	111	0.13	9	150	125	308	0.51	0.033	0.015	11.9	0	0.02	0.251	0.103	0.524	0.057	8	0	8	8	
01016	CHEESE,CO	82.48	72	12.39	1.02	1.39	2.72	0	2.72	61	0.14	5	134	86	405	0.38	0.028	0.003	9	0	0.021	0.165	0.128	0.215	0.068	12	0	12	12	
01017	CHEESE,CR	54.44	342	5.93	34.24	1.32	4.07	0	3.21	98	0.38	9	106	138	365	0.51	0.019	0.011	2.4	0	0.02	0.125	0.145	0.57	0.035	11	0	11	11	
01018	CHEESE,ED	41.56	357	24.99	27.8	4.22	1.43	0	1.43	731	0.44	30	536	188	812	3.75	0.036	0.011	14.5	0	0.037	0.389	0.082	0.281	0.076	16	0	16	16	
01019	CHEESE, FE'	55.22	264	14.21	21.28	5.2	4.09	0	4.09	493	0.65	19	337	62	917	2.88	0.032	0.028	15	0	0.154	0.844	0.991	0.967	0.424	32	0	32	32	
01020	CHEESE, FO	37.92	389	25.6	31.14	3.79	1.55	0	1.55	550	0.23	14	346	64	800	3.5	0.025	0.014	14.5	0	0.021	0.204	0.15	0.429	0.083	6	0	6	6	
01021	CHEESE,GJI	13.44	466	9.65	29.51	4.75	42.65	0		400	0.52	70	444	1409	600	1.14	0.08	0.04	14.5	0	0.315	1.382	0.813	3.351	0.271	5	0	5	5	
01022	CHEESE,GC	41.46	356	24.94	27.44	3.94	2.22	0	2.22	700	0.24	29	546	121	819	3.9	0.036	0.011	14.5	0	0.03	0.334	0.063	0.34	0.08	21	0	21	21	
01023	CHEESE,GR	33.19	413	29.81	32.34	4.3	0.36	0	0.36	1011	0.17	36	605	81	714	3.9	0.032	0.017	14.5	0	0.06	0.279	0.106	0.562	0.081	10	0	10	10	
01024	CHEESE,LIN	48.42	327	20.05	27.25	3.79	0.49	0	0.49	497	0.13	21	393	128	800	2.1	0.021	0.038	14.5	0	0.08	0.503	0.158	1.177	0.086	58	0	58	58	
01025	CHEESE.MC	41.01	373	24.48	30.28	3.55	0.68	0	0.5	746	0.72	27	444	81	600	3	0.032	0.011	14.5	0	0.015	0.39	0.093	0.21	0.079	18	0	18	18	
01026	CHEESE.MC	50.01	300	22.17	22.35	3.28	2.19	0	1.03	505	0.44	20	354	76	627	2.92	0.011	0.03	17	0	0.03	0.283	0.104	0.141	0.037	7	0	7	7	
01027	CHEESE.MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01	575	0.2	21	412	75	710	2.46	0.022	0.009	16.1	0	0.016	0.27	0.094	0.071	0.062	8	0	8	8	
01028	CHEESE.MO	53.78	254	24.26	15.92	3.27	2.77	0	1.13	782	0.22	23	463	84	619	2.76	0.025	0.01	14.4	0	0.018	0.303	0.105	0.079	0.07	9	0	9	9	
01029	CHEESE.MC	45.54	301	24.58	19.72	3.8	6.36	0	2.24	716	0.23	29	537	131	682	3.61	0.034	0.041	26.8	0	0.059	0.348	0.144	0.415	0.111	27	0	27	27	
01030	CHEESE.MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12	717	0.41	27	468	134	628	2.81	0.031	0.008	14.5	0	0.013	0.32	0.103	0.19	0.056	12	0	12	12	
01031	CHEESE NE	63.11	253	9.15	22.78	1.37	3.59	0	3.19	117	0.13	10	138	152	334	0.82	0.027	0.011	3	0	0.022	0.155	0.21	0.575	0.041	14	0	14	14	
01032	CHEESE PA	22.65	420	28.42	27.84	7.18	13.91	0	0.07	853	0.49	34	627	180	1804	4.2	0.04	0.071	34.4	0	0.026	0.358	0.08	0.45	0.081	6	0	6	6	
01033	CHEESE PA	29.16	392	35.75	25.83	6.04	3.22	0	0.8	1184	0.82	44	694	92	1376	2 75	0.032	0.02	22.5	0	0.039	0 332	0.271	0.453	0.091	7	0	7	7	-
01034	CHEESE PO	45.45	352	23.78	28.2	2.04	0.57	0	0.57	650	0.43	24	360	136	534	26	0.022	0.011	14.5	0	0.014	0.24	0.06	0.21	0.053	18	0	18	18	
01035	CHEESE PR	40.95	351	25.58	26.62	4 71	2.14	0	0.56	756	0.52	24	496	138	876	3 22	0.026	0.01	14.5	0	0.014	0.321	0.156	0.476	0.073	10	0	10	10	
01036	CHEESE BY	71.7	174	11.26	12.98	1.02	3.04	0	0.27	207	0.32	11	158	105	84	1.16	0.020	0.006	14.5	0	0.013	0.195	0.104	0.213	0.043	12	0	12	12	
01037	CHEESE RK	74.41	174	11.20	7.91	1.02	5.14	0	0.31	207	0.44	15	193	105	99	1.10	0.034	0.01	14.5	0	0.013	0.195	0.078	0.242	0.043	12	0	12	12	
01038	CHEESE BO	30.91	287	21.0	26.94	6.72	3.62	n n	0.72	1054	0.77	41	76/1	85	1432	2.54	0.034	0.02	14.5	0	0.021	0.103	0.072	0.428	0.02	7	0	7	7	-
1 10 4 1 3 0	1 Yest Charling Line Parks			31.0		0.74	3.03			41019			/00/	00		4.30	0.03	0.02	1.12.12			0.37								

Figure 32. USDA Abbreviated Excel file with all nutrients

Ç				recipes_data - Excel						
1E INSERT PAGE LA	YOUT FORMUL	AS DATA REVIEW	N VIEW							
t Painter	na • 14 • A* A* 	= = ≫ • = = = € ± Align	₩ Wrap Text Merge & Cente ment	General S - % ,	Delete Format Cells AutoSum → Array Image: Construction of the state of the					
$\therefore \checkmark f_x$	Albondigas Sou	p Recipe								
В	с	D	E	F						
Title	PrepareTime	CookTime	YieldServe	ingredients						
icken Peanut Stew Recipe	20 minutes	l hour, 55 minutes	Serves 6-8	2-3 pounds chicken legs, thighs and/or wings3 Tbsp vegetable oill large yellow or white onion, slicedA 3-inch piece of ginger, peeled and chinced6-S garlic cloves, chopped roughly2-3 pounds sweet potatoes, peeled and cut into chunkal 15-ounce can of crushed tomatoes1 quart chicken stock1 cup peanut butterl cup roasted peanuts1 Tbsp ground coinader 1 teaponc cayenne, or to tasteSalt and black pepper1/4 to 1/2 cup of chopped cilantro	Methodl Heat the vegetable oil in a large soop po brown them in the oil. Don't crowd the pot, so do ti in the oil for 3-4 minutes, stirring often and scrapi sastA© another 1-2 minutes, then add the sweet J tomatose, pearub butter, pearuts, coriander and cays if needed. Cover the pot and simmer gently for 92 bone and the sweet potatoes are tender. 4 Remove to Remove and discard the skin if you want, or chop i back in the pot. 5 Adjust the seasoning for salt and should be experent. Stir in the					
ndigas Soup Recipe	15 minutes	50 minutes	Serves 6-8	2 Tbop olive oill large onion, choppedl large gatic clove, mineed3 quarts of chicken stock or beef stock OR water OR a mixture of both (we usually use half stock half water as the meatballs will create their own stock)*1/2 cup of fomato sauce1/2 lb of string beams, strings and ends removed, cut into 1 inch pieceal large carrots, peelde and sized1/2 cup of raw white rice1 pound ground beef1/4 cup of chopped frash mint leaves1/4 cup of chopped paraley1 raw egg1 1/2 teaspoon salt1/4 teaspoon black pepperA dash of cayenne (optional)1 1/2 cup of frozen or fresh peaal teaspoon of dried oregano, crambled, or 1 Tbop fresh chopped oreganoSalt and pepper1/2 cup chopped fresh cilantro*ff using prepared stock, be sure to use gluten-free.	1 Heat oil in large heavy-bottomed pot (5-qt) over a and cook a minute more. Add broth mixture and bean.2 Prepare the meatballs. Mix rice into meat, into 1-inch meatballs. 3 Add the meatballs to the s towards the end of the 1/2 hour. Add a few pinc taste.G					
1, Shrimp, and Pasta Recipe	10 minutes	20 minutes	Serves 2 as a main or 4 to 6 as a side.	8 ounces long, thin pasta such as spaghetti or fettuciniSalt1/2 cup vegetable oil (grapeseed or canola)3 thinly sliced cloves garlic1 ounce (about 2 medium to large) dried ancho chiles, rinsed, seeded and deveined1/2 pound 21-25 count raw shrimp, peeled, deveined, and tails removed, the shrimp pieces cut into thirdsBlack pepperFreshly grated ParmesanLime or lemon juice, fresh squeezed	1 Put a large pot or saited water on to boil for the cook until al dente. While the water is coming to a Heat oil in a small skillet on medium heat. Once it slotted spoon to a large bowl.3 Thinly slice the anc and slice crosswise). Add the sliced chiles to the hoit bowl with the grafic. Do not over-cook the chiles garlic infused oil. Increase the heat to high, cook: Remove from heat. Add the shrimp and oil to the I the shrimp garlic, chiles, and oil. Sprinkle with					
				8 ounces angel hair pastaSalt 1/4 cup olive oil2 Tbsp finely chopped fresh herbs such	I Bring a large pot of salted water to a boil (1 Tbig the olive oil in a small saucepan on medium heat. Å a parsley to the oil. Å Cook for one minute or until (

Figure 33. Excel file with extracted recipe data

References

- Abburu, S. (2012). A survey on ontology reasoners and comparison. *International Journal of Computer Applications*, *57*(17), 33-39.
- Adlin, T., & Pruitt, J. (2010). *The essential persona lifecycle*. San Francisco, Calif.: Morgan Kaufmann.
- Barreras, A., & Gurk-Turner, C. (2003). Angiotensin II receptor blockers. Proceedings (Baylor University. Medical Center), 16(1), 123. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200815/
- BBC. (2015). BBC Ontologies Food Ontology. Bbc.co.uk. Retrieved 29 August 2015, from http://www.bbc.co.uk/ontologies/fo
- Berners-Lee, T., Hendler, J., & Lasilla, O. (2001). *The Semantic Web. Scientific American*. Retrieved 13 August 2015, from http://www.scientificamerican.com/article/the-semanticweb/
- Brown, D. (2011). Communicating design (2nd ed.). Berkeley, CA: Peachpit Press/New Riders.
- Bushra, R., Aslam, N., & Khan, A. (2011). Food Drug Interactions. *Oman Medical Journal*, 26(2), 77-83. http://dx.doi.org/10.5001/omj.2011.21
- Caine, J. (2012). *Pinning Down the Impact of Pinterest Compete Pulse. Blog.compete.com*. Retrieved 30 July 2015, from https://blog.compete.com/2012/06/28/pinning-down-theimpact-of-pinterest/
- Cantais, J., Dominguez, D., Gigante, V., Laera, L., & Tamma, V. (2005). An example of food

ontology for diabetes control. In *Proc. of the Int. Semantic Web Conf. 2005 workshop on Ontology Patterns for the Semantic Web.*

- Çelik, D. (2015). FoodWiki: Ontology-Driven Mobile Safe Food Consumption System. The Scientific World Journal, 2015, 1-22. http://dx.doi.org/10.1155/2015/475410
- Centers for Disease Control & Prevention (CDC). (2014). *High Blood Pressure (Hypertension) Risk Factors* | *cdc.gov. Cdc.gov.* Retrieved 25 November 2015, from http://www.cdc.gov/bloodpressure/risk_factors.htm
- Datahub.io, (2015). *AGROVOC the Datahub*. Retrieved 29 August 2015, from http://datahub.io/dataset/agrovoc-skos
- DeMiguel, J., Plaza, L., & Diaz-Agudo, B. (2008). ColibriCook: A CBR System for Ontology-Based Recipe Retrieval and Adaptation. In *ECBR Workshops*.
- Dentler, K., Cornet, R., ten Teije, A., & de Keizer, N. (2011). Comparison of reasoners for large ontologies in the OWL 2 EL profile.
- Drugs.com. (2015a). *Amlodipine and Alcohol / Food Interactions Drugs.com*. Retrieved 15 December 2015, from http://www.drugs.com/food-interactions/amlodipine.html
- Drugs.com. (2015b). *Lisinopril and Alcohol / Food Interactions Drugs.com*. Retrieved 15 December 2015, from http://www.drugs.com/food-interactions/lisinopril.html
- Emarketer.com. (2013). *Canadians Take Food Fact-Finding Online eMarketer*. Retrieved 30 July 2015, from http://www.emarketer.com/Article/Canadians-Take-Dood-Fact-Finding-Online/1010428

- Ermilov, I., Auer, S., & Stadler, C. (2013). Csv2rdf: User-Driven csv to rdf mass conversion framework. In *ISEM* (pp. 4-6).
- Eurofir.org. (2015). *All about EuroFIR AISBL*. Retrieved 29 August 2015, from http://www.eurofir.org/?page_id=3
- Gothelf, J. (2012). Using Proto-Personas for Executive Alignment | UX Magazine. Uxmag.com. Retrieved 28 August 2015, from http://uxmag.com/articles/using-proto-personas-forexecutive-alignment
- Graca, J., Pinto, M., Loureiro, V., Monteiro, P., & Anunciacao, O. (2005). Ontology Building Process: the wine domain. In *Proc. of the 5th Conf. of EFITA*.
- Gruber, T. (2009). Ontology (Computer Science) definition in Encyclopedia of Database Systems. Tomgruber.org. Retrieved 30 July 2015, from http://tomgruber.org/writing/ontology-definition-2007.htm
- Hepp, M. (2008). Good Relations: An Ontology for Describing Products and Services offers on the Web. In *Proc. of EKAW'08. LNCS.* (pp. 332-347). Springer.

Hooland, S., & Verborgh, R. Linked data for libraries, archives and museums.

- Kolchin, M., & Zamula, D. (2013). Food Product Ontology: Initial Implementation of a Vocabulary for Describing Food Products. In *14th International Conference of FRUCT Association* (p. 191).
- Li, H., & Ko, W. (2007). Automated Food Ontology Construction Mechanism for Diabetes Care. In *Proc. of the 6th Int. Conf. on Machine Learning and Cybernetics*. Hong Kong.

- Mayoclinic.org. (2015). *High blood pressure (hypertension) Treatments and drugs Mayo Clinic*. Retrieved 30 July 2015, from http://www.mayoclinic.org/diseases-conditions/high-blood-pressure/basics/treatment/con-20019580
- National Center for Biomedical Ontology. (2015). *Bionutrition Ontology Poly Classes* | *NCBO BioPortal. Bioportal.bioontology.org*. Retrieved 29 August 2015, from http://bioportal.bioontology.org/ontologies/BNO/?p=classes&conceptid=http%3A%2F%2F www.owl-ontologies.com%2FOntology1361987617.owl%23Poly
- National Institutes of Health (NIH). (2015a). *Medical Subject Headings Home Page*. *Nlm.nih.gov*. Retrieved 29 August 2015, from https://www.nlm.nih.gov/mesh/
- National Institutes of Health (NIH). (2015b). *RxNorm Overview*. *Nlm.nih.gov*. Retrieved 15 December 2015, from http://www.nlm.nih.gov/research/umls/rxnorm/overview.html
- Noy, N., & McGuinness, D. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Medical Informatics Technical Report, SMI-2001-0880. Retrieved from http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinnessabstract.html
- Parapontis, I. (2012). *Semantic Web LD Stack3*. Retrieved from Web of Data Stories: http://ioannis.parapontis.com/2012/12/07/semantic-web-technology-stacks/semanticweb-ld-stack3/
- Pew Research Center. (2013). Health Fact Sheet. Pew Research Center: Internet, Science & Tech. Retrieved 14 August 2015, from http://www.pewinternet.org/fact-sheets/health-factsheet/

PubMed Health. (2015). High Blood Pressure - National Library of Medicine - PubMed Health. Retrieved 15 August 2015, from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0024199/

- Raebel, M. (2011). Hyperkalemia Associated with Use of Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers. *Cardiovascular Therapeutics*, 30(3), e156e166. http://dx.doi.org/10.1111/j.1755-5922.2010.00258.x
- recipe. (2015). *Merriam-Webster's Collegiate Dictionary*. Retrieved from http://www.merriamwebster.com/dictionary/recipe
- Ribeiro, R., Batista, F., Pardal, J., Mamede, N., & Pinto, H. (2006). Cooking an Ontology. *Artificial Intelligence: Methodology, Systems, And Applications*, 4183, 213-221. http://dx.doi.org/10.1007/11861461 23
- Sam, M., Krishnadi, A., Wang, C., Gallagher, J., & Hitzler, P. (2014). An Ontology Design
 Pattern for Cooking Recipes Classroom Created. In *Proceedings of the 5th Workshop on Ontology and Semantic Web Patterns (WOP2014) co-located with the 13th International Semantic Web Conference (ISWC 2014)* (pp. 49-60). Riva Del Garda: CEUR-WS.org.
- Segaran, T., Evans, C., & Taylor, J. (2009). Programming the Semantic Web. Beijing: O'Reilly.
- Sirin, E., Parsia, B., Grau, B., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical OWL-DL reasoner. Web Semantics: Science, Services and Agents On The World Wide Web, 5(2), 51-53. http://dx.doi.org/10.1016/j.websem.2007.03.004
- Snae, C., & Bruckner, M. (2008). FOODS: A Food-Oriented Ontology-Driven System. In 2nd IEEE International Conference on Digital Ecosystems and Technologies (pp. 168-176).

IEEE.

- Suksom, N., Buranarach, M., Thein, Y., Supnithi, T., & Netisopakul, P. (2010). A Knowledgebased Framework for Development of Personalized Food Recommender System. In Proc. of the 5th Int. Conf. on Knowledge, Information and Creativity Support Systems.
- U. S. Food and Drug Administration. (2015a). Drugs@FDA Frequently Asked Questions. Fda.gov. Retrieved 15 December 2015, from http://www.fda.gov/Drugs/InformationOnDrugs/ucm075234.htm
- U.S. Department of Health and Human Services. (2001). Final MNT regulations. CMS 1169. *Federal Register*, 66(12), 55246.
- U.S. Food and Drug Administration. (2015b). National Drug Code Directory. Fda.gov.
 Retrieved 29 August 2015, from http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm
- USDA. (2011). *NDL/FNIC* Food Composition Database Home Page. Ndb.nal.usda.gov. Retrieved 15 December 2015, from http://ndb.nal.usda.gov
- W3.org. (2004). *OWL Web Ontology Language Reference*. Retrieved 15 December 2015, from http://www.w3.org/TR/owl-ref/
- W3.org. (2015a). *OWL Web Ontology Language Overview*. Retrieved 30 July 2015, from http://www.w3.org/TR/owl-features/
- W3.org. (2015b). *W3C Semantic Web FAQ*. Retrieved 6 October 2015, from http://www.w3.org/RDF/FAQ

- Wikipedia. (2015a). *ICD-10*. Retrieved 29 August 2015, from https://en.wikipedia.org/wiki/ICD-10
- Wikipedia. (2015b). *Web scraping*. Retrieved 17 November 2015, from https://en.wikipedia.org/wiki/Web_scraping
- World Health Organization. (2015). WHO | International Classification of Diseases (ICD). Who.int. Retrieved 29 August 2015, from http://www.who.int/classifications/icd/en/