Implementing a Taxonomy for the Development of a Database for a Type 2 Diabetes Mellitus Preventative Care Personalized Learning System

Dorian Davis, Tanisha Richards and Steven Jiang
Department of Industrial and Systems Engineering
North Carolina A&T State University
Greensboro, NC 27411, USA
davis13@aggies.ncat.edu, tarichal@aggies.ncat.edu and xjiang@ncat.edu

Abstract
Type 2 Diabetes Mellitus (T2DM), the seventh leading cause of death in the United States, is a chronic disease that has plagued the United States adult population and has increased their risk of other debilitating health conditions. In order to reduce the epidemic of T2DM, the United States healthcare system must transform to a more proactive preventative care system. To assist in the transformation, a taxonomy for T2DM preventative care was developed to identify and categorize factors that are significant in the prevention of the disease. Medical databases and subject matter experts were used to classify the variables into seven domains and sub-categories. The taxonomy was implemented in the development of a personalized learning system that requires a compilation of databases that house the T2DM knowledge and the user profile, resulting in users receiving personalized and adaptive information and recommendations specific to their individual differences.

Keywords
Taxonomy, Database, Personalized Learning System, Type 2 Diabetes Mellitus

1. Introduction
Type 2 Diabetes Mellitus (T2DM), the most common form of diabetes is a chronic, but yet, reversible disease that has affected nearly 27 million of the United States adult population (Centers for Disease Control and Prevention, 2014). It is an illness that concerns all ages and brings a heavy economic burden to the person as well as the healthcare industry. If not discovered early and managed properly, T2DM can cause significant morbidity and mortality. The disease has been linked to kidney failure, stroke, high blood pressure, lower limb amputation and other chronic diseases. Reducing these occurrences starts with being aware of the risks. The most common risk factors for developing T2DM are associated with pre-diabetes, age, obesity, family history of diabetes, history of gestational diabetes, impaired glucose metabolism, physical inactivity, and race/ethnicity (Centers for Disease Control and Prevention, 2014). Unfortunately, the minority population (African Americans, Hispanics/Latinos, American Indians, some Asians, and Native Hawaiians/Pacific Islanders) and low socio-economic status communities are at high risk for developing the disease (Centers for Disease Control and Prevention, 2014). Although T2DM in children and adolescents is uncommon, it is being diagnosed more frequently among the United States population and specifically the low socio-economic status community. To negate the disease from progressing, it requires lifelong self-care.

Self-care is defined as taking the necessary action to protect one’s life, health and well-being (Toobert, Hampson, & Glasgow, 2000). Several studies suggest that lifestyle changes such as improved diet and physical activity would prevent or delay the onset of T2DM (Bassuk & Manson, 2005; Diabetes Prevention Program Research Group, 2002; Saaristo et al., 2010; Tuomilehto et al., 2001). Saaristo et al (2010) conducted a study with over ten thousand high risk participants (majority of the participants were obese) to understand how to implement prevention strategies based on knowledge and expertise from clinical trials. Their study focused on specific lifestyle changes such as weight, meal frequency, nutrition, exercise and smoking reduction. In one year, results revealed a 69% risk
reduction in the group who lost more than 5% of their body weight compared to the other groups who maintained or gained weight. Therefore, prevention awareness is imperative in fighting T2DM.

Pew Internet Project research study shows that 72% of U.S. American adults have used the Internet to access health information (Fox & Duggan, 2013). Unfortunately, existing T2DM web-based information systems overwhelm their users with information overload (Davis & Jiang, 2015). Utilizing the appropriate approach such as a personalized learning system will aid in the design of an effective T2DM healthcare information system that will not only reduce information overload, but also allow information to be personalized and adapted to a user’s individual differences.

Personalized learning is an approach borrowed from the educational science discipline where instruction is tailored, based on individual differences or needs (Perry et al., 2008). It is designed to fit the user’s needs, background and learning style. The “traditional” educational system is centered on a universal approach to learning with little to no consideration of individual differences (e.g. age, gender, cultural background). However, personalized learning provides an environment for users who learn under different conditions (e.g. visual, auditory, etc.) and at a different pace to actively participate in the development of their own learning.

Currently, there are computerized instruction such as web-based educational systems (e.g. tutorial systems) that are designed for storing and delivering instructional content and facilitating the interaction between instructors and students. Exactly what and how the instructional content is delivered is very important and is a critical component in developing personalized learning systems. Thus, the aim of this study is to develop a hierarchical taxonomy for a web-based personalized learning system (PLS) for T2DM preventative care and its implementation in development of the system’s database. While other T2DM informational websites have focused their content on managing the chronic disease, this proposed PLS aims to place emphasis on providing knowledge of how to prevent T2DM. Thus, the taxonomy is used to describe some of the most critical factors for T2DM preventative care. It does not however, attempt to establish all instruction of diabetes education. The intent is to advance the knowledge of T2DM, where users can make informed decisions about their health.

2. Method

This paper presents a method that provides guidance in developing and implementing a taxonomy into a database for a personalized learning system designed for T2DM preventative care. This taxonomy will serve as a model for future taxonomies aimed at the prevention of other chronic illnesses/diseases. The goal is to develop a set of domains consisting of a set of subcategories that describe the objects specifically to T2DM preventative care. Overall, a taxonomy should be concise and contain enough dimensions and categories to be of interest. The taxonomy is used to develop the database for the PLS. Figures 1 and 2 provide an illustration of the approach used to develop the taxonomy and the database, respectively. The following sections discuss the approach in detail.

2.1 Knowledge Collection

A taxonomy is the practice and science of classification. It is groupings of practical related data elements that capture and represent concepts. To identify the data elements for the T2DM preventative care, a comprehensive search of medical and educational databases (MEDLINE, Science Citation Index, ERIC, Google Scholar and PsyInfo), diabetes education Internet sources (American Diabetes Association, WebMD, Mayo Clinic and Center for
Disease Control and Protection) were searched and diabetes subject matter experts (SME) were utilized. The subject matter experts that were consulted were a Nursing Professor and a Physician Assistant. Although this study places emphasis on preventative care, it was discovered that the methods to prevent and manage T2DM were identical to some extent. Therefore, the databases were searched specifically for research to prevent the onset of T2DM and how to manage the disease using these specific keywords: type 2 diabetes preventative care, type 2 diabetes prevention, type 2 diabetes management, type 2 diabetes care and chronic illness management. The search resulted in more than 200 articles. After eliminating the articles that were not relevant, fifteen articles were obtained. The elimination criteria used was research on other forms of diabetes and articles that were more than five years old.

The articles collected were analyzed and reviewed with the SMEs for accuracy. The SMEs were also consulted for additional information that may not have been discovered in the literature. For this study, a hierarchical taxonomy was selected to organize the knowledge. Hierarchical refers to the terms or variables relation to each other. The next section discusses the analysis process and the utilization of the taxonomy to organize the knowledge.

2.2 Developing the Taxonomy

Figure 1 presents the approach that structures the taxonomy in terms of domains and subcategories. The four steps are discussed further in the next subsections.

2.2.1 Examine T2DM Knowledge

The knowledge collected from the literature and the SMEs was examined to identify distinguishing characteristics to establish the domains. This process consisted of four steps. Step 3 and 4 are iterative steps.

Step 1: Examine the results, limitations and future research recommendations from existing studies. This step assesses the effectiveness of the study and has implications for providing more effective factors.
Step 2: Identify similarities and dissimilarities for T2DM preventative care.
Step 3: Generate a list of comprehensive terms. The list seeks to capture significant factors that are discussed in the existing research.
Step 4: Consult SMEs for the purpose of refinement. The process for refinement consisted of (1) seeking the advice of SMEs to eliminate redundancy among the initial list of terms; (2) identifying the correct and common terminology used in healthcare; and (3) obtaining additional and current information that may have been missed in the literature.

Upon completion of Steps 1 and 2, a set of terms were generated. This list consisted of over 100 terms associated with T2DM care. The terms were first arranged in four domains (T2DM basics, risk factors, complications and prevention). The SMEs were then consulted for modification. This process was repeated three times. Terms such as ‘exercise’ and ‘nutrition’ were changed to ‘physical fitness’ and ‘food and nutrition’, respectively. The experts suggested that changing ‘exercise’ to ‘physical fitness’ would account for cultural factors pertaining to the meanings and norms around specific behavior (Aronson & Oman, 2004). Some of the terms were also revised to accommodate diverse groups. Thus, terms were selected based on the potential users’ knowledge. For example, ‘hereditary’ was changed to ‘genetics’ and ‘way of life’ was changed to ‘lifestyle’.

The initial domains were modified and expanded to include ‘genetics’, ‘healthcare’, ‘lifestyle’ and ‘pre-diabetes’. The refinements resulted in establishing the following seven domains discussed below: Genetics, Healthcare, Lifestyle, Pre-Diabetes, Symptoms, Complications and Preventative Measures. The domains selected are the point of origin for T2DM preventative care. They represent the risk factors, symptoms and complications related to T2DM. Although some of the terms in the main categories are non-modifiable (i.e. genetics), it is important for one to be aware that non-modifiable variables (e.g. family history of T2DM) can significantly increase their chance of developing T2DM.

Domain 1: Genetics
T2DM has a strong hereditary component. A person is at high risk if his/her mother, father, brother or sister has been diagnosed with the disease. According to the American Diabetes Association website the risk of developing T2DM is:

- 1 in 7, if one parent was diagnosed with T2DM before the age of 50
• 1 in 13, if one parent was diagnosed after the age of 50
• 1 in 2, if both parents have diabetes

Scientists have also linked gene mutations to T2DM. Although not everyone who carries a mutation will develop T2DM, the risk increases if a person has more than one of the following mutation: production of glucose, production of insulin, how glucose levels are sensed in the body and regulation of insulin.

Another way hereditary affects diabetes is family inheritance of having a sedentary lifestyle. If parents are not physically active or do not have healthy eating habits, these behaviors will most likely be passed to the next generation.

In 2011, the mean age for T2DM diagnosis occurred at the age of 54 years, but most recently physicians are seeing a trend of diagnosis amongst children and teenagers (Dabelea et al., 2014). As for race, the American Diabetes Association (American Diabetes Association, 2014b) published the following rates of T2DM by race/ethnicity:
• 7.6% of non-Hispanic whites
• 9.0% of Asian Americans
• 12.8% of Hispanics
• 13.2% of non-Hispanic blacks
• 15.9% of American Indians/Alaskan Natives

Domain 2: Healthcare
Healthcare is an important risk factor associated with T2DM. Regardless of age, sex, race or ethnicity, neglecting personal care can lead to chronic conditions such as T2DM. Healthcare can be defined as the diagnosis, treatment and prevention of diseases and other physical and mental impairments (American Diabetes Association, 2014a; Centers for Disease Control and Prevention, 2014). There is a lot that one can do to contribute to better healthcare. It can be delivered through your own personal care or physician care. Personal care includes attending to personal hygiene, eating a proper diet, getting the proper sleep, being physically active and getting an annual physical or as required by a primary care physician. The role of a primary care physician which also includes physician assistant or nurse practitioner consists of health promotion, disease prevention, health maintenance, counseling, patient education, diagnosis and treatment of acute and chronic illnesses. In addition, primary care physicians encourage active participation in one’s personal healthcare.

It is also important for a person to know their family health history as it is the first genetic test and a risk factor for T2DM. Being aware of family health history increases the individual’s knowledge and is very important information for physicians. This is because it captures hereditary trends and patterns of diseases that can help the physician to make important decisions about a person’s health to reduce risks for diseases or other illnesses.

Domain 3: Lifestyle
Genetics and healthcare play an extremely important role in T2DM diagnosis and prevention, but lifestyle choices are just as important because they can affect the outcome of one’s health. Research has shown that a person’s lifestyle can increase or decrease the diagnosis of T2DM. This includes diet, physical activity, stress management and healthcare. Healthcare in this instance includes oral care, vision care and foot care as infections or other symptoms that can be a sign of T2DM. Smoking or use of tobacco can also increase the risk of complications that are associated with T2DM. In addition, alcohol usage can either increase or decrease the blood glucose level; thus, it is important to drink alcohol or drink mixers in moderation.

Lack of physical activity can lead to obesity, which is also a precursor to other chronic conditions such as high blood pressure and high cholesterol. American Diabetes Association reports that if one loses 7% of their body weight, it will lower the risk for T2DM (American Diabetes Association, 2014a): this includes exercising moderately such as a brisk walk for 30 minutes. A diet filled with low carbohydrates, low fat, reduced sodium and high fiber will lessen the likelihood of T2DM (Evert et al., 2014).

Domain 4: Pre-Diabetes
Individuals diagnosed with pre-diabetes have a blood sugar level higher than normal, but not high enough for a diagnosis of T2DM. The Center for Disease Control (Centers for Disease Control and Prevention, 2011) reports that
79 million U.S. Americans have pre-diabetes, out of the 79 million, 5-10% will progress to diabetes (Tabák, Herder, Rathmann, Brunner, & Kivimäki, 2012). Unfortunately, there are no clear symptoms for pre-diabetes, therefore one’s healthcare and lifestyle are critical factors that will either influence or prevent the diagnosis of pre-diabetes.

Domain 5: Symptoms
Some individuals never experience symptoms or the symptoms are very mild where they can be easily overlooked. Some of the most known symptoms include frequent urination, extreme thirst or hunger. These symptoms can easily be mistaken for other conditions. Thus, patient awareness of symptoms can lead to early detection and treatment.

Domain 6: Complications
As previously mentioned, a person may never experience any symptoms related to T2DM, but can develop complications due to lack of diagnosis. In addition, the severity of T2DM can be ignored in the early stages when an individual is not experiencing any side effects or complications. However, T2DM is a disease that remembers. This means that the body will remember the improper care given to it. Over a period of time, the disease will affect several organs resulting in complications such as heart disease, stroke, nerve damage and lower limb amputation.

Domain 7: Preventative Measures
The main goal of the PLS is to provide personalized knowledge of health education related to T2DM where individuals are well informed and engaged on preventative measures. However, the PLS system can also serve as a tool to educate individuals who may have been diagnosed with pre-diabetes or at an early stage of T2DM, on how to avoid T2DM complications by maintaining a healthy weight, making healthier food selections and avoiding stress.

2.2.2 Examine Domain Characteristics / Identify Misleading Knowledge
After the seven domains were established, the domain characteristics were examined to identify subcategories for each domain. The SMEs were consulted and the initial subcategories were again refined. The investigator and SMEs reviewed the taxonomy for misleading knowledge. However, no misleading knowledge was found. Figure 3 provides an illustration of the final taxonomy for T2DM preventative care.

Figure 3: T2DM Preventative Care Taxonomy
2.3 Database Development

The hierarchical taxonomy presented within this research serves as the foundation for the relational database, a collection of data items organized into one or more related tables with a unique key for each row (Kroenke & Auer, 2010). The objective of the database within this research is to store the user’s information (e.g. demographics, learning preference), search/browser history and the knowledge content for T2DM preventative care. T2DM knowledge however, will not be generically presented. When a selection is made in the personalized learning system, the user will receive personalized information and recommendations that are specific to their needs based on their user profile. As a graphical illustration, the model presented in Figure 4 shows the components of the database system for the personalized learning system.

![Database Model for PLS-T2DM Preventative Care](image)

The database application is the computer program that serves as the communication agent between the user and the data management system (DBMS). The DBMS is the system’s software that creates, retrieves, updates and manages the data. The user profile database houses all relevant information (i.e. age, gender and race) pertaining to the user. The user search history database collects the browsing history of all users. This information is used to provide additional recommendations to the user based on other users search history with similar profiles. The knowledge module contains all the T2DM knowledge relevant to preventative care. The learning preference database stores the learning preference instrument, which determines how the information is personalized and adapted for the user. Personalization and adaptation is determined based on the user’s individual differences (conative, affective and cognitive factors). The databases were created using the four phases below for database design (Kung, Kung, & Gardiner, 2012).

**Phase 1: Requirement analysis – Data requirements**

**Phase 2: Conceptual design – Entity relationship diagram of the collection of requirements and results needed from the database**

**Phase 3: Logical design – Description of the structure of the database (e.g. relational)**

**Phase 4: Physical design – Description of the implementation (e.g. tables)**

**Phase 1: Requirement analysis**

Using the taxonomy, interviews were conducted with subject matter experts and additional literature was researched to determine the description of the data to be generated for the User Profile and Knowledge Module Database, the details of how the data is to be used (Database Application and Users Search History Database) and any additional requirements for the databases (e.g. Learning Preference Instrument).

It was determined that via the user interface, users will create a profile that collects their personal demographics (age, gender, race, educational level, weight, height, etc..) and personal and family health history (e.g. history of obesity, high blood pressure and high cholesterol). This information will be stored in the User Profile Database. In addition, users will be required to take a short assessment to establish their learning preference. The learning preference instrument (LPI) will be housed in the Learning Preference Database. Post response, information is then sent to the User Profile Database for storage.

During this phase, it was understood that additional literature search would be needed to provide the detail information for the Knowledge Module Database, which stores the T2DM preventative care knowledge from the taxonomy. The knowledge must uniquely adapt to the learning preference of each user. Therefore, the Knowledge
Module Database must offer data on a multitude of hypermedia platforms. This will require algorithms to be generated (Database Application). The algorithms will determine the appropriate data and technique to deliver the information from the Knowledge Module Database to the user. Since the personalized learning system was designed as a hybrid recommender system (combines two or more recommender systems approaches), the algorithms are required to be created specifically for knowledge-based and collaborative filtering recommender systems. Recommender systems are used as personalized web-based application and to reduce information overload (Ricci, Rokach, & Shapira, 2011).

Phase 2: Conceptual design
In the development of databases, the next phase was to evaluate the entities and relationships represented in the taxonomy. The hierarchical structure of the taxonomy identified the parent child relationship amongst the information. There are relationships that are directly connected one to one, signifying for each parent entity, there is exactly one child. Other relationships among parent to child are represented as 1:0, 0:0, 0:1, 1:M, M:1 or M:M. In database design, M is defined as many. The 7 domains and subcategories in the taxonomy will serve as parent and child entities, respectively. A partial view of the conceptual design is represented in Figure 5.

Phase 3: Logical Schema
Every relation has a schema, which describes the columns or fields. A schema is a description of the data collected using a given data model. In this research, the logical schema was developed in terms of relational tables based on the requirements from the conceptual design. Using a bottom up approach, normalization was done to eliminate redundancy and dependency among the relational tables (Kung et al., 2012). Normalization is the process of evaluating the relationships of the logical schema design eliminating unnecessary table relations to improve the efficiency of the design (Kroenke & Auer, 2010). Through the normalization process, appropriate candidate keys, primary keys and foreign keys were designated for each attribute.

Phase 4: Physical Design
Post normalization, the final phase in the development process is physical design, which institutes the system conceptualized by the three phases prior. Information has been established within the appropriate tables maintaining normalized relationships. In addition, the algorithms have been written to synthesize the entire system as a whole. It is within this phase the database becomes functional.

3. Discussion and Conclusion
The taxonomy developed in this study has been implemented in a database designed for a personalized learning system for T2DM preventative care. The primary use of the system is to provide personalized and adaptive T2DM preventative care information to users based on their individual differences (conative, affective and cognitive...
factors). The taxonomy and database provide the T2DM preventative care knowledge that will educate and engage users in taking responsibility in making informed decisions about their lifestyle and healthcare to prevent chronic illnesses. Accuracy of the data is determined by current literature and subject matter experts. In addition, the database will have to be updated as new data is validated. The taxonomy and database from this study can promote uniformity in the design of taxonomies across a range of preventative care for other chronic diseases and illustrate how a database facilitates the delivery of personalized and adaptive information through an innovative technology that promotes, engages and educates users on reducing unhealthy behaviors.

Acknowledgements
The authors would like to thank the Center for Compact and Efficient Fluid Power (CCEFP) for partially funding this research.

References

Biography

Dorian Davis is a doctoral candidate in the Department of Industrial and Systems Engineering at North Carolina A&T State University in Greensboro, North Carolina, USA. She earned her B.S. in Industrial Engineering from North Carolina A&T State University and her M.S. from Purdue University in West Lafayette, Indiana, USA. Dorian has worked at Ford Motor Company, 3M Corporation, Chrysler Corporation and Henry Ford Community College. She has led research projects with the National Science Foundation and the Center for Compact and Efficient Fluid Power. Dorian is a member of Alpha Pi Mu Industrial Engineering Honor Society, Tau Beta Pi Engineering Honor Society, Golden Key International Honour Society and Phi Kappa Phi Honor Society. She is also a member of the Institute of Industrial Engineers, Human Factors and Ergonomics Society and the American Society for Engineering Education. Her research interests include human factors engineering, human computer interaction, personalized learning and healthcare systems.

Tanisha A. Richards is a full-time, undergraduate Electrical Engineering student at North Carolina Agricultural & Technical State University, Greensboro, North Carolina. She is a member of the National Society of Black Engineers (NSBE), Golden Key International Honour Society, and Alpha Kappa Mu Honor Society, in addition to other campus organizations. Within her undergraduate matriculation, Tanisha has completed two co-op sessions with Rockwell Collins, Inc. in Cedar Rapids, Iowa where she worked on both the Platform Hardware Team and the Flight Deck Panels Hardware Team. During this time, Tanisha was responsible for the design and development of the Flow Detector, specialized production test equipment designed to determine unit airflow during assembly on the manufacturing floor. She has also worked as an undergraduate research assistant where she designed and implemented a functional system database as well as graphical user interface. Upon graduation, Tanisha plans to continue her engineering education in Industrial & Systems Engineering pursuing a M.S. degree. Her research interest includes application of the user-centered design process to systems of education.

Steven Jiang is currently an associate professor in the Department of Industrial and Systems Engineering at North Carolina A&T State University in Greensboro, North Carolina, USA. He holds a B.S. degree in Mechanical Engineering from East China Institute of Technology, M.S. in Manufacturing Engineering from Nanjing University of Science & Technology and Ph.D. in Industrial Engineering from Clemson University. His research interests are modeling human performance in complex systems, usability engineering and visual analytics.