A CORE REFERENCE HIERARCHICAL PRIMITIVE ONTOLOGY FOR ELECTRONIC MEDICAL RECORDS SEMANTICS INTEROPERABILITY

by

Ziniya Zahedi
B.S. December 2012, Old Dominion University
M.E.M. May 2015, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ENGINEERING MANAGEMENT AND SYSTEMS ENGINEERING

OLD DOMINION UNIVERSITY
August 2020

Approved by:

T. Steven Cotter (Director)

Charles B. Daniels (Member)

C. Ariel Pinto (Member)

Mustafa Canan (Member)

ABSTRACT

A CORE REFERENCE HIERARCHICAL PRIMITIVE ONTOLOGY FOR ELECTRONIC MEDICAL RECORDS SEMANTICS INTEROPERABILITY

Ziniya Zahedi
Old Dominion University, 2020
Director: Dr. T. Steven Cotter

Currently, electronic medical records (EMR) cannot be exchanged among hospitals, clinics, laboratories, pharmacies, and insurance providers or made available to patients outside of local networks. Hospital, laboratory, pharmacy, and insurance provider legacy databases can share medical data within a respective network and limited data with patients. The lack of interoperability has its roots in the historical development of electronic medical records. Two issues contribute to interoperability failure. The first is that legacy medical record databases and expert systems were designed with semantics that support only internal information exchange. The second is ontological commitment to the semantics of a particular knowledge representation language formalism. This research seeks to address these interoperability failures through demonstration of the capability of a core reference, hierarchical primitive ontological architecture with concept primitive attributes definitions to integrate and resolve non-interoperable semantics among and extend coverage across existing clinical, drug, and hospital ontologies and terminologies.

Let us be practical and analytical with this dedication, as I am an Analyst. 60% of this thesis is dedicated to my Mom and Dad, who have shaped me for life, who always had faith in me and taught me that the harder we work, the luckier we get. 30% of this thesis is dedicated to my better half, Dr. Mahmud as without him this journey would not even have started. And 10% is dedicated to people who have always doubted and criticized me, as without their criticism I would not have this drive to be successful. That is a full 100%.

ACKNOWLEDGMENTS

Earning a PhD is not for the faint of heart. One must be persistent and committed throughout the full journey. But most importantly, one needs to have a support system to get going. This endeavor would not have been possible without some individuals. First and foremost, my PhD advisor Dr. T. Steven Cotter, who held my hands throughout this full process and mentored me each step of the way. His untiring efforts to keep me motivated deserve special recognition. My committee members, Dr. Cesar A. Pinto, Dr. Charles Daniels, and Dr. Mustafa Canan, who have supported me tremendously. Their patience and hours of guidance made this cumbersome task painless.

My parents, who have been by my side since forever. They have always motivated me and prepared me for any challenges that life throws at me. Their sacrifices made me what I am today and without them I would not be here.

My dear husband, my better half, the love of my life, Dr. Mahmud; without his consistent push (sometimes to the extent that made me furious but eventually brought out the best in me) I would not have probably even started this journey in the first place. I have always doubted my capabilities, but he always knew what I could achieve and here I am today with a smile on my face.

There are times when I have thought that I am a very unlucky person. But I am actually very fortunate and blessed when it comes to my surroundings. My parents, my husband, my extended family who would do anything for me, the greatest professor and mentor, and very supportive committee members; all these awesome people made me what I am today and I cannot go a day without acknowledging and thanking them.

NOMENCLATURE

HI	Human Intelligence, (No Units)
MI	Machine Intelligence, (No Units)
AI	Artificial Intelligence, (No Units)
DL	Descriptive Logic, (No Units)
A	Abductive Meaning, (No Units)
D	Deductive Structure, (No Units)
HCI	Human-Computer Interaction, (No Units)
HMI	Human-Machine Interaction, (No Units)

TABLE OF CONTENTS

Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Theoretical Formulation

Currently, electronic medical records (EMR) cannot be exchanged among hospitals, clinics, laboratories, pharmacies, and insurance providers or made available to patients outside of local networks. Hospital, laboratory, pharmacy, and insurance provider legacy databases can share medical data within a respective network and limited data with patients. The lack of interoperability has its roots in the historical development of electronic medical records.

Two issues contribute to interoperability failure. The first is that legacy medical record databases and expert systems were designed with semantics that support only internal information exchange. The second is ontological commitment to the semantics of a particular knowledge representation language formalism. Uschold and Gruninger (1996) observe that ontological design for interoperability involves a tradeoff: "… making too many ontological commitments can limit extensibility, making too few can result in the ontology being consistent with incorrect or unintended worlds (i.e., models)." The universality of knowledge representation semantics was not considered in legacy medical record databases and expert systems, which severely limits extensibility needed for interoperability.

Hierarchical primitive ontologies present the potential to resolve complex conceptual semantic spaces like those in electronic patient medical records. Recognizing the implications of primitive ontology theory for ontology engineering, Rector (2003) proposed normalization and modularization of proper ontologies (Welty and Guarino, 2001) to yield hierarchical primitive ontologies. Stuckenschmidt, Parent, and Spaccapietra (2009) provided a survey of ontology

partitioning and modularization approaches to identify and connect primitive classes. However, no work has investigated building hierarchical primitive ontologies to integrate semantics of existing biomedical ontologies and terminologies.

1.2 Purpose

This research seeks to demonstrate the capability of a core reference, hierarchical primitive ontological architecture with integrated primitive concept ontology extraction and concept attributes decomposition to integrate and resolve non-interoperable semantics among and extend coverage across existing clinical, drug, and hospital ontologies and terminologies. A primitive concept is defined as follows:

Definition: *Every primitive concept is its own semantic hypernym and must be uniquely specified by its set of "is-a" existential primitive attributes.*

This research contributes to the interoperability and transferability of electronic patient medical records and, thus, contributes to societal quality of health. The proposed project investigated the potential for increased patient electronic medical records semantics interoperability coverage through development of a patient medical records core reference, primitive ontology hierarchy.

1.3 Problem

The capability for accurate transmission of patient medical information and records within and among hospitals, clinics, laboratories, pharmacies, and insurance providers does not currently exist due to lack of interoperable medical terminology semantics.

CHAPTER 2

BACKGROUND OF THE STUDY

2.1 Socio-Technical Medical Records Literature Review

Medical Records 1920s to 1960s

Prior to the 1920s, medical records existed only in the form of narratives documenting symptom and outcome observations and documentation of prior successful cures. With scientific advancements of the 20th century, physicians realized that to improve the diagnosis and treatment of illnesses they needed to have a standard way of documenting and communicating medical information with other physicians. To accomplish standardization, the American College of Surgeons (ACOS) established the American Association of Record Librarians (AARL) in 1928 to "… elevate the standards of clinical records in hospitals and other medical institutions" (AHIMA, 2018). The Association has authorized three name changes: (1) in 1938 it became the American Association of Medical Record Librarians (AAMRL) and focused its work on the creation of standards and regulations for medical records; (2) in 1970 its name changed to the American Medical Record Association (AMRA), and the organization extended its standardization activities to include community health centers and other health service providers; and (3) in 1991 it became the American Health Information Management Association (AHIMA), with the new name reflecting the transition to data-driven decision making in healthcare. In the 1960s, the AAMRL drove standardization of paper-based medical records, and standardization of electronic records has continued through the AMRA and AHIMA.

Medical Records 1960s and 1970s

The primary driver toward electronic medical records was the passage into law of Medicare and Medicaid in 1965. The law required hospitals to collect and document healthcare services provided for reimbursement. Although computers were being increasingly used for billing and accounting, paper-based records remained the primary documentation mechanism. As computers became affordable, hospital department specific databases were coded to support patient registration and billing and laboratory and pharmacy records. Initial EMRs were developed by and used within academic medical facilities, but none of the electronic systems translated all the information in paper-based medical records into electronic form.

Medical Records 1980s

Diagnosis-related Groups (DRG) were introduced in the early 1980s to determine Medicare payment schedules for medical service "products" within case groups. The state of New Jersey experimented with implementing DRGs in its hospital systems for three years. Full integration was never achieved. In parallel to development of DRGs, the Master Patient Index (MPI) was introduced by Wiedemann (2010) to be used across healthcare departments for sharing patient information. The MPI is an indexed database of patients within a healthcare provider linked together by a medical record number identifier. Even with the advancement of DRGs and the MPI, by the end of the 1980s hospital departments still could not share patient information with each other let alone external clinics, pharmacies, insurance providers, or patients.

Medical Records 1990s

By the early 1990s, most EMRs were still a hybrid of paper and electronic data deployed on a combination of mainframe and personal computers (Evans, 2016). The complexity and

inadequacies of the mixed paper-electronic medical records was the driver behind the Institute of Medicine's call to shift to a complete electronic medical record system (Institute of Medicine, 1997). However, other medical professionals noted that the initial cost of a completely computerized EMR system was prohibitive and advocated that only key data be computerized as a complement to the paper-based system (Regan, 1991).

Advances in computing technology and the Internet made online access to health information possible. At the same time, competition in healthcare and the health insurance industries drove consolidation of hospitals into health systems competing on delivery of integrated health care (Ginsburg, 2005). Efforts were initiated in the medical profession to transition from paper-based to electronic medical records. Networks of EMR workstations were linked to create and process inpatient orders, but creation of electronic orders required more physician time than the traditional paper charts, broke down physician-nurse communication based around the paper-based system, and actually induced errors putting patient health and life at risk (Wachter, 2017). Similarly, initial implementation of nurse workstations failed due to excessive manual data entry time. Data entry errors and poor-quality data limited the usefulness of early EMRs and put patients at risk (Tierney, et. al., 1993). Despite the noted implementation and interoperability problems, the massive amounts of health care data also proved valuable for epidemiological studies (Hierholzer, 1992). Recognizing the potential informational value, the medical community pressed forward with EMR implementation.

Medical Records 2000 to Present

By the late 1990s, EMR implementation had not overcome the interoperability barriers. On the other hand, the merger of individual hospitals into health care systems drove the need for

information interoperability. Integrated EMRs provided the potential for improved decision making and reduction of the incidence of errors.

In 2004, President Bush established the Office of the National Coordinator for Health Information Technology (ONCHIT) with the goal of implementing electronic health records (EHRs), nationwide within ten years. While there was bipartisan support for healthcare EMRs, the US Congress allocated no funding for ONCHIT. President Bush reallocated $42 million from within the Department of Health and Human Services budget to fund ONCHIT (Wachter, 2017). Under its first director, the ONCHIT set forth its primary goal of planning and designing the implementation of a National Health Information Network (NHIN) to promote electronic health information exchange among HIEs. Realizing that a NHIN could not be achieved without healthcare information standardization, the ONCHIT made grants to the American National Standards Institute (ANSI) to coordinate the creation of Health Information Technology Standards and to create the Health Information Security and Privacy Collaborative. The ONCHIT also awarded a grant to a collaboration among the American Health Information Management Association (AHIMA), the Healthcare Information and Management Systems Society (HIMSS) and the National Alliance for Health Information Technology (NAHIT) to create and administer the Certification Commission for Health Information Technology (CCHIT). Since 2006, CCHIT has been the sole certifying agency for EMR software applications (Gur-Arie, 2013).

By the time Barack Obama entered office in 2009, progress toward EMR implementation was not realized. NAHIT had voluntarily dissolved itself. President Obama re-initiated implementation of electronic medical records as a part of the American Recovery and Reinvestment Plan (ARRP) with a goal of access of all citizens to their electronic medical

records by 2014 (Manos, 2014). The Health Information Technology for Economic and Clinical Health Act (HITECH Act) was part of the ARRP. The HITECH Act objective was to motivate the implementation of EMRs and to support EMR technology improvement by providing monetary incentives for demonstration of use of EMRs. The monetary incentives were offered from 2011 to 2015 after which time penalties were imposed for failing to demonstrate EMR use. EMR adoption grew as a result of the renewed support. By 2015, 96% of hospitals and 87% of physician practices were using EHRs. The renewed emphasis did not overcome the original implementation and interoperability problems and induced other problems (Evans, 2016). Adler-Milstein (2017) notes that the major technical issue still to be overcome is interoperability; specifically, "Why can't (EMR) systems talk to each other? The substantial increase in electronic health record adoption across the nation has not led to health data that can easily follow a patient across care settings." Adler-Milstein's research suggests that the reason for interoperability failure is technological and multidisciplinary. Technological challenges include standardization of medical terminology semantics, software applications, and healthcare provider procedures. Multidisciplinary challenges center on balancing national policy versus private EMR software vendors' profitability. "Though billions in monetary incentives fueled EHR adoption itself, they only weakly targeted interoperability."

2.2 Patient Medical Records Interoperability Literature Review

The recent acceleration in the deployment of electronic health record (EHR) systems has precipitated the emergence of a few dominant terminologies widely adopted in the clinical community. Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and the Logical Observation Identifiers, Names, and Codes (LOINC®) are the two that have become

international standards. The RxNorm,terminology is used in the United States, but similar national drug terminologies exist in other countries (e.g., the NHS Dictionary of medicines and devices (dm+d) (2018) in the U.K., the Australian Medicines Terminology (AMT) (2018) in Australia). SNOMED CT, LOINC, and RxNorm have been used and referenced in many articles over time, but none of the articles discussed how they could contribute in building an interoperable system. This work will discuss the history and structure of these terminologies briefly before moving to a detailed investigation.

2.2.1 Brief History of SNOMED CT

The Structured Nomenclature of Pathology (SNOP) was initiated in 1965. As illustrated in Figure 1, versions of SNOMED have been developed both in terms of content structure and representation.

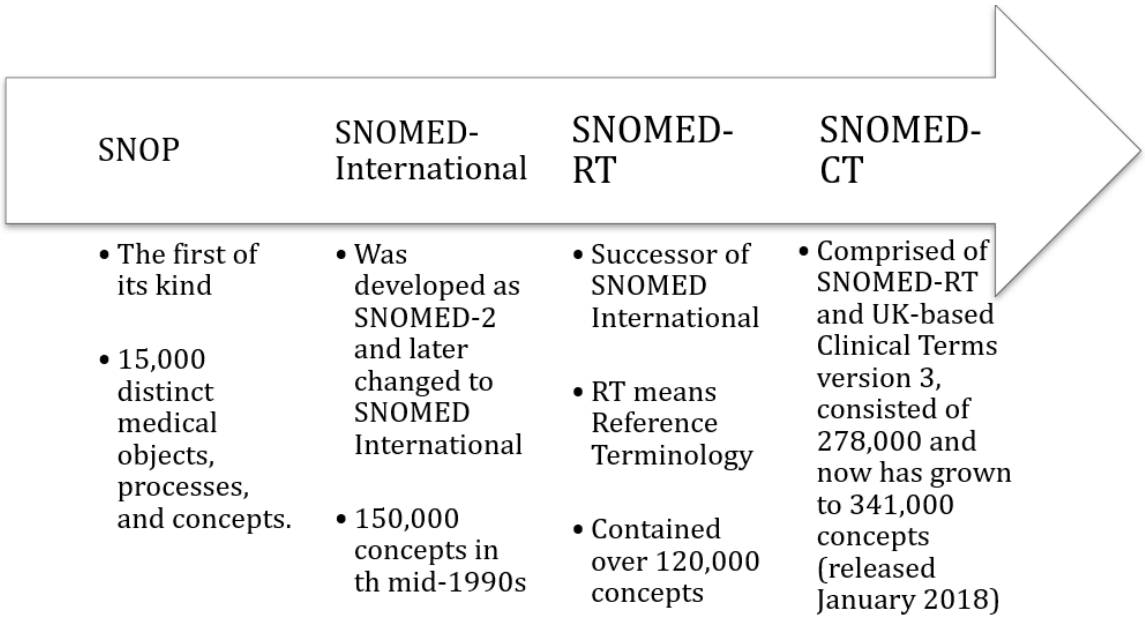


Figure 1: Development of SNOMED CT.

(Source: Dunham, 1978; Spackman, 1997; and Wang et al., 2001)

SNOP and SNOMED-International versions used multi-axial systems, but SNOMED-RT abandoned the self-standing axes and started using description logic. SNOMED CT continued to use the same logic as its underlying representation. SNOMED CT was first released in January 2003, and since then the updated versions have been released twice a year. The January 2018 release contains 341,000 active concepts, 1,062,000 active relationships and 1,156,000 active descriptions. The largest categories of concepts in SNOMED CT are disorders (22%), procedures (17%), body structures (11%), clinical findings other than disorders (10%), and organisms (10%) (Bodenreider et al., 2018).

SNOMED has always been kept simple enough so that it can be used widely by clinicians. The relationships between concepts and allowed values are determined and specified by the concept model. SNOMED CT is now being used by over 32 countries (as of May 2018) with a population over 2 billion.

2.2.2 Brief History of RxNorm

At the beginning of the 21st century, there was no standardized drug terminology (Sperzel et al., 1998). While many companies provided clinical information, their codes for drugs were all different. For example, the same transdermal patch delivering 0.583 milligrams of nicotine per hour for 24 hours (e.g., to help with smoking cessation) is referred to in three of the major drug knowledge bases with the varying codes and names listed in Table 1 (Bodenreider et al., 2018).

Table 1: Differences in Drug Names and Codes.

Codes	Drug Names
2707	nicotine 14 mg/24 hr transdermal film, extended release
102712	Nicotine 14 MG/24 HR Transdermal Patch, Extended Release
016426	NICOTINE 14 mg/24 hour TRANSDERM PATCH, TRANSDERMAL 24 HOURS

Differences in capitalization and abbreviation are problematic when the system is trying to communicate. The lack of drug code standardization generated the need to create RxNorm. RxNorm makes the drug terminologies interoperable. RxNorm was introduced in 2002 through the Unified Medical Language System (UMLS), a terminology integration system, and was established as independent terminology in 2004 (Bodenreider, 2004). RxNorm files are publicly available and downloaded about 1,000 times each month. RxNav (the browser that allows users to explore RxNorm from a variety of names and codes including proprietary names and codes (RxNav, 2018)) has over 2,000 unique users and serves some 500,000 queries annually. The RxNorm API has over 20,000 unique users and serves some 800 million queries annually. The main use cases of RxNorm are electronic prescribing, information exchange, formulary development, reference value sets, and Analytics.

2.2.3 Brief History of LOINC

The Regenstrief Institute, a non-profit medical research organization associated with Indiana University initiated Logical Observation Identifiers, Names, and Codes (LOINC®) in 1994. LOINC is clinical terminology for identifying health measurements, observations, and documents. LOINC was first released in May 1995 when it contained only terms for laboratory

testing. By December 1996, it had already added about 1500 clinical terms including vital signs for measurements, ECG measures, etc. LOINC had 60 releases in the last 20 years, and it has grown in other domains as well such as radiology (Vreeman et al., 2005), standardized survey instruments and patient-reported outcomes measures (Vreeman et al., 2010), clinical documents, nursing management data (Frazier el al., 2001), and nursing assessments (Dentler et al., 2011).

A semantic data model that contains six majors and up to four minors is used by LOINC to create specified concepts. The attributes are:

1. Component (e.g., what is measured, evaluated, or observed),
2. Kind of property (e.g., mass, substance, catalytic activity),
3. Time aspect (e.g., 24-hour collection),
4. System type (e.g., context or specimen type within which the observation was made),
5. Type of scale (e.g., ordinal, nominal, narrative),
6. Type of method (e.g., procedure used to make the measurement or observation) (Bodenreider et al., 2018).

LOINC has been adopted widely in the United States and internationally. There are more than 60,000 registered users from 170 countries, and it has been translated into 18 variants of 12 languages (Vreeman et al., 2012). More than 30 countries have adopted LOINC as a national standard.

2.2.4 Brief History of SNOMED CT, RxNorm, and LOINC Integration

After EMRs were introduced in 1994, different electronic systems communicated with each other by sending clinical information using the messaging systems called Health Level Seven (HL7) or ASTM 1238 (American Society for Testing and Materials). This created decoding problems as the terminologies were not granular enough and were focused more on

coding for billing. HL7 did not deliver the expected clinical results, so the need for a standardized terminology consisting of interoperable parameters emerged. To improve interoperability, the collaboration between the developers of SNOMED CT, RxNorm, and LOINC has increased over the past few years. SNOMED CT is being leveraged as the building blocks of LOINC for a more consistent clinical and laboratory observation. The new international drug model in SNOMED- CT facilitates the development of compatible drug models in RxNorm for better consistency. Even though this collaboration has focused on improving interoperability, cross-coverage among these terminologies is still low.

Research into Medical Terminologies Integration

The U.S. National Library of Medicine, Lister Hill National Center for Biomedical Communications has led the research effort for the integration, dissemination, and quality assurance of drug ontologies and biomedical ontologies. According to Oliver Bodenreider (2018), Senior Scientist and Chief of the Cognitive Science Branch, "Despite the best efforts of human editors and the use of formalisms, such as description logics, content errors remain frequent in biomedical terminologies, which justifies the development of multiple approaches to identifying these problems" (p. 4).

There have been many quality assurance (QA) effort studies, but these studies merely focused on the main problem area where errors occur more frequently within the subsets of terminologies (Ochs, et al., 2015; Ochs, et al., 2013; Zhang, et al., 2017). Even though these efforts were somewhat accommodating to human reviewers, they are still not truly effective. As part of the "Medical Ontology Research" project, Bodenreider's team has explored quality assurance and interoperability issues in a variety of biomedical terminologies including drug terminologies (RxNorm), standard clinical terminologies (SNOMED CT, LOINC), and

specialized terminologies, such as HPO – the Human Phenotype Ontology and the Orphanet terminology for rare diseases. They have reviewed 32 investigations that were performed as part of the project. Half of the investigations revealed quality assurance issues for which they developed some auditing and evaluation methods, and half were interoperability issues related. Structural, semantic, lexical, and transformation methods were applied to audit terminology quality. Structural methods use the taxonomic structure of concept lattices. Semantic methods use description logic-based concept definitions. Lexical methods were based on term properties. Other evaluation methods included transforming the representation of a terminology to a different formalism (semantic architecture, frames, rules, and ontologies) and evaluating for compliance to that formalism, evaluating terminologies to specified principles, and mapping to other ontologies.

Bodenreider's application of structural-lexical methods to SNOMED CT extracted 6,801 non-lattice subgraphs that matched four primary lexical patterns. A random sample of 59 small subgraphs out of 2,046 amenable to visual inspection showed that all 59 contained errors as confirmed by terminology experts. The most frequent error was missing "is-a" relationships. An investigation of partial mappings between the Human Phenotype Ontology (HPO) and SNOMED CT revealed that there were 7,358 HPO concepts that did not completely map to SNOMED CT. A reference list of mappings between the Disease Ontology (DO) and SNOMED CT showed that 2,453 of the 6,931 DO concepts had no mapping to SNOMED CT (Bodenreider, 2018).

In summary, the quality assurance processes developed by the U.S. National Library of Medicine have proven effective in identifying a limited number of errors with precision. However, the quality assurance processes still rely heavily on human evaluation and are thus

slow and reactive relative to medical terminology development. Most important, current quality assurance processes are not able to identify the root cause of interoperability errors.

2.3 Interoperability Limitations of Existing Medical Ontologies and Terminologies

Semantic interoperability deficiencies (inconsistent semantics, missing semantics, missing links, and incomplete coverage) in patient medical record terminologies and ontologies can be traced to differences in biomedical terminology standards, limited terminology coverage, static mappings among biomedical terminologies, and missing hierarchical relationships across biomedical terminologies.

Barbarito (2012) points out that the everyday workflow in information technologies (ITs) have a certain degree of independence. This independence may be the cause of difficulty in interoperability between information systems standards. Thus, interoperability failures result from a lack of standard coding system in data dictionary (Lau and Shakib, 2005). Most of the time, the electronic data collected do not follow any standard code or structure, which causes communication problems between healthcare providers. Data standardization means that the same set of codes needs to be used throughout a system. For example, in the domain of "sex", it could be decided to code the sex of male as "1", female as "2", and unknown as "3". This domain will always consist of three members, "male", "female" and "unknown", and will be coded by following this standard, thus forming a vocabulary for data standardization. If all data about sex is coded consistently according to this vocabulary, the data should always be understandable and usable for analysis. Standard vocabularies will be the pathway to create interoperability between systems. Both Barbarito (2012) and Lau and Shakib (2005) offer data standardization as a solution. The Lombardy case mentioned by Barbarito (2012) shows the whole process and how

this region in Italy became interoperable by just following a twofold approach. First, the political and operative push towards the adoption of the Health Level 7 (HL7) standard within each hospital failed to interlink databases among hospitals. Second, providing a technological infrastructure for data sharing based on regionally recognized interoperability specifications failed to provide interoperability across regions. Data standardization means terminologies communicate with each other seamlessly without failing to understand each other's codes.

Bodenreider (2010) studied 13 different terminologies and ontologies over a 12 year period for terminology coverage. Some of the notable studies include:

- Unified Medical Language System (UMLS): Bodenreider found thousands of inconsistent concepts throughout the system even though those were not indicative of any errors. A pattern of false synonymy was found which could create "real" errors.
- RxNorm: This is a vast terminology that relies on human editors. Multiple inconsistencies and missing links were identified, and 62% of the inconsistencies were fixed as of January 2009.
- SNOMED: A limited number of coverage errors were detected which defeated the Quality Assurance Mechanisms that were in place. Some of the errors were fixed.

Bodenreider established that the terminologies themselves are inconsistent because of the lack of standardization and coverage. Until the terminology coverages are fixed from within, the interoperability issues will continue to exist.

Cholan and Bodenreider (2018) sought to identify the gaps and similarities between clinical research value sets and healthcare quality value sets. They have gathered the lists of value sets from Clinical Data Interchange Standards Consortium (CDISC) which was developed for clinical data exchange used by the Food and Drug Administration and from Value Set

Authority Center (VSAC), which maintains value sets for clinical quality measures. VSAC uses codes and terms from standard terminologies like SNOMED CT, RxNorm, and LOINC. After mapping and evaluating the interoperability between VSAC and CDISC, the authors found limited interoperability between the two. There is a different number of value sets in CDISC, and each value set has limited to no coverage by SNOMED CT or LOINC. Biomedical terminologies are dynamic with changes in term definitions, dropping terms, adding terms, and local extensions requiring constant monitoring and revisions to maintain the static mappings up to date (Lau and Shakib, 2005). Without constant monitoring static patient data may become non-interpretable. For example: standard vocabularies may retire or delete certain codes. If patient data is stored using the retired or deleted code it will no longer be interoperable with other systems. Thus, constant updating and monitoring are required to maintain interoperability of static data sets.

Bodenreider (2016) conducted a study to identify missing hierarchical relationships from logical definitions of concept names in SNOMED CT. The study inferred hierarchical *subClassOf* relationships among the concepts using the ELK reasoner and compared the derived hierarchy to the original SNOMED CT hierarchy. From manual comparison of the hierarchies, the study identified 559 (3.5%) potentially missing out of a total of 15,833 hierarchical relationships. Of the 559 potentially missing hierarchical relationships, 436 (2.8%) were found to be valid. Cui, et al. (2017), introduced a hybrid structural-lexical method for systematically identifying missing hierarchical relationships in SNOMED CT. They extracted all non-lexical subgraphs using the scalable MapReduce algorithm. Four lexical patterns associated with a specific error type indicating missing hierarchical relationships were identified. They found 6,801 non-lattice subgraphs matching these lexical error patterns out of which 2,046 were admissible to manual inspection. A random sample of 100 patterns was taken. Of the sample,

59 were reviewed in detail by domain experts, and all 59 contained hierarchical errors. The domain experts identified missing "is-a" errors due to incomplete or inconsistent modeling of the concept to be the most frequent.

In summary, this literature review identified the following issues with EMR interoperability.

- The transition from paper-based to electronic medical records did not identify interoperability issues and increased the risk of diagnosis and treatment errors due to the breakdown of physician-nurse communication. Specifically, there are human consequences and impacts arising from medical terminology interoperability failures.
- Despite national investments toward implementing electronic health records over the last thirty years, significant interoperability issues remain.
- Semantic interoperability deficiencies in patient medical record terminologies and ontologies can be traced to differences in terminology standards, limited terminology coverage, static mappings among terminologies, false synonymy, and missing hierarchical relationships across biomedical terminologies.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Overall Research Design: The Hierarchical Ontology Architecture

The word "ontology" originated within Philosophy to mean a systematic explanation of "being." Within knowledge and ontology engineering, ontology means a set of concept categories, their attributes, and axiomatic relationships within and between them that specifies a knowledge area or domain. This work defines ontology as a set of logical concepts and axioms that specify their interrelationships designed to account for a discipline's body of knowledge.

Roussey, et. al., (2011) argue that a four-level hierarchy of ontologies based on language expressivity and formality, Figure 2, is necessary to fully specify a knowledge discipline.

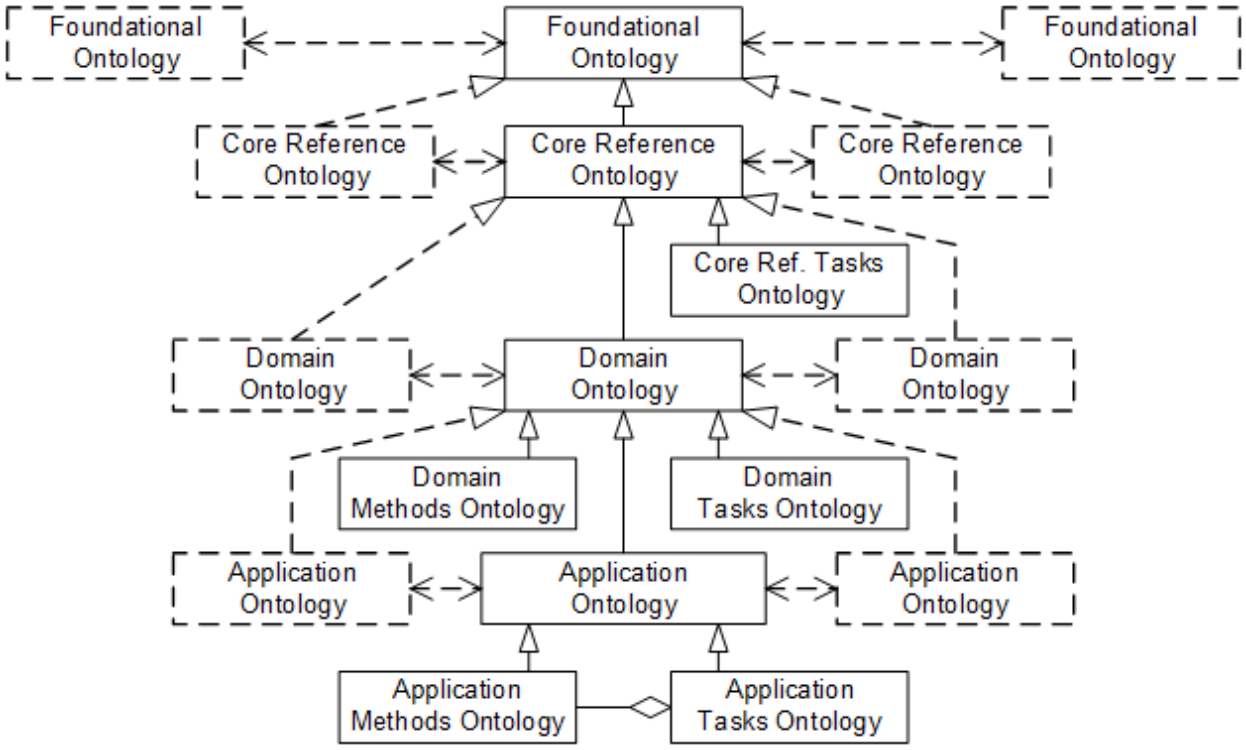


Figure 2: Ontology Hierarchy (Rousey, et al.).

- A *top-level foundational ontology* specifies a broad taxonomic and axiomatic structural scope of low granularity for a general body of knowledge. It specifies the taxonomic and axiomatic basis for underlying core reference ontologies and domain ontologies. Foundational ontologies are designed and implemented using a top-down approach and general methodologies such as BFO, Cyc, DOLCE, GFO, PROTON, and SUMO (Mascardi & Paolo, 2007).
- A *core reference ontology* provides the taxonomic and axiomatic scope structure of finer granularity for a sub-discipline within a body of knowledge by integrating differing domain viewpoints. Core reference ontologies are designed and implemented using a top-down approach with reference to the discipline's foundational ontology using a general methodology such as SENSUS (Jones, Bench-Capon, & Visser, 1998).
- A *domain ontology* provides the specific taxonomic and axiomatic structure necessary to organize knowledge about a discipline. Domain ontologies are designed and constructed using a middle-out approach with reference to the relevant core reference ontology using a general methodology such as SENSUS.
- An a*pplication* or *local* ontology provides the specific taxonomic and axiomatic structure necessary to organize specific competency knowledge within a discipline's domain. Application ontologies are designed and constructed using a bottom-up approach with reference to the relevant domain ontology using a specific methodology such as CommonKADS, DILIGENT, Enterprise Model Approach, KACTUS, KBSI IDEF5, METHONTOLOGY, or TOVE (Corcho, Fernandez-Lopez, & Gomez-Perez, 2003) (Cristani & Cuel, 2005).

- A *task* ontology provides the taxonomic and axiomatic structure necessary to specify the design of the components, methods, diagnosis, and satisfaction criteria to solve a particular problem. A *task* ontology selects appropriate *methods* via the methods ontology for a particular problem (Chandrasekaran and Josephson, 1997).
- A *methods* ontology provides the taxonomic and axiomatic structure necessary to specify a collection of analyses and sub-analyses, control information for passing information among and invoking analyses and sub-analyses, and control information for problem solution (Chandrasekaran and Josephson, 1997).

Obrst (2010) argues that for engineering purposes, an ontological architecture may need to be layered within levels in order to represent consistent and coherent theories.

> … upper ontologies are most abstract making assertions about constructs … that apply all lower levels …. Mid-level ontologies are less abstract and make assertions that span multiple domain ontologies. (p. 29)

Assuming only primitive ontologies, Obrst's layered hierarchical architecture is represented in Figure 3. In Figure 3, a line direct link, primitive propagation indicates that a lower-level ontology is a proper subcategory of a higher-level ontology category, and an open arrow, primitive-modular link indicates that a lower-level ontology references a higher-level ontology.

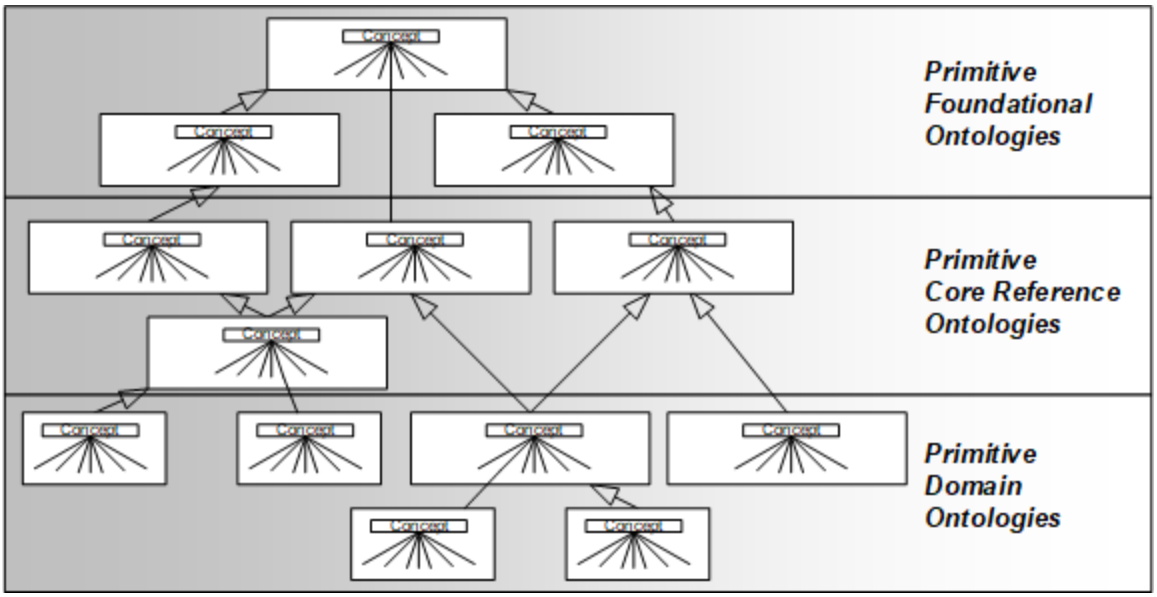


Figure 3: Representation of Obrst's Layered Hierarchical Primitive Ontology Architecture.

3.2 Sample Collection - Establishing the Corpus

This research used SNOMED CT glossary textual definitions downloadable from https://confluence.ihtsdotools.org/display/DOCGLOSS/textual+definition, RxNorm normalized names and codes standardized list downloadable from https://www.nlm.nih.gov/research/umls/rxnorm/index.html, and LOINC core definitions downloadable from https://loinc.org/downloads/. Primary-foreign key relations were numerically encoded and not usable for this research purpose.

3.3 The Core Reference Ontology Development Method

The first problem addressed was the selection of the ontology development method that produces a hierarchy of primitive ontologies. General ontology learning methods are clustering, syntactic similarity, extraction patterns, hierarchical decision tree, semantic lexicon construction, and information content. Since ontology learning is a relatively new field, only two standards

have been applied for evaluation of learned ontologies: human expert evaluation and comparing the learned ontology to a previously learned gold-standard ontology. Neither were available for this research. Rather, this research applied text mining and analysis within a SENSUS-like method to develop the primitive patient electronic medical records semantics integration ontology. The primitive semantics integration ontology was verified using Gomez-Perez's (1996, 1999, 2001, 2004) method for evaluating and verifying taxonomies and ontologies against Welty and Guarino's (2001) definitions of a proper ontology and Rector's (2003) normalization and modularization criteria for primitive ontology structure.

The second problem addressed was what primitive breadth is necessary and sufficient to assure semantic translation among ontologies and terminologies with minimal human intervention. Coverage was applied as the metric to evaluate core reference primitive breadth.

The third problem addressed was identifying the limits of ontological semantics completeness such that incomplete or missing hierarchical branches can be identified. Concept lattices of the learned ontology were developed and tested for core reference ontological closure and completeness using Formal Concept Analysis.

To address these problems, the general strategy for building the patient electronic medical records interoperability ontology was to apply text mining as the logical basis for identifying seed terms (primitive concepts) and hierarchical path interrelationships within a SENSUS-like ontology method and to verify ontological properness applying Welty and Guarino's (2001) criteria, normalization and modularity applying Rector's criteria, and completeness, closure, coupling, and cohesion using Formal Concept analysis. The outcome objective of this strategy is a human understandable theoretical basis for the core reference

ontology and a machine readable hierarchical taxonomic logic shareable across the medical terminology SNOMED CT, RxNorm, and LOINC domains.

The extraction and definition of electronic medical records core reference primitive concepts followed the text mining steps of the semantic axiomatic set theory method set forth in Cotter, Mahmud, and Zahedi (2020). For the extraction and definition of EMR primitive concepts, the Qualitative Data Analysis (QDA) portions of the method were not necessary because the medical terminology was already embedded in the medical terminologies included in the ontological semantic axiomatic set theory design. The modified EMR primitive concept extraction method is set forth as follows.

Primitive Concept Extraction Method Process 1: Primitive concept taxonomic seed terms and axiomatic relationships extraction.

1. Conduct a structured literature search in the knowledge discipline of interest.
2. Build a corpus of peer reviewed articles, professional society papers, consensual terminologies, government documents, etc., that spans the discipline's body of knowledge. This research used existing formal medical terminologies as the corpus.
3. Perform text mining to extract manifest and latent candidate primitive concept categories and correlations among them as candidates for primitive concept nouns or noun phrases and axiomatic relationships at the relevant ontology level.
4. For each primitive concept candidate noun seed term, identify it in WordNet's or the domain-specific terminology's noun hypernym-hyponym hierarchy.
    a. If the candidate noun seed term is the hypernym concept primitive, specify its definition, "is-a" existential primitive attributes and "has-a" state-modification attributes in the ontology dictionary and synset terms in the ontology thesaurus.

b. If the candidate noun seed term is not the hypernym primitive concept but is a synonym of the hypernym concept primitive, compare the hypernym concept primitive's WordNet or terminology definition to the candidate noun term's usage in the discipline's corpus. If the WordNet or terminology hypernym concept primitive's meaning can be substituted for the noun candidate term with no loss in discipline-specific meaning, specify the noun candidate term as the ontology concept primitive and specify the WordNet or terminology hypernym concept primitive's definition and attributes in the ontology dictionary and the WordNet or terminology synset terms, using the WordNet hypernym concept primitive as the synonym, in the ontology thesaurus.

c. If the candidate noun seed term is a hyponym of a WordNet or terminology hypernym, extract the candidate noun seed term's definition or intended meaning from the discipline's corpus. If the WordNet or terminology hypernym concept primitive's meaning can be substituted for the noun candidate term with no loss in discipline-specific meaning, specify the noun candidate term as the ontology concept primitive, and specify the WordNet or terminology hypernym concept primitive's definition and attributes in the ontology dictionary and the WordNet or terminology synset terms, using the WordNet hypernym concept primitive as the synonym, in the ontology thesaurus.

5. If a candidate noun seed term is not included in WordNet's or the terminology's noun hypernym-hyponym hierarchy or its synonym or hyponym discipline-specific meaning cannot be substituted for the WordNet or terminology hypernym's definition:

a. Gather candidate noun seed terms into a two-way contingency table by joint frequencies, estimate and rank terms by marginal frequencies (rank 1 = highest frequency, rank 2 = next highest frequency, etc.), and apply Santus, Lenci, Lu, and Walde's [67] $SLQS(w_1, w_i)$ entropy measure to each ranked term relative to the rank 1 term to determine semantic generality. Determine differences in rank by plotting each $SLQS(w_1, w_i)$ entropy measure, beginning with $SLQS(w_1, w_1) = 0$, versus rank on a Scree plot.

b. A smooth Scree plot with no obvious inflection point indicates the strict order of generality with the rank 1 term being the hypernym of the candidate terms. For this case, the correlations between the primitive concept term and other primitive concept terms are those estimated from text mining.

c. A Scree plot with two or more terms at and above the first inflection point on the Scree plot indicates equivalence of generality of those terms. Integrate the semantically equivalent terms into a latent primitive term that communicates the discipline's intended meaning. Integrate the semantically equivalent terms' individual correlations with the other primitive terms into a weighted correlation $Cor(L_p, P_j) = \Sigma_{ij} (f_i \, Cor(L_i, P_j) / \Sigma_i f_i)$, where Li is each semantically equivalent term in the new latent primitive concept term and P_j are the other primitive concept terms correlated with each L_i.

Primitive Concept Extraction Method Process 2: Encoding the ontology and linking primitive concept seed terms.

6. Encode the noun primitive concepts and the axiomatic relationships in an ontology editor and test for controlled natural language consistency and consistency to OWL 2.

7. Test the noun "is-a" existential primitive attributes for Welty and Guarino's (2001) proper taxonomy characteristics of rigidity, identity, unity and dependence. Test the structure of the noun primitive concepts and the axioms for Rector's (2003) normalization and modularity. Test the noun-attribute relationships for completeness and closure through the construction of a Formal Concept Analysis lattice (1999). Meeting these criteria ensures that the primitive ontology meets Gómez-Pérez's evaluation criteria for inconsistency errors and incomplete errors.

For the core reference EMR ontology, the MedTerms Medical Dictionary at https://www.medicinenet.com/script/main/alphaidx.asp?p=o_dict was used as a supplement to the WordNet dictionary in Primitive Concept Extraction Method Process 1.

3.4 Verifying the Primitive Ontology

The developed core reference patient medical records ontology was coded in Fluent Editor using controlled natural language. During encoding, concept classes and attributes definitions were verified using Fluent Editor's Validate RL+ for consistency with the World Wide Web Consortium (W3C®) Web Ontology Language OWL2 semantic profiles.

In the second verification step, Gomez-Perez's (1996, 1999, 2001, 2004) process for evaluating and verifying taxonomies and ontologies was applied to assess meeting Gruber's (1995) ontological design criteria of clarity, coherency, extendibility, minimal encoding bias, and minimal ontological commitment was applied. Formally, Gomez-Perez's process evaluates for the following errors.

- Inconsistency errors

- *Circularity errors* result from a concept being defined as a semantic specialization or generalization of itself. Taxonomic circularity errors are tested by the distance criteria. No circularity exists at a distance 0; that is, the concept is a unique concept. Circularity errors of distance 1 … n means that a concept has a semantically equivalent definition in subclass 1 … n.
- *Partition errors* result from disjoint decompositions.
  - *Common classes in disjoint decompositions* occur when there is a partition of a concept class A $\{a_1, a_2, \ldots, a_n\}$ into class A $\{a_1, a_2, \ldots, a_i\}$ and class B $\{a_j, a_k, \ldots, a_n\}$.
  - *Common instances in disjoint decompositions* occur when several instances belong to more than one class of a disjoint decomposition.
  - *External instances in exhaustive decompositions* occur when there is an exhaustive decomposition of all concept classes and some instances of a class A $\{a_j, a_k, \ldots, a_n\}$ do not belong to any class.
- *Semantic or instance errors* result from an incorrect semantic or instance classification.

- *Incomplete errors* result from the over-specification or imprecise specification of a concept class.
  - *Incomplete concept classification* results from an incomplete decomposition of the knowledge in a concept class.
  - *Partition errors* result when disjoint and exhaustive knowledge among classes is incompletely defined.

- *Disjoint knowledge omission* occurs when a set of subclasses is omitted in the taxonomy.
- Exhaustive knowledge omission occurs when a class is decomposed into two or more subclasses that carry the same knowledge.

- *Redundancy errors* occur in a taxonomy when there is more than one axiomatic hierarchical definition of a subclass relationship or there exists more than two classes or instances with the same formal definition.
  - Redundancies of Subclass-Of relations.
  - Redundancies of Instance-Of relations.
  - Identical formal definitions of two or more classes.
  - Identical formal definitions of two or more instances.

The result of applying Gomez-Perez's criteria is verification that the core reference hierarchical primitive ontological taxonomy is composed of maximally separated, axiomatically logical conceptual categories.

The third verification step was verification of a proper ontology structure by applying Guarino and Welty's (2000) and Welty and Guarino's (2001) subsumption criteria for concept "is-a" attributes and Rector's (2003) criteria for hierarchical "is-kind-of" attribute relationships. Welty and Guarino specify that for arbitrary properties (attributes), the statement "ψ subsumes ϕ, to mean that, necessarily:"

$$\forall x\ \phi(x) \rightarrow \psi(x) \qquad (1)$$

Welty and Guarino develop "is-a" attribute proper subsumption on the philosophical concepts of *rigidity*, *identity*, *unity*, and *dependence*. Refer to Guarino and Welty (2000) and Welty and Guarino (2001) for the proofs linking these philosophical concepts to "is-a" attribute proper

subsumption. To assure accuracy of specification, this work directly quotes Guarino and Welty's "is-a" attribute proper subsumption definitions.

Rigidity depends on the concept of essentiality. Welty and Guarino (2001, p. 57) define three levels of rigidity:

> ***Definition 1:*** A *rigid property* is a property that is essential to *all* its (concept's) instances, i.e., a property ϕ: $\Box(\forall x t\ \phi(x, t) \rightarrow \Box\forall t'\ \phi(x, t'))$.
>
> ***Definition 2:*** A *non-rigid property* is a property that is not essential to *some* of its (concept's) instances, i.e., a property ϕ: $\Diamond(\exists x, t\ \phi(x, t) \cap \Diamond (\exists t' \neg \phi(x, t'))$.
>
> ***Definition 3:*** An *anti-rigid property* is a property that is not essential to *all* its (concept's) instances, i.e., a property ϕ: $\Box(\forall x t\ \phi(x, t) \rightarrow \Diamond(\exists t' \neg \phi(x, t'))$.

where $\Box\phi$ means necessarily true in all possible worlds and $\Diamond\phi$ means possibly true in at least one possible world. Rigid properties are designated with +R, non-rigid properties with -R, and anti-rigid properties with ~R.

Welty and Guarino (2011, pp. 58-59) refer to the philosophical concept of *identity* as the ability to distinguish a specific instance of a concept class from other instances of the same class by means of at least one of its characteristic properties. Welty and Guarino (2011, pp. 58-59) define "… an *identity condition* (*IC*) for an arbitrary attribute property □ …as a suitable relation □□satisfying:"

$$\phi(x) \cap \phi(y) \rightarrow (\rho(x, y) \leftrightarrow x = y) \qquad (2)$$

This definition admits the following definitions of identity:

> ***Definition 4:*** An IC is a *sameness* formula Σ that satisfies either of the following conditions assuming the predicate E for actual existence.
>
> $$\Box(E(x, t) \cap \phi(x, t) \cap E(y, t') \cap \phi(y, t') \cap x = y \rightarrow \Sigma(x, y, t, t') \qquad (3)$$

$$\Box(E(x, t) \cap \phi(x, t) \cap E(y, t') \cap \phi(y, t') \cap \Sigma(x, y, t, t') \rightarrow x = y) \quad (4)$$

Definition 5: Any property *carries* an IC iff it is subsumed by a property supplying this IC, including the case where it supplies the IC itself. This property is marked as +I attribute.

Definition 6: A property □ *supplies* and IC iff (i) it is rigid, (ii) there is an IC for it, and (iii) the same IC is not carried by *all* the properties subsuming □. Therefore, +O attribute.

Definition 7: Any property carrying an IC is called a *sortal*.

A property carrying an IC is designated as +I (–I otherwise), and any property supplying an IC is designated as +O (–O otherwise).

Conversely, Welty and Guarino (2011, p. 55) note that unity is "… the problem of distinguishing the *parts* of an instance from the rest of the world by means of a *unifying relation* that binds the parts, and only the parts together." Based on this concept, Welty and Guarino (2011, pp. 59-60) define unity as:

Definition 8: An object *x is a whole under* ω iff ω is a relation such that all the members of a certain division *x* are linked by ω, and nothing else is linked by ω.

Definition 9: A property ϕ *carries a unity condition* (UC) iff there exists a single relation ω such that each instance of ϕ is *necessarily* a whole under ω.

Definition 10: A property has *anti-unity* if every instance of the property is not necessarily a whole.

Welty and Guarino recognize three types of unity– (1) *Topological* based on a physical relationship; (2) *Morphological* based on some combination of topological unity and shape; and (3) *Functional* based on functional purpose. Any attribute property carrying an UC is designated

as +U (–U otherwise). Any attribute property that has anti-unity is designated as ~U, but ~U implies –U.

Welty and Guarino (2011) distinguish between *intrinsic* and *extrinsic* properties based on whether they depend on the properties of their own concept entities and instances or the properties of other concept entities and instances. An intrinsic property is inherent to the concept entity or instance, whereas an extrinsic property is at least partially dependent on the properties of other concept entities or instances. Welty and Guarino (2011, p. 60) define dependence as:

> ***Definition 11:*** A property ϕ is externally dependent on a property ψ if, for all its instances x, necessarily some instances of ψ must exist, which is neither a part nor a constituent of x:

$$\forall x \, \Box(f(x) \rightarrow \exists y \, \psi(y) \cap \neg P(y, x) \cap \neg C(y, x)) \quad (5)$$

An externally dependent attribute property is designated as +D (–D otherwise).

At the core reference ontology level, Welty and Guarino define a proper taxonomy as one that possesses the combinations of *rigidity*, *identity*, *unity*, and *dependence* as illustrated in Table 2.

Table 2: Core Reference Ontological Property Kinds.

	Property Combination			
Meta-Property	Rigidity	Identity	Unity	Dependence
Category	+R	+O, -I	+U	+D
				-D
Role	~R	+O, -I	+U	+D
Attribute	~R	+O, -I	+U	-D
	-R			+D
				-D

To assure a primitive taxonomy, Rector (2003) added the criteria of *modularity* and *explicitness* to Guarino and Welty's criteria for a proper taxonomy. Rector set forth a two-step normalization. First, assure a proper ontology relative to Welty and Guarino's criteria. Second, normalize the ontology to assure a primitive architecture. Rector defines a primitive taxonomy as one that has *"… independent disjoint skeleton … restricted by simple trees"* (p. 1). The essence of Rector's normalization proposal is that a primitive ontology "… should consist of disjoin homogeneous trees" (p. 2).

- Each concept can have one and only one primitive parent.
- Each categorical branch of a primitive ontology must be logical and homogeneous.
- Each primitive ontology must clearly distinguish self-standing concepts and explicit partitioning among self-standing concepts.
- Subsumption of each primitive concept by one and only one other primitive concept.

To normalize a proper ontological taxonomy, Rector proposed applying relational database normal forms. Formal definitions of normal forms are set forth as follows (Vieria, 2007, 157-158).

- First Normal Form (1NF): Eliminate repeating duplicate groups of data [concepts] to guarantee Atomicity (data [concept attributes] that are self-contained and independent).
- Second Normal Form (2NF): Every row of data [instance] in a 1NF table [primitive ontology] must be unique and depend only on the table's whole key [the concept's attributes].
- Third Normal Form (3NF): A table [primitive ontology] must be in 2NF and no column data in any row [sub-concept] can have any dependency [equivalent attributes] on any other non-key column [sub-concept] (i.e., data in one column cannot be derived from the data in any other column [sub-concept attributes in one hierarchical branch cannot be derived from another sub-concept hierarchical branch]).
- Boyce-Codd Normal Form (BC-NF):
  - All candidate keys are composite keys [all composite concepts are derivable only from independent parent concepts or other composite concepts themselves derived ultimately from independent parent concepts].
  - There is more than one candidate key [composite concept].
  - The candidate keys [composite concepts] each have at least one column [concept] that is in common with another candidate key [concept].

- Fourth Normal Form (4NF): No data column [sub-concept] may depend on another column [sub-concept] other than a primary key column and depends on the whole primary key [class concept or composite concept].
- Fifth Normal Form (5NF): A table [proper ontology] must be in 4NF, and if a table is decomposed further to eliminate redundancy and anomaly, when the decomposed tables [primitive ontologies] are re-joined by means of candidate keys [concepts], the original data [concept attributes] may not be lost and no new records [concept attributes] must arise.

In seeking to assure a primitive ontological architecture, Rector's goals were ontology re-use, maintainability, and evolution. Development of a hierarchical primitive ontological architecture at each ontological level also assures meeting Gruber's criteria of clarity, coherency, extendibility, minimal encoding bias, and minimal ontological commitment.

Rector noted the following issues that must be addressed in transforming a proper ontology to a primitive ontology.

- The notion of a "primitive concept" and "primitive sub-concepts" hierarchically dependent on only their respective primitive parent concept can be difficult to demonstrate.
- Whether or not a concept should be part of a primitive ontology might be better expressed by metaknowledge; however, not all ontology languages permit reasoning over metaknowledge. Rector advocates that the criterion for concept normalization include specifications of "self-standing" and "partitioning" concepts.

- The notions of ontology normalization and ontology views are not established in ontology theory. Rector advocates a provision for concept axes to demonstrate separation.
- Provide concept indexing pointers. If an ontology is modular, the same information will point to only one primitive branch. Under this approach, concept lattices inferred from normalized and well modularized ontologies will be complete and closed under Formal Concept Analysis.

This research assured normalization to achieve primitive hierarchical dependence through restricted definition of each primitive concept's primitive "is-a" attributes to meet the criteria of coverage, completeness, and closure.

Formal Concept Analysis has long been applied in knowledge discovery (Poelmans, Elzinga, & Dedene, 2010) knowledge processing (Poelmans, Ignatov, Kuznetsov, & Dedene, 2013), and ontology learning (Cimiano, Hotho, and Staab, 2005). The Complete Lattice definition, Closure Operator definition, and Basic Theorem of Concept Lattices (Ganter and Wille, 1999) are necessary and sufficient to demonstrate the formalism of hierarchical primitive ontology branches within concept lattices.

> ***Complete Lattice Definition*:** An ordered set $V := (V, \leq)$ is a **lattice** if for any two elements x and y in V the supremum $x \vee y$ and the infimum $x \wedge y$ always exist. V is called a **complete lattice** if the supremum $\vee X$ and the infimum $\wedge X$ exist for any subset of X of V. Every complete lattice V has a largest element $\vee V$ called the **unit element** of the lattice, denoted by $\mathbf{1}_{\vee}$. Dually, the smallest element $\mathbf{0}_{\vee}$ is called the **zero element** (Ganter and Wille, 1999; p. 5).

***Closure Operator Definition*:** A closure operator φ on G is a map assigning a closure $\varphi X \subseteq G$ to each subset $X \subseteq G$ under the following conditions:

(1) $X \subseteq Y \Rightarrow \varphi X \subseteq \varphi Y$, monotony.

(2) $X \subseteq \varphi X$, extensity.

(3) $\varphi\varphi X = \varphi X$, idempotency.

Closure Theorem: If $\mathcal{U}$ is a closure system on G then

$$\varphi_{\mathcal{U}} X := \cap \{A \in \mathcal{U} \mid X \subseteq A\} \tag{6}$$

defines a closure operator on G. Conversely, the set

$$\mathcal{U}_{\varphi} := \{ \varphi X \mid X \subseteq G\} \tag{7}$$

of all closures of a closure operator φ is always a closure system, and

$$\varphi\mathcal{U}_{\varphi} = \varphi \qquad \text{and} \qquad \mathcal{U}\varphi_{\mathcal{U}} = \mathcal{U} \tag{8}$$

Proof provided by Ganter and Wille (1999, p. 8).

***Basic Theorem on Concept Lattice*s**: The concept lattice $\mathcal{B}(O$ objects, A attributes, I relations) is a complete concept lattice in which infimum and supremum are given by:

$$\wedge_{t \in T} (O_t, A_t) = (\cap O_t , (\cup A_t)'') \tag{9}$$

$$\wedge_{t \in T} (O_t, A_t) = ((\cup O_t)'', \cap A_t) \tag{10}$$

A complete lattice $\boldsymbol{V}$ is isomorphic to $\mathcal{B}(O, A, I)$ if and only if there are mappings $\gamma : O \rightarrow \boldsymbol{V}$ and $\mu : A \rightarrow \boldsymbol{V}$ such that $\gamma(O)$ is supremum-dense in $\boldsymbol{V}$, $\mu(A)$ is infimum-dense in $\boldsymbol{V}$, and oIa is equivalent to $\gamma o \leq \mu a$ for all $o \in O$ and all $a \in A$.

Proof provided by Ganter and Wille (1999, pp. 20-22).

Algebraic decomposition of closed and complete concept lattices provides the means for identifying hierarchical primitive ontology branches within concept lattices. This research

adapts the formal definitions of cohesion and coupling from software engineering (Lindig and Snelting, 1997) to define modular primitive concepts.

Modular Concept Object Definition: A modular concept object (*MCO*) consists of a set of set of objects $o \subseteq O$ and a set of attributes $a \subseteq A$ such that $\forall a \in A$, $o \in O$: $(o, a) \in V \Rightarrow a \in A$ and $\forall o \in O$, $a \in A$: $(o, a) \in V \Rightarrow o \in O$, where the MCO $\subseteq O \times A$.

Thus, in a modular concept object, all objects *O* have only attributes *A*, and all attributes *A* only describe objects *O*.

In order to map a modular concept object to Rector's proper ontology normal forms, we need a definition of the term "cohesion." Cohesion indicates the strength of relationship among modular objects *O* in an MCO via shared attributes *A*.

Cohesion Definition: A MCO (o, a) has *maximal cohesion* if $\forall o \in O, a \in A : (o, a) \in V$. A MCO $((o, \bar{o}), (\bar{a}, o))$ has *normal cohesion* if $\exists\, \bar{o} \in O\ \forall a \in A : (\bar{o}, a) \in V$ and $\exists\, \bar{a} \in A\ \forall o \in O : (o, \bar{a}) \in V$.

Maximal cohesion means that two or more concept objects within an MCO are described by the same attributes. Conversely, two sets of attributes maximally interfere if they describe the same concept objects. Normal cohesion means that concept objects in an MCO are not described by the same attributes (each concept object is described by at least one attribute not used by the other objects in the MCO).

Coupling indicates the strength of relationship among modular concept objects via shared objects *O* and attributes *A*.

Coupling Definition 1: Let $O_1 \in MCO_1$ and $O_2 \in MCO_2$ be two modular concept objects and let $a \in A$ be an attribute. MCO_1 and MCO_2 be are coupled via *a*, iff $a \in O_1 \cap O_2$.

Coupling Definition 2: Let $A_1, A_2 \in A$ be two sets of disjoint attributes, and let $o \in O$ be an object. Then $A_{1,2}$ interfere via o, iff $o \in A_1 \cap A_2$.

Coupling definition 1 states that two conceptual objects are coupled if they require the same global attribute (or some intersection of global attributes) to define their respective existence. Similarly, two sets of attributes interfere if they are used to define the existence of the same conceptual object.

The Complete Lattice and Closure Operator definitions, Basic Theorem of Concept Lattices, cohesion and coupling definitions can be combined with tree structures from graph theory to specify the properties of a proper, normalized primitive ontology.

Basic Tree Theorem: Let *T* be a graph *G* with *n* vertices. Then, *T* has the following properties:

(i) *T* is a *tree*;

(ii) *T* contains no cycles and has $n - 1$ edges;

(iii) *T* is connected and has $n - 1$ edges;

(iv) *T* is connected and each edge is a bridge;

(v) Any two vertices of *T* are connected by exactly one path; and

(vi) *T* contains no cycles, but the addition of any new edge creates exactly one cycle (proofs provided by Wilson, 1996, p. 44).

A forest is a collection of connected trees that itself forms a tree with no cycles.

Forest Corollary: If *G* is a forest with n vertices and *k* components, then *G* has $n - k$ edges (Wilson, 1996, p.44).

Spanning Forest Theorem: If *T* is any spanning forest of a graph G, then

(i) Each cutset of G has an edge in common with T; and

(ii) Each cycle of G has an edge in common with the complement of T (proofs provided by Wilson, 1996, p. 45).

Under the assumption of maximal cohesion within only concept object sets, each MCO(*O*, *A*) cross table corresponds to maximal primitive ontology rectangles in attributes. Absence of couplings or interferences of attributes among concept leads to a pure, modular primitive ontological tree structure.

3.5 Potential Research Benefits

The primary benefit of this research is a first demonstration of the capability of a core reference, hierarchical primitive ontological architecture and concept attributes definitions to integrate and resolve non-interoperable semantics among and extend coverage across existing clinical, drug, and hospital ontologies and terminologies.

3.6 Potential Methodology Risks and Limitations

The primary risks of this research were set forth as problems needing resolution in section 3.3 above. As part of the SENSUS-like ontology development method, algorithms will have to be developed to identify (1) the primitive depth necessary and sufficient to assure semantic translation among ontologies and terminologies with minimal human intervention and (2) ontological semantics completeness such that incomplete or missing hierarchical branches can be identified. The primary limitation with this research is the inability to access SNOMED CT, RxNorm, and LOINC directly, having instead to use only their glossary textual definitions, normalized names and codes, and core definitions. Since primary-foreign key relations were numerically encoded and not usable for this research purpose, some *a priori* specified axiomatic

interrelationships among categories and terms may not be fully discovered by this methodology. Conversely, it is expected that latent axiomatic interrelationships not currently encoded among SNOMED CT, RxNorm, and LOINC will be discoverable by this hierarchical primitive ontology development methodology.

Similarly, this research did not address identification and encoding of modular ontological branches. In his work, Rector did not succinctly delineate primitive from modular hierarchies. Modular concepts are those that are common knowledge units across knowledge domains and, hence, not restricted to hierarchical primitive "is-a" attribute propagation. This work's restrictive primitive concepts "is-a" attributes definitions extend Rector's definitional criteria such that primitive concepts propagate naturally within the breadth of their combined "is-a" attributes through "has-a" attributes state modifications. Conversely, modular concepts are linked through restricted sets of "is-a" attributes which act as primary-foreign key relationships between atomic, self-contained but related units of knowledge. Future research is needed to develop axiomatic definitions and to extend the hierarchical primitive concept ontology development method to partition primitive from modular concepts and properly propagate them hierarchically.

CHAPTER 4

RESULTS

4.1 Taxonomy Classes/Categories

There are two steps to identify the taxonomy classes/categories: (1) The SNOMED CT, LOINC, and RxNorm terminologies were collected in plain text format in a corpus folder. (2) Text mining was performed using the R statistical software package "tm" to identify the classes and categories. Detailed R code and term explanations relevant to the text mining can be found in Appendix A.

The most frequent terms that appeared from the text mining are:

- English– 1045658,
- Oral– 318479,
- Drug– 250376,
- Clinic– 239966,
- Active– 177078,
- Tablet– 175466
- Solution– 113492
- Substance– 109371 and
- Topic– 102873.

To get more detailed information, the lower frequency was set to 49000, and common English words (use, random, english, find, first, however) were removed and cleaned. Figure 4 represents the frequency of words.

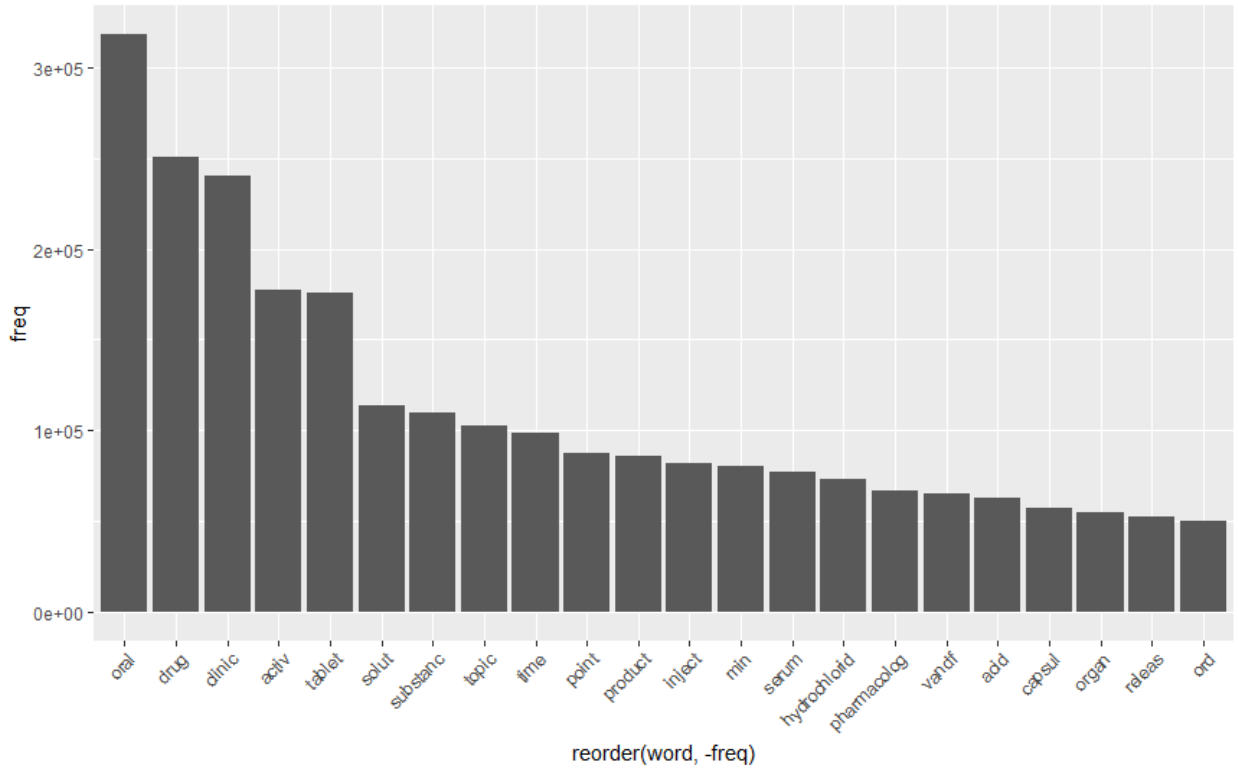


Figure 4: Frequency of Words by Order.

The words “minimum” and “additives” were kept as they relate to drug additives and minimum dosage. In parallel, to create a taxonomic structure for the ontology, hclust (cluster dendrogram) and CLUSPLOT were plotted and analyzed. By changing the sparsity and the means of the document-term matrix, multiple plots were plotted to analyze in depth and to interpret the results in text mining.

The hierarchical clustering (hclust) as shown in Figures 5, 6, and 7 are based on agglomerative hierarchical clustering strategy that works with the following logic (Mahmud, 2018):

Step 1: Assigning each observation to its own cluster.

Step 2: Identifying the pair of clusters that are closer to each other by Euclidian distance and then merging them. This means there is now one cluster less than before.

Step 3: Computing the Euclidian distance between the new cluster and each of the old clusters.

Step 4: Repeating step 2 and step 3 until it reaches a single cluster containing all the documents.

Cluster dendograms at 5%, 10%, 15%, 20%, and up to 45% sparsity were created to explore the taxonomic categories. The full sequence of diagrams are presented in Appendix B. In Figure 5, the dendrogram shows an hclust plot at 10% non-sparsity. This means 10 percent zero terms are removed from the document-term matrix (dtm). Following the Euclidean distance method and "complete" method in hclust plot, this figure shows hierarchical plot of nodes and leaves. As the sparse terms changed from 10% to 15% in Figure 6, nothing changed visibly except the cluster pattern. When 15% changed to 20% (Appendix B), a cluster mass of more terms appeared in the diagram. However, at this point it was a lot more noise than the usable terms.

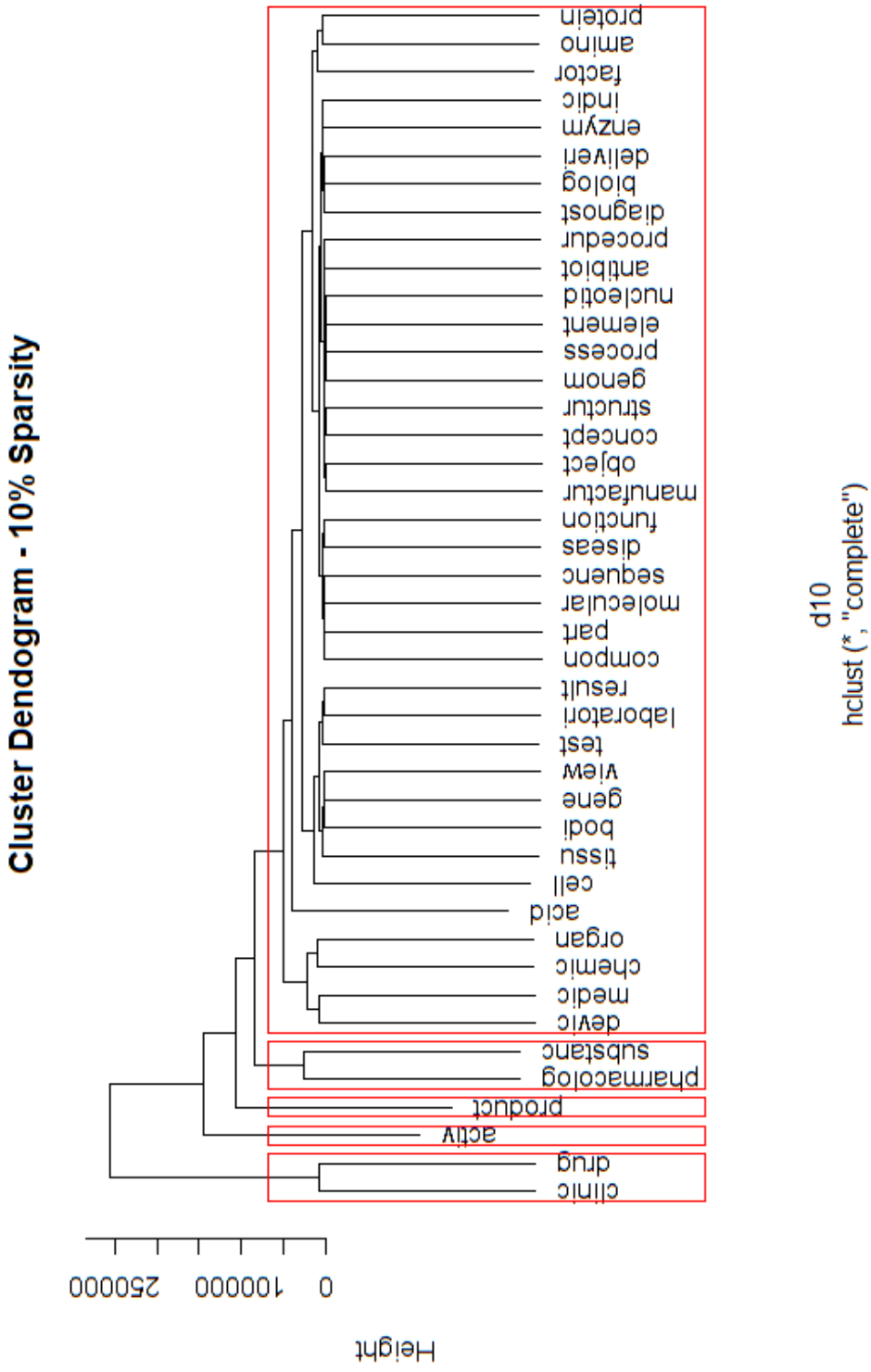


Figure 5: Cluster Dendrogram for 10% Sparsity.

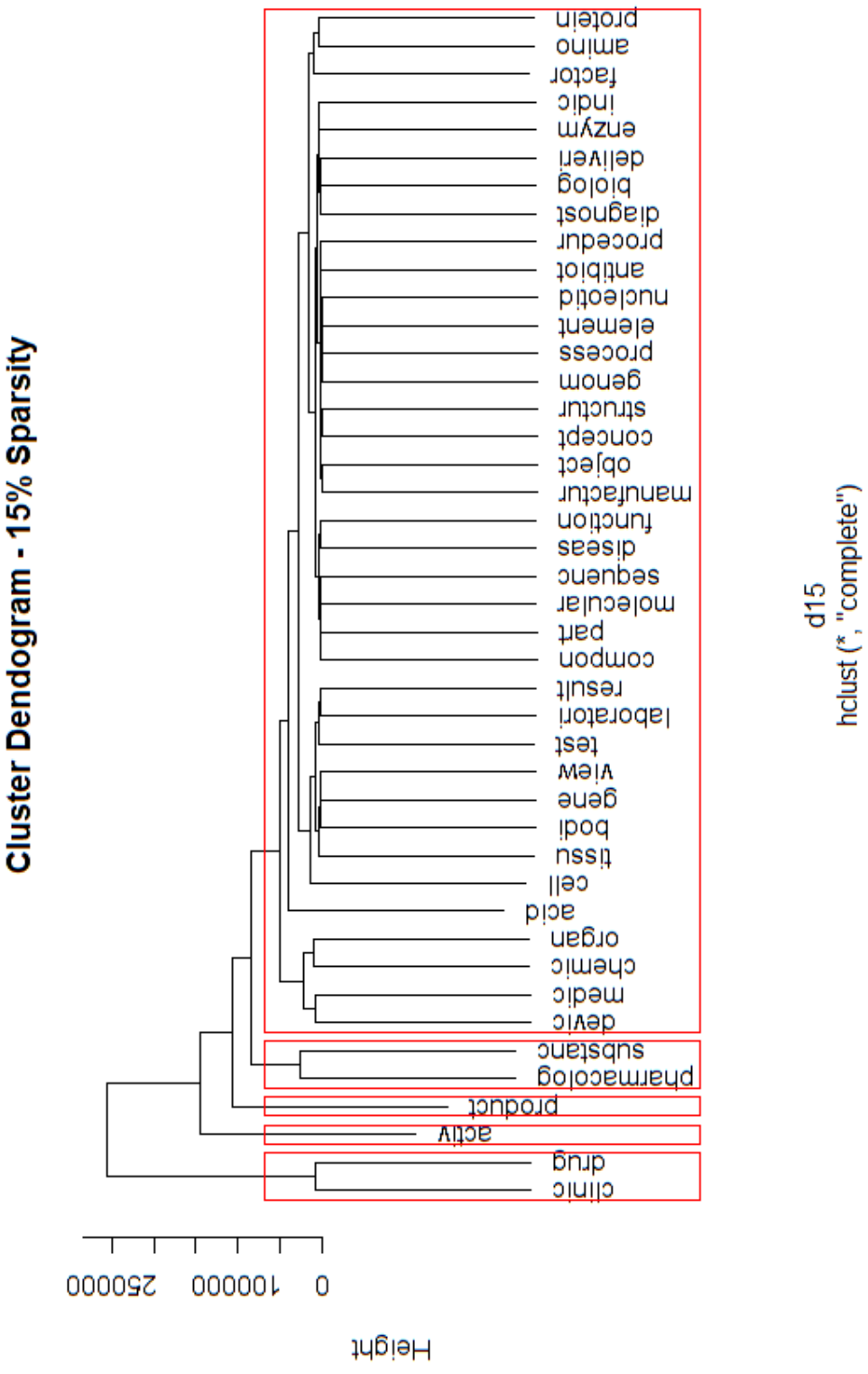


Figure 6: Cluster Dendrogram for 15% Sparsity.

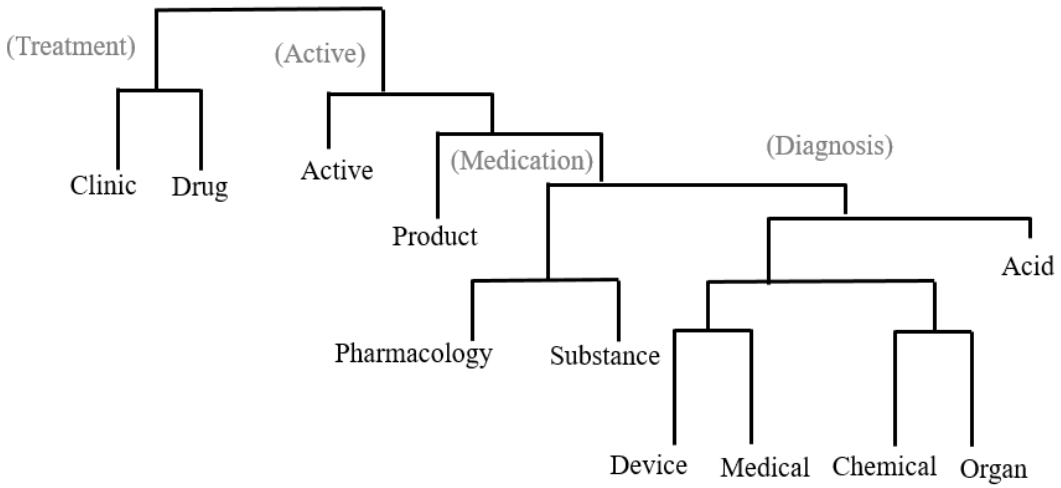


Figure 7: Summarized Cluster Dendogram.

In the above clustering analyses, the number of clusters was not pre-specified, and further analyses are needed to evaluate the data. For in-depth analyses K-means clustering where the number of clusters is pre-specified was performed. Cluster plots at 5%, 10%, 15%, and 20% sparsity with 3 through 9 means were plotted to explore the potential number of independent taxonomic categories. The full sequence of plots is presented in Appendix C.

This analysis generates K-corpus clusters, and the logic and algorithm that were used herein are similar to Mahmud's (2018) which were used for building the Foundational Ontology. The steps are below (Mahmud, 2018).

Step 1: Assigning the document randomly to k bins.

Step 2: Computing the location of the centroid of each bin.

Step 3: Computing the distance between each document and each centroid.

Step 4: Assigning each document to the bin corresponding to the centroid closest to it.

Step 5: Terminating the computation if no document is moved to a new bin. Otherwise, go to step 2.

Figures 8, 9, 10, 11, 12, 13, 14 and 15 show K-means clustering for the analyzed corpus for 4, 5, 6, and 7 clusters (K-means) with 10% and 15% sparsity respectively. The cluster plots shown in these figures work in a mathematical space whose dimensionality equals the number of concept terms in the corpus. In this case, SNOMED CT has 352,567, LOINC has 92,369, and RxNorm has 1,044,971 distinct concepts, which are substantial numbers, so it was neither feasible nor possible to visualize using normal means. To visualize, Principle Component Analysis (PCA) is applied to reduce the number of dimensions to two (component 1 and component 2) for 3, 4, 5, 6, 7, 8, and 9 clusters (in this analysis) in such a way that the reduced dimensions explain as much of the variability as possible among the clusters. The variability explained with 5% sparsity was 99.84%, but the plots are full of noise. Sparsity 10 and 15 provided plots that are acceptable with the variability of 96.44%.

Figures 8 and 9 have four clusters (K=4) with 10% and 15% sparsity respectively, and most of the core terms appeared in cluster numbers 2, 3, and 4. Figures 10 and 11 have five clusters (K=5) with 10% and 15% sparsity respectively, and most of the core terms appeared in cluster numbers 1, 2, 4, and 5. Figures 12 and 13 have six clusters (K=6) with 10% and 15% sparsity respectively, and most of the core terms appeared in cluster numbers 2, 3, 4, and 5. Figures 14 and 15 have seven clusters (K=7) with 10% and 15% sparsity respectively, and most of the core terms appeared in cluster numbers 1, 2, 5, and 6. For K=4, CLUSPLOT has four clusters, and one of them is noise. The rest of the clusters do not have the terms in clear formation. For K=5, CLUSPLOT has five clusters, and the formation becomes clearer. The term “Active” got its own cluster. For K=6, CLUSPLOT has six clusters, and the formation is almost similar to K=5. It has two noise clusters while K=5 only had one noise cluster. For K=7, CLUSPLOT has seven clusters and cluster 7 has terms “Medical” and “Devices” separated out

with few other noise terms. K=3, K=8, and K=9 clusters were also analyzed. These can be found in Appendix C.

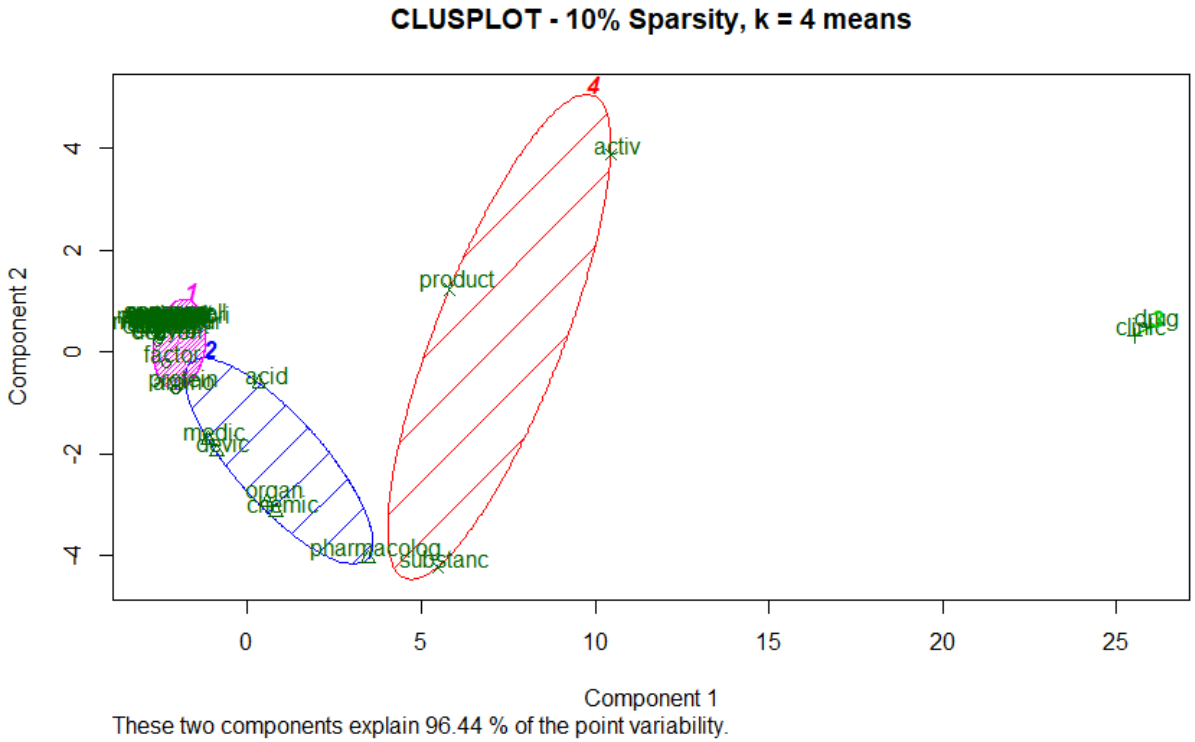


Figure 8: CLUSPLOT for 10% Sparsity, K=4 means.

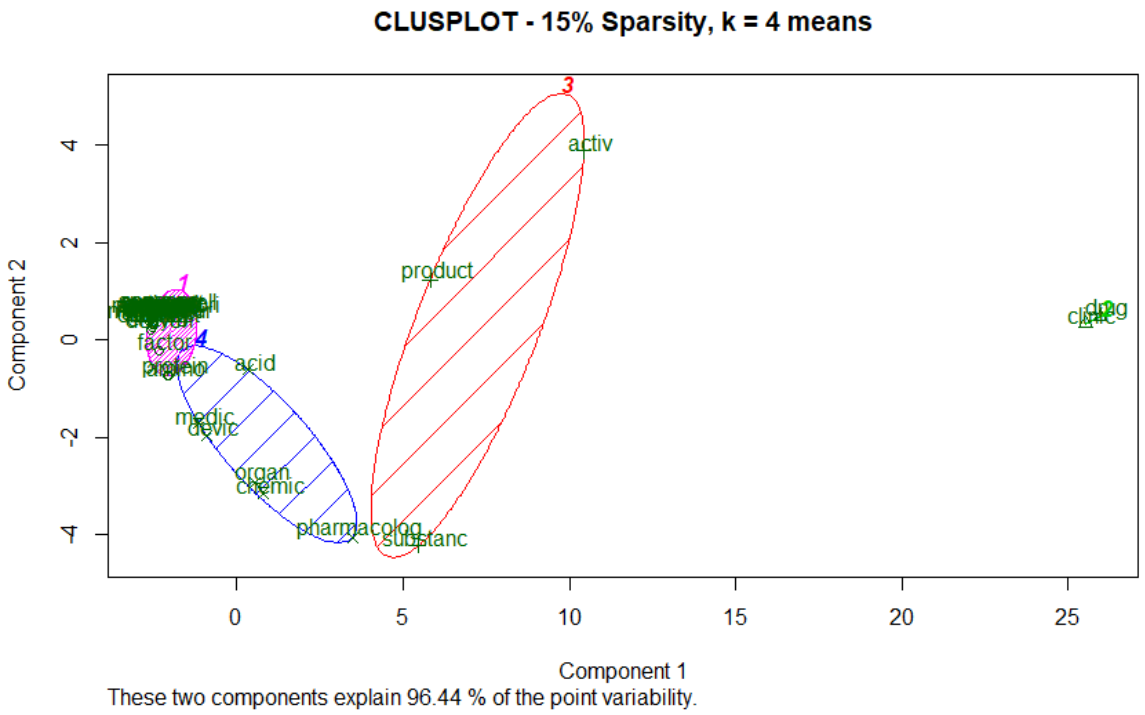


Figure 9: CLUSPLOT for 15% Sparsity, K=4 means.

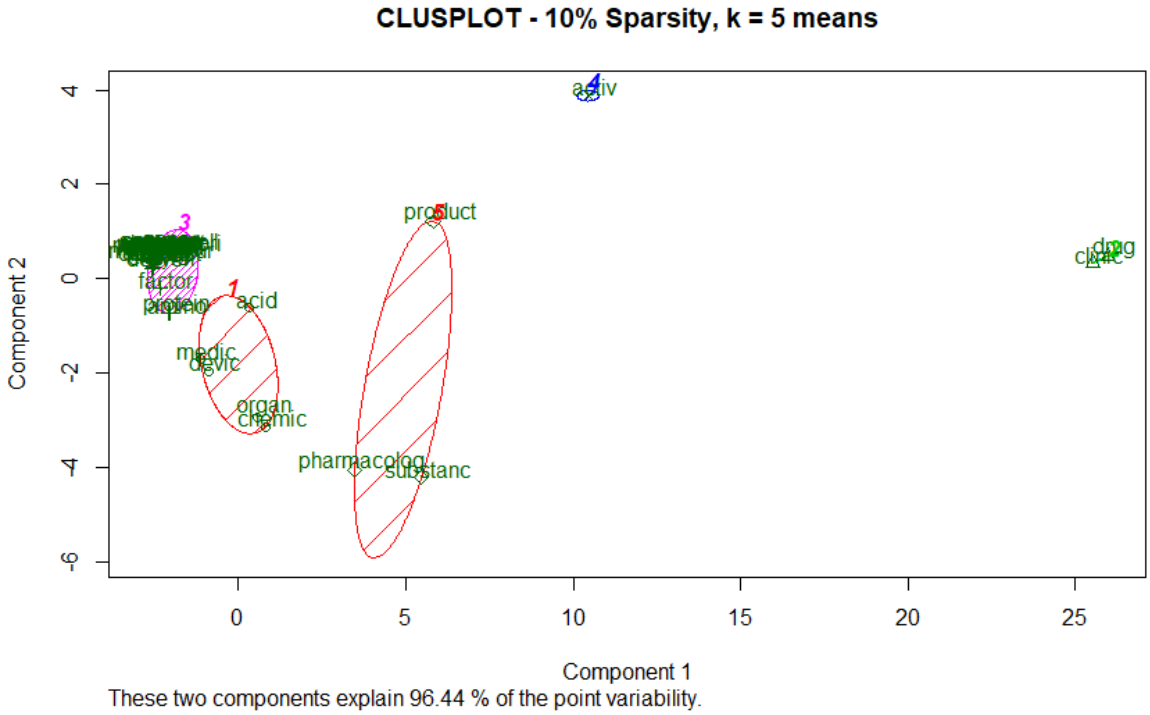


Figure 10: CLUSPLOT for 10% Sparsity, K=5 means.

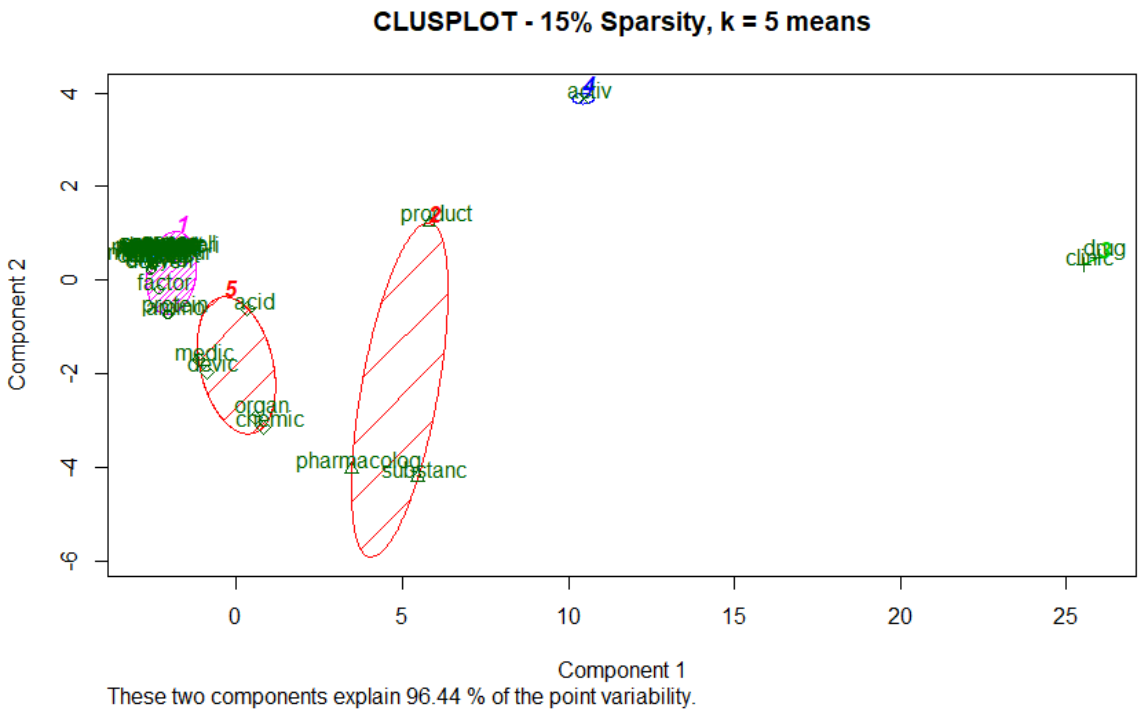


Figure 11: CLUSPLOT for 15% Sparsity, K=5 means.

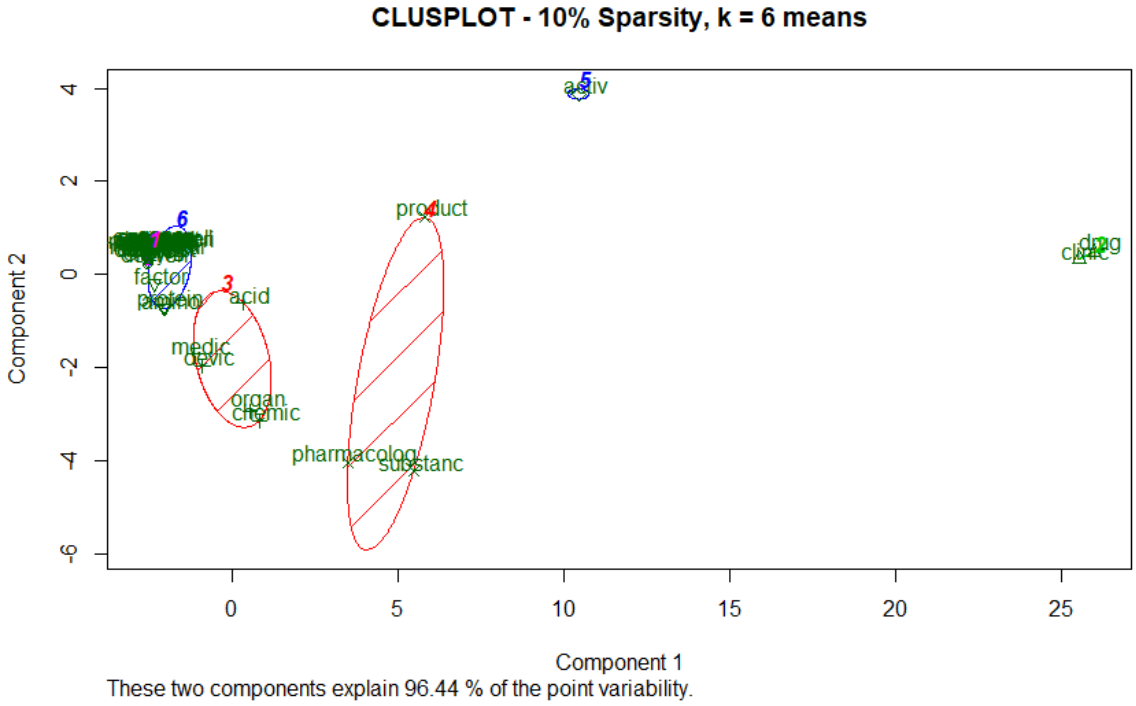


Figure 12: CLUSPLOT for 10% Sparsity, K=6 means.

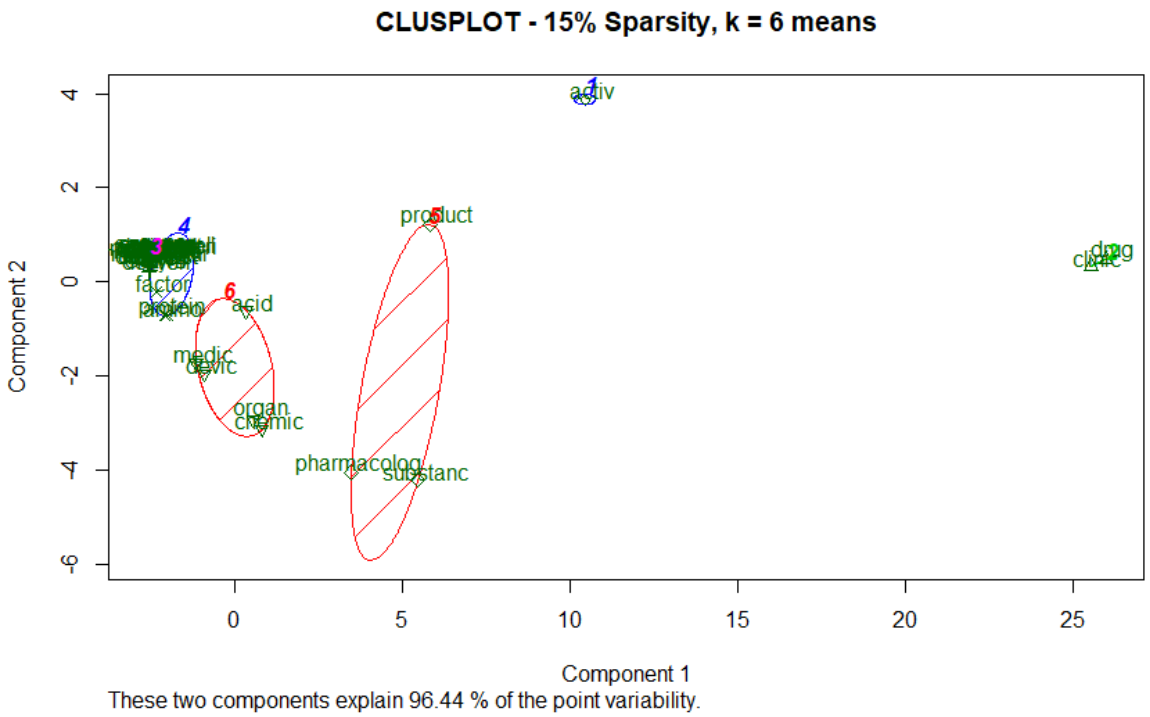


Figure 13: CLUSPLOT for 15% Sparsity, K=6 means.

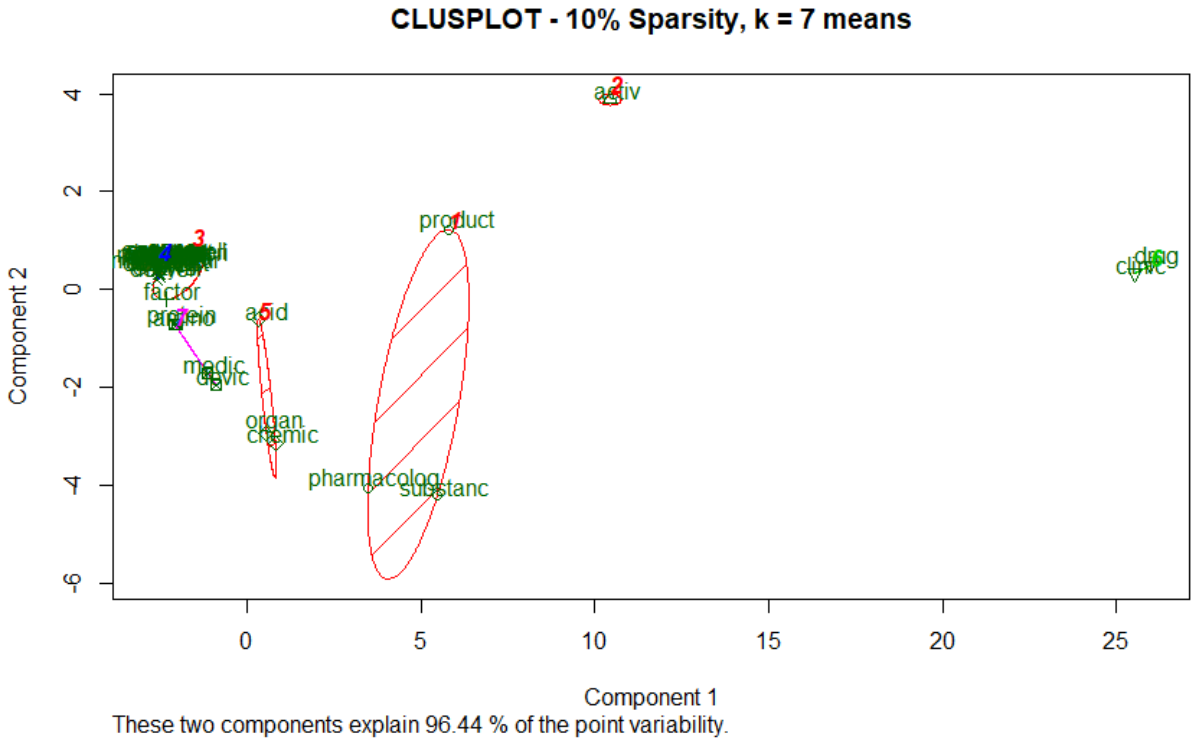


Figure 14: CLUSPLOT for 10% Sparsity, K=7 means.

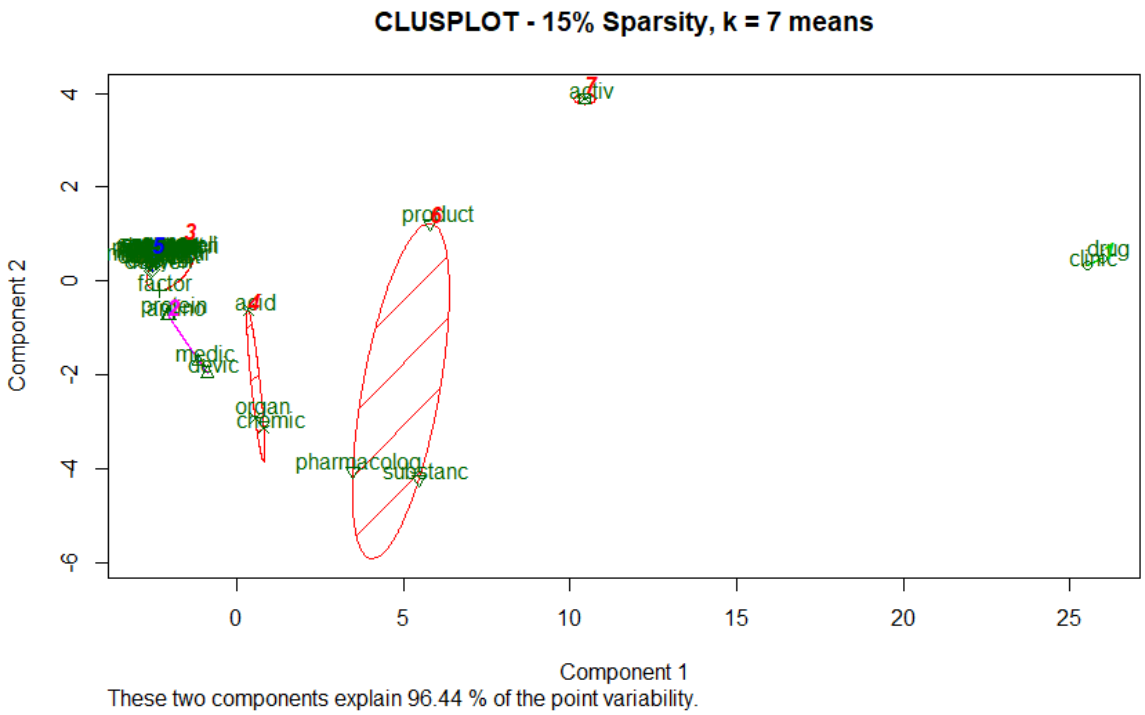


Figure 15: CLUSPLOT for 10% Sparsity, K=7 means.

To create the taxonomic hierarchy, both the cluster dendrogram and CLUSPLOT were evaluated side-by-side. This also allowed identification of (i) the core terms and (ii) potential relationships among the terms (within the same and between different clusters). The cluster dendrogram provides an overall picture of the terms appearing in the corpus in hierarchy (and possible clusters to form). Figures 10 to 13 show that the three clusters (for K=5, the clusters were 2, 3, and 4 and for K=6, the clusters were 2, 4, and 5) contain all the major terms. The only change from K=5 to K=6 was the noise cluster #3 from K=5 became noise cluster #3 and #4 in K=6. From the above analyses, it was determined that K=5 with 10% sparsity is the stabilized version.

4.2 Ontological Relationships

Now that core taxonomic terms are identified, the next step is to find the taxonomic relationships among the terms within and outside of the clusters. To achieve that, a relationship matrix was created for finding associations within "tm" text mining by pulling out the correlations between the specified term frequency distributions and the frequency distributions in other terms in tm text mining (Table 3).

For this analysis, the minimum correlation 0.80 was used, because the joint SNOMED CT, RxNorm, and LOINC corpus were pre-specified terminologies. The correlation coefficient of 1.0 in the table is strongly correlated (being +1 is perfectly positively correlated, and 0 is not correlated) and is marked in orange. Drug and Clinic are correlated with a correlation coefficient of +1 (Table 3) which corresponds to them being in the same cluster in CLUSPLOTS (Figures 8 – 15). Clinic and Pharmacology has a +1 correlation co-efficient which defines strong axiomatic relationship.

All the correlation coefficients between the terms (Table 3) are strongly correlated ranging from 0.93 to 1.0. The lowest correlation coefficient amongst the terms in Table 3 is 0.93 which is between Medical and Organ. Active, Acid, and Product are frequently used words and top-level terms but are independent axioms as they are not strongly correlated to other top-level terms. The strong correlation amongst terms are logical as they are taken out of structured medical terminologies.

Table 3: Association Matrix

Classes	Clinic	Drug	Active	Acid	Product	Pharmacology	Substance	Device	Medical	Chemical	Organ
Clinic		1.00				1.00	0.97	1.00	0.96	1.00	0.98
Drug	1.00					1.00	0.98	1.00	0.96	1.00	0.99
Active											
Acid											
Product											
Pharmacology	1.00	1.00					0.97	1.00	0.95	1.00	0.98
Substance	0.97	0.98				0.97		0.97	1.00	0.97	0.95
Device	1.00	1.00				1.00	0.97		0.96	1.00	0.99
Medical	0.96	0.96				0.95	1.00	0.96		0.95	0.93
Chemical	1.00	1.00				1.00	0.97	1.00	0.95		0.98
Organ	0.98	0.99				0.98	0.95	0.99	0.93	0.98	

Table 4: Axiomatic Relationships between EMR Core Reference Ontology Primitive

Composite	**Primitive**	**Dependency**	**Axiom**
Treatment	Clinic	within	Clinic is strongly correlated with Drug.
		between	Clinic is strongly correlated with Pharmacology.
		between	Clinic is strongly correlated with Substance.
		between	Clinic is strongly correlated with Device.
		between	Clinic is strongly correlated with Medical.
		between	Clinic is strongly correlated with Chemical.
		between	Clinic is strongly correlated with Organ.
	Drug	within	Drug is strongly correlated with Clinic.
		between	Drug is strongly correlated with Pharmacology.
		between	Drug is strongly correlated with Substance.
		between	Drug is strongly correlated with Device.
		between	Drug is strongly correlated with Medical.
		between	Drug is strongly correlated with Chemical.
		between	Drug is strongly correlated with Organ.
Active	Active		
Medication	Product		
	Pharmacology	between	Pharmacology is strongly correlated with Clinic.
		between	Pharmacology is strongly correlated with Drug.
		within	Pharmacology is strongly correlated with Substance.
		between	Pharmacology is strongly correlated with Device.
		between	Pharmacology is strongly correlated with Medical.
		between	Pharmacology is strongly correlated with Chemical.
		between	Pharmacology is strongly correlated with Organ.
	Substance	between	Substance is strongly correlated with Clinic.
		between	Substance is strongly correlated with Drug.
		within	Substance is strongly correlated with Pharmacology.
		between	Substance is strongly correlated with Device.
		between	Substance is strongly correlated with Medical.
		between	Substance is strongly correlated with Chemical.
		between	Substance is strongly correlated with Organ.

Table 4: Axiomatic Relationships between EMR Core Reference Ontology Primitive (continued)

Composite	Primitive	Dependency	Axiom
Diagnosis	Acid		
	Device	between	Device is strongly correlated with Clinic.
		between	Device is strongly correlated with Drug.
		between	Device is strongly correlated with Pharmacology.
		between	Device is strongly correlated with Substance.
		within	Device is strongly correlated with Medical.
		within	Device is strongly correlated with Chemical.
		within	Device is strongly correlated with Organ.
	Medical	between	Medical is strongly correlated with Clinic.
		between	Medical is strongly correlated with Drug.
		between	Medical is strongly correlated with Pharmacology.
		between	Medical is strongly correlated with Substance.
		within	Medical is strongly correlated with Device.
		within	Medical is strongly correlated with Chemical.
		within	Medical is strongly correlated with Organ.
	Chemical	between	Chemical is strongly correlated with Clinic.
		between	Chemical is strongly correlated with Drug.
		between	Chemical is strongly correlated with Pharmacology.
		between	Chemical is strongly correlated with Substance.
		within	Chemical is strongly correlated with Device.
		within	Chemical is strongly correlated with Medical.
		within	Chemical is strongly correlated with Organ.
	Organ	between	Organ is strongly correlated with Clinic.
		between	Organ is strongly correlated with Drug.
		between	Organ is strongly correlated with Pharmacology.
		between	Organ is strongly correlated with Substance.
		within	Organ is strongly correlated with Device.
		within	Organ is strongly correlated with Medical.
		within	Organ is strongly correlated with Chemical.

Table 4 above shows logical axiomatic relationships within and between EMR core reference ontology design composite and primitive concepts, and the composite primitives inherit their primitive properties from their primitives.

4.3 EMR Core Reference Ontology Specification

Table 5 shows the specifications for EMR core reference ontology design. For this ontology design, only nouns have been used as primitive concepts.

Table 5: Specification of EMR Primitive Concepts.

Candidate Noun Term	WordNet Hypernym(s)	Hypernym in EMR Core Reference Ontology	Definition	Is-a Attributes	Synonyms
Clinic	Medical Institution	Clinic	Medical specialists' practice.	Medicine Practice	Dispensary
Drug	Medical substance	Drug	Matter that is used as a medicine or narcotic	Medicine Matter	Pharmaceutical
Active	Active agent	Active	Medical agent capable of producing a physiological response.	Agent Medicine Physiological Response	Pathology
Acid	Acid	Acid	Any of various water-soluble substances having a pH less than 7 and reacting with a base to form a salt	Chemical Matter pH	Anti-alkaline
Product	Product	Product	Matter formed as a result of a chemical reaction.	Chemical Reaction Matter	Chemical substance
Pharmacology	Pharmacology	Pharmacology	The science or study and application of drugs: their nature, properties, preparation, uses and effects.	Effects Medicine Treatment	Nonsurgical medicine

Table 5: Specification of EMR Primitive Concepts (continued).

Candidate Noun Term	WordNet Hypernym(s)	Hypernym in EMR Core Reference Ontology	Definition	Existential Attributes (is-a attributes)	Synonyms
Substance	Matter	Substance	Matter of a particular kind or constitution; the real physical matter of which a person or thing consists.	Matter	Substamtia
Device	Instrument	Device	A physical item used in medical treatment.	Instrument Medicine Treatment	Instrument
Medical	Examination, Scrutiny	Medical	The study or practice of medicine.	Medicine Practice Study	Aesculapian, medicinal
Chemical	Matter	Chemical	Matter produced by a reaction involving changes in atoms or molecules	Matter Reaction	Chemic
Organ	Organ	Organ	A fully differentiated, structural unit in a living entity that is specialized for some particular function.	Function Structure Unit	Unit, Element, Part

Table 5 shows that the nouns are primitive concepts in terms of EMR. Clinic's hypernym could be medical institution, but medical institution is a broader term, and it is a hypernym at the foundational ontology level. At the core reference level, clinic is the hypernym.

Drug could be medical substance (WordNet hypernym), but medical substance is broader and is primitive at the foundational level. Thus, in the EMR core reference ontology, Drug is the

primitive as it means “matter that is used as a medicine or narcotic” without any ambiguity. The same applies for Substance, Device, Medical, and Chemical.

For substance, “constituent” could be the hypernym, but it is at the foundational level. Constituent means an artifact that is one of the individual parts of which a composite entity is made up; especially a part that can be separated from or attached to a system. Thus, it could be "spare components for cars" or "a component or constituent element of a system", but none of these definitions apply to EMR. Substance in EMR means “matter of a particular kind” or constitution or “the real physical matter of which a person or thing consists.” Substance is also the term that is widely used in the medical field (as it is one of the most frequent terms that appeared in the medical terminologies) instead of constituent. For example: DNA is the substance of our genes. Thus, at the EMR core reference level, Substance is the primitive.

The hypernym for device could be instrumentation, but that is also a broader term. The meaning of Device is a “physical item used in medical treatment,” which is the hypernym in EMR. A device in EMR means medical devices or applicators used for medical purposes, but instrumentation mostly refers to music. The definition of instrumentation in WordNet refers to the instruments called for in a musical score or arrangement for a band or orchestra. Device is a frequent term that appeared in the terminologies and is the hypernym for the EMR core reference ontology.

Medical in terms of EMR core reference ontology means a physical examination without any ambiguity; thus, it is the primitive concept for EMR core reference ontology. Its hypernym could be examination but does not apply to the EMR otology as the examination refers to school/college administered examinations as well.

Like medical, chemical does not have any ambiguity when used in EMR core reference ontology, but its hypernym "material" has other meanings. The synonym of material is stuff and is used in different fields such as in engineering, building, production, and so on. Chemical in the medical terminology means the tangible substance or material produced by or used in a reaction involving changes in atoms or molecules.

The rest of the terms which are active, acid, product, pharmacology, and organ are already primitives according to WordNet.

Table 6: Attributes of EMR Primitive Concepts.

Candidate Noun Term	**Definition**	**Role**	**Existential Attributes (is-a attributes)**	**State-Modification Attributes (has-a attribute**
Clinic	Medical specialists' practice.	Health facility	Medicine Practice	Profession Curing Generalist Specialist
Drug	Matter that is used as a medicine or narcotic	Nonsurgical treatment	Medicine Matter	Profession Curing Pharmaceutic
Active	Medical agent capable of producing a physiological response.	Energetic	Agent Medicine Physiological Response	Causal Profession Curing Body Pathology
Acid	Any of various water-soluble substances having a pH less than 7 and reacting with a base to form a salt	Matter with an excess of hydrogen atoms	Chemical Matter pH	Chemic Pharmaceutic 0 to 7 potential hydrogen
Product	Matter formed as a result of a chemical reaction.	Formulation	Chemical Reaction Matter	Chemic Decomposition Synthesis Pharmaceutic

Table 6: Attributes of EMR Primitive Concepts (continued).

Candidate Noun Term	Definition	Role	Existential Attributes (is-a attributes)	State-Modification Attributes (has-a attribute
Pharmacology	The science or study and application of drugs: their nature, properties, preparation, uses and effects.	Study and application of drugs.	Effects	Pharmacology
Substance	Matter of a particular kind or constitution; the real physical matter of which a person or thing consists.	Elemental matter.	Matter	Pharmaceutic
Device	A physical item used in medical treatment.	Physical use	Instrument Medicine Treatment	Tool Profession Curing Diagnosis Prognosis
Medical	The study or practice of medicine.	Healing practice	Medicine Practice Study	Profession Curing Generalist Specialist Understanding
Chemical	Matter produced by a reaction involving changes in atoms or molecules	Composition of atoms or molecules.	Matter Reaction	Pharmaceutic Decomposition Synthesis
Organ	A fully differentiated, structural unit in a living entity that is specialized for some particular function.	Functional unit of an entity.	Function Structure Unit	Transformation Composition Element

Existential Attributes:

Existential attributes are essential for the existence of a concept. In the absence of any of these attributes, the concept would fall apart. These attributes are associated with “is-a” relationships with the concept.

State-Modification Attributes:

State-Modification attributes are required to explain a certain state of the concept. These attributes are not essential for the existence of a concept and associated with "has-a" relationships with the concept.

For each of the core primitive taxonomic terms, a list of attributes is documented in Table 6. A few attributes in the table may sound similar but have different meanings. Conversely, some attributes need more elaboration. For example, the taxonomic term "Drug" has *Medicine* and *Pharmaceutic* attributes. Drug is used in the *profession* of *Medicine* for *curing* a disease and *Pharmaceutic* plays a role in creating and distributing those cures. Thus, *Medicine* is listed as "is-a" attribute and *Pharmaceutic* as "has-a" attribute. Another example could be Organ. For Organ *Unit* is listed as an "is-a" attribute, and *Element* is listed as a "has-a" attribute. Even though they may sound similar, *Unit* is a whole of something, while *Element* is a part of something.

4.4 EMR Core Reference Ontology Design

The taxonomic classes of Figure 7, the axiomatic relationship defined in Table 4, and the attributes defined in Table 6 were encoded into an EMR core reference design ontology in Fluent Editor using its controlled natural language (CNL). Fluent Editor's controlled natural language (CNL) is a restricted English for human communication that encodes ontology semantics consistent with and translatable into description logic, SWRL rules, and OWL standards. Thus, ontologies encoded in Fluent Editor's CNL meet Gruber's criteria of clarity, coherency, extendibility, minimal encoding bias, and minimal ontological commitment. To conform strictly

with minimal ontological commitment, only the following hierarchical and axiomatic relationships were used.

Hierarchical: "is-a" existential.

"has-a" state modification.

Axiomatic: "be strongly correlated with" in accordance with definitions derived from Table 3.

Figure 16 shows the ontology developing window, and Figure 17 shows the taxonomic and axiomatic relationships that were encoded following CNL. Figure 17 demonstrates Taxonomic hierarchy from "thing." A "thing" can be either a "physical-thing" or an "abstract-thing." A physical-thing has presence in time and space whereas an abstract-thing does not have such presence.

The ontologies were materialized in OWL2-RL+ and validated with the OWL2-RL+ reasoned. The Fluent Editor CLN EMR core reference design ontology encoding is presented in Appendix D.

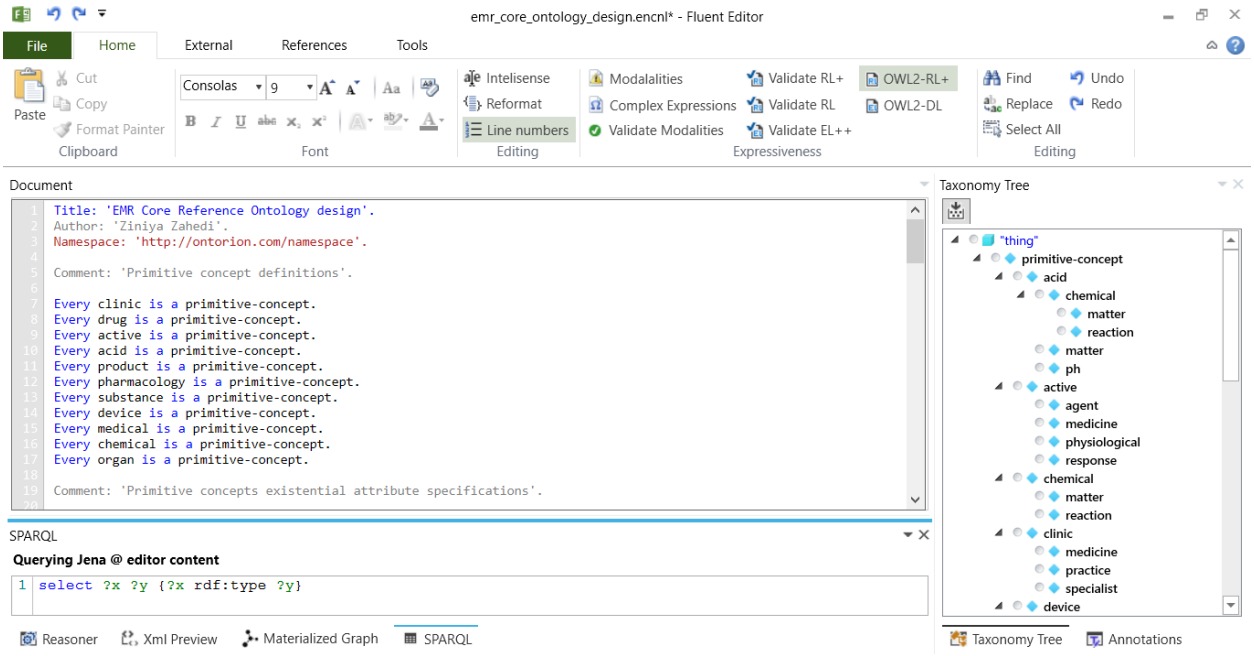


Figure 16: Fluent Editor Development Window.

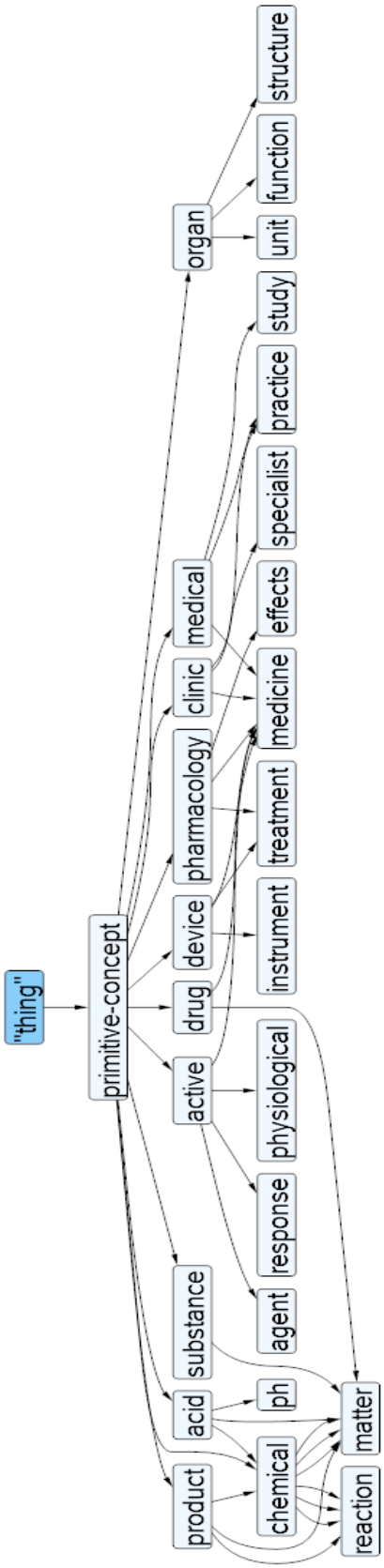


Figure 17: Fluent Editor- EMR Core Reference Ontology Design

4.5 Proofs of Ontological Concept-Attribute Relationships

Assessment of the core reference primitive ontology against Welty and Guarino's (2001) subsumption criteria for concept "is-a" attributes is set forth in Table 7. The properties of each 'is-a" attribute meet the category criteria specified in Table 2. Table 7 also demonstrates that each primitive concept acts as a primary key for its "is-a" attributes meeting Rector's (2003) normalization criteria necessary and sufficient for modularity and explicitness.

Table 7: Core Reference Primitive Ontology Design "is-a" Attribute Properties.

Candidate Noun Term	**Existential Attributes (is-a attributes)**	**Attribute Property**	**Property Combination**			
Clinic	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Practice	The exercise of a profession.	+R	+O, -I	+U	-D
Drug	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Matter	An entity that has physical existence.	+R	+O, -I	+U	-D
Active	Agent	Capable of producing a certain effect.	+R	+O, -I	+U	-D
	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Physiological	Of or consistent with an organism's normal functioning.	+R	+O, -I	+U	-D

Table 7: Core Reference Primitive Ontology Design "is-a" Attribute Properties (continued).

Candidate Noun Term	Existential Attributes (is-a attributes)	Attribute Property	Property Combination			
	Response	A bodily process occurring due to the effect of some antecedent stimulus or agent.	+R	+O, -I	+U	-D
Acid	Chemical	Material produced by or used in a reaction involving changes in atoms or molecules.	+R	-O, +I	+U	+D
	Matter	An entity that has physical existence.	+R	+O, -I	+U	-D
	pH	The number of moles of hydrogen ions per cubic decimeter that provides a measure on a scale from 0 to 14 of the acidity or alkalinity of a solution.	+R	+O, -I	+U	-D
Product	Chemical	Material produced by or used in a reaction involving changes in atoms or molecules.	+R	-O, +I	+U	+D
	Reaction	A process in which one or more substances are changed into others.	+R	+O, -I	+U	-D
	Matter	An entity that has physical existence.	+R	+O, -I	+U	-D
Pharmacology	Effects	Act to bring into existence.	+R	+O, -I	+U	-D
	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Treatment	Therapy	+R	+O, -I	+U	-D
Substance	Matter	An entity that has physical existence.	+R	+O, -I	+U	-D

Table 7: Core Reference Primitive Ontology Design "is-a" Attribute Properties (continued).

Candidate Noun Term	**Existential Attributes (is-a attributes)**	**Attribute Property**	**Property Combination**			
Device	Instrument	An instrumentality invented for a particular purpose.	+R	+O, -I	+U	-D
	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Treatment	Care provided to improve a situation.	+R	+O, -I	+U	-D
Medical	Medicine	Learned profession that is mastered in a medical school and devoted to curing diseases and injuries.	+R	+O, -I	+U	-D
	Practice	The exercise of a profession.	+R	+O, -I	+U	-D
	Study	A branch of knowledge.	+R	+O, -I	+U	-D
Chemical	Matter	An entity that has physical existence.	+R	+O, -I	+U	-D
	Reaction	A process in which one or more substances are changed into others.	+R	+O, -I	+U	-D
Organ	Function	What something is used for.	+R	+O, -I	+U	-D
	Structure	A complex entity constructed of many parts.	+R	+O, -I	+U	-D
	Unit	A specific measure of amount.	+R	+O, -I	+U	-D

Table 7 shows that all attribute properties of EMR core reference ontology are classified as +R, +O, –I, +U, and –D except for Acid-Chemical and Product-Chemical. In section 3.4, the third verification step was for a proper ontology structure by applying Guarino and Welty's

(2000) and Welty and Guarino's (2001) subsumption criteria for concept "is-a" attributes and Rector's (2003) criteria for hierarchical "is-kind-of" attribute relationships. Below are the assessment criteria (details are in section 3.4).

1. Rigid properties are designated with +R, non-rigid properties with -R, and anti-rigid properties with ~R.
2. A property carrying an Identity (IC) is designated as +I (–I otherwise), and any property supplying an Identity (IC) is designated as +O (–O otherwise).
3. Any attribute property carrying a Unity (UC) is designated as +U (–U otherwise). Any attribute property that has anti-unity is designated as ~U, but ~U implies –U.
4. An externally dependent attribute property is designated as +D (–D otherwise).

In Table 7, all the attribute properties are rigid (+R), not carrying (-I) but supplying IC (+O), carrying UC (+U), and externally independent (-D) except *Acid-Chemical* and *Product-Chemical*. For *Acid-Chemical* and *Product-Chemical,* IC, UC, and dependability are different than the rest. For both cases, the attribute property is, "material produced by or used in a reaction involving changes in atoms or molecules" which implies that *Acid* and *Product* are externally dependent on *Chemical*, and without Chemical, these two are nonexistent while for other attributes that is not the case. For example: *Chemical-Matter*'s attribute property is, "an entity that has physical existence" which implies that *Matter* is not externally dependent on *Chemical* and can exist by itself.

Concept lattices were developed to assess modularity, completeness, cohesion, coupling, and closure. In Figure 17, concepts (objects) are marked in the white boxes, and attributes are marked in the grey boxes. When a concept node contains a blue filled upper semicircle, it means that there is an attribute attached to this concept. When there is a black filled lower semicircle, it

means that there is only a concept attached. When there is a white filled upper semicircle, it means the attributes of that concept are attached to more than one concept.

Figure 18 graphically demonstrates the conformance to Formal Concept Analysis's Complete Lattice Definition, Closure Operator Definition, Basic Theorem on Concept Lattices, and the Spanning Forest Theorem. EMR core reference ontology concept lattices in Figures 19 through 29 graphically demonstrate conformance to the Modular Concept Object Definition, Cohesion Definition, Coupling Definitions, and the Primitive Ontology Definition.

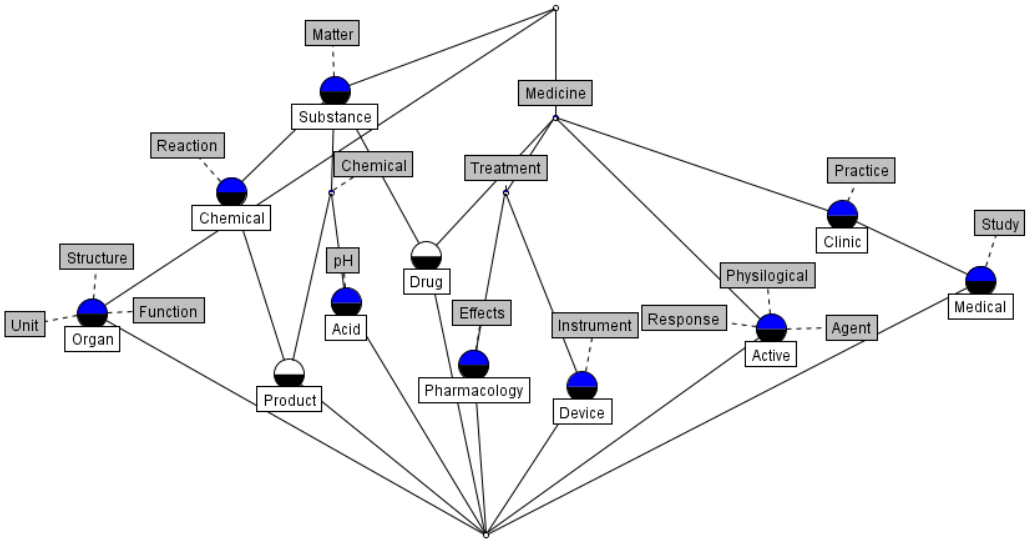


Figure 18: EMR Core Reference Ontology Primitive Concept Lattice for Existential Attributes.

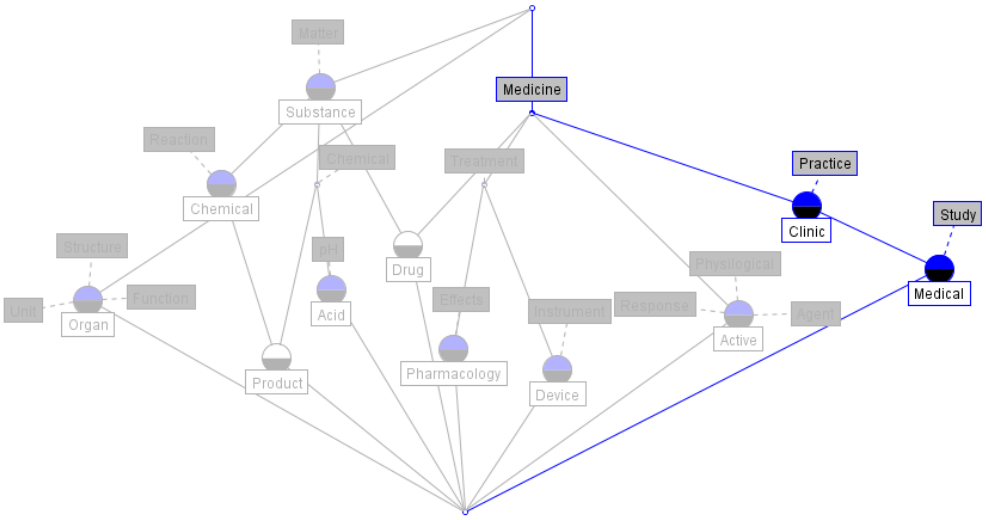


Figure 19: Lattice Path for Clinic.

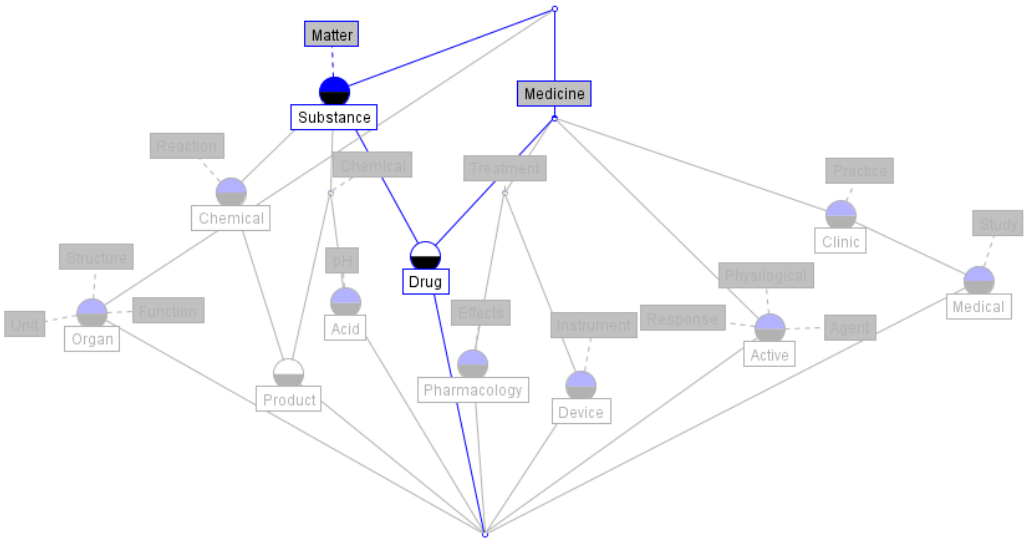


Figure 20: Lattice Path for Drug.

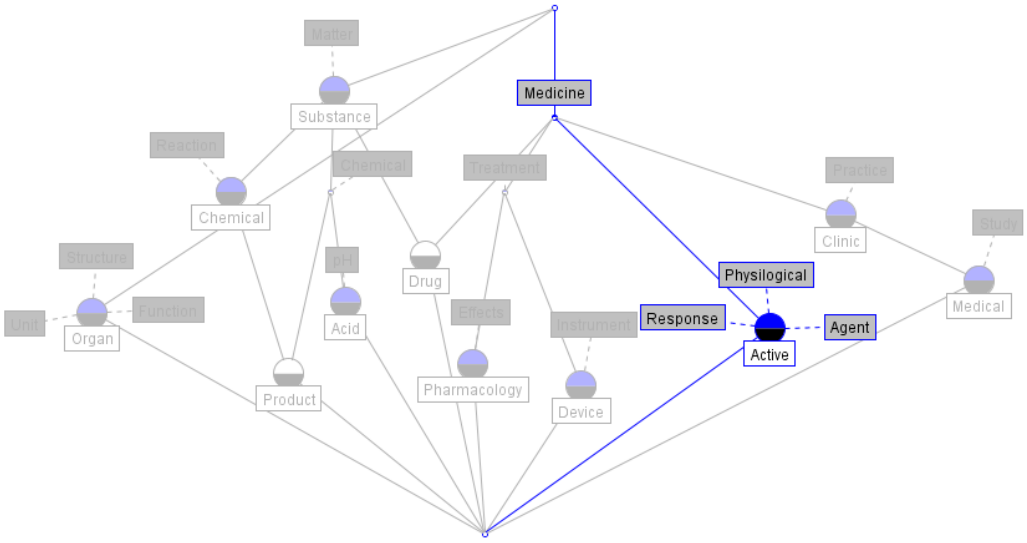


Figure 21: Lattice Path for Active.

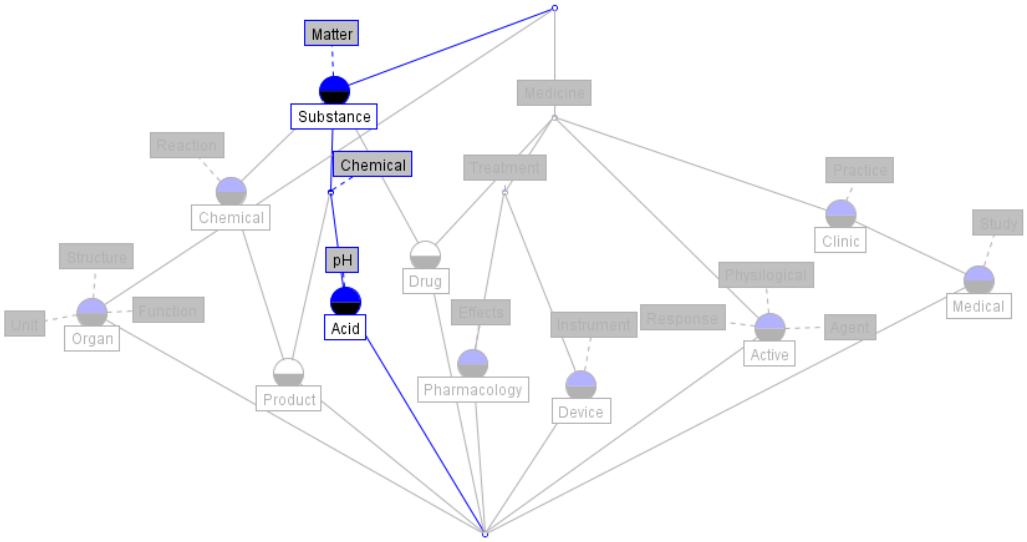


Figure 22: Lattice Path for Acid.

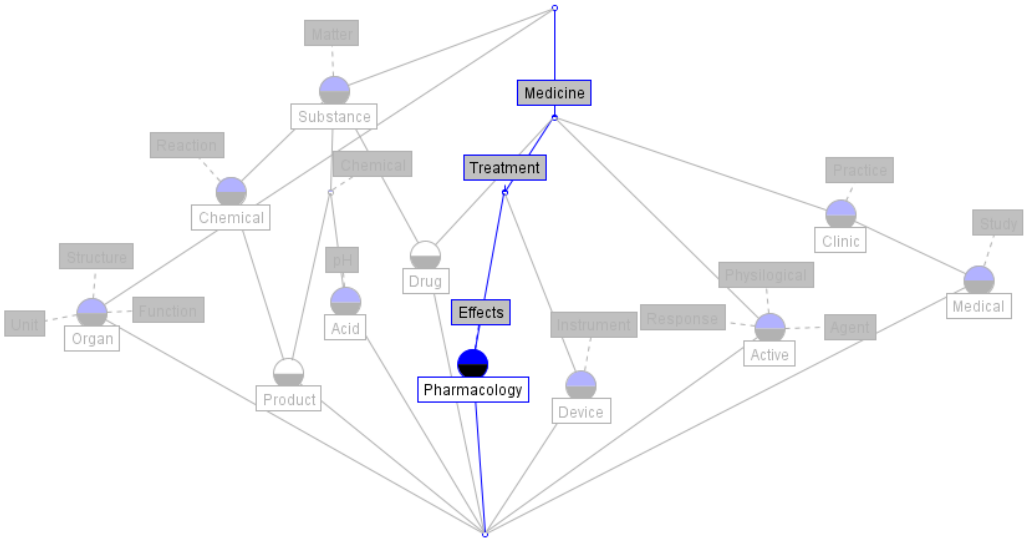


Figure 23: Lattice Path for Pharmacology.

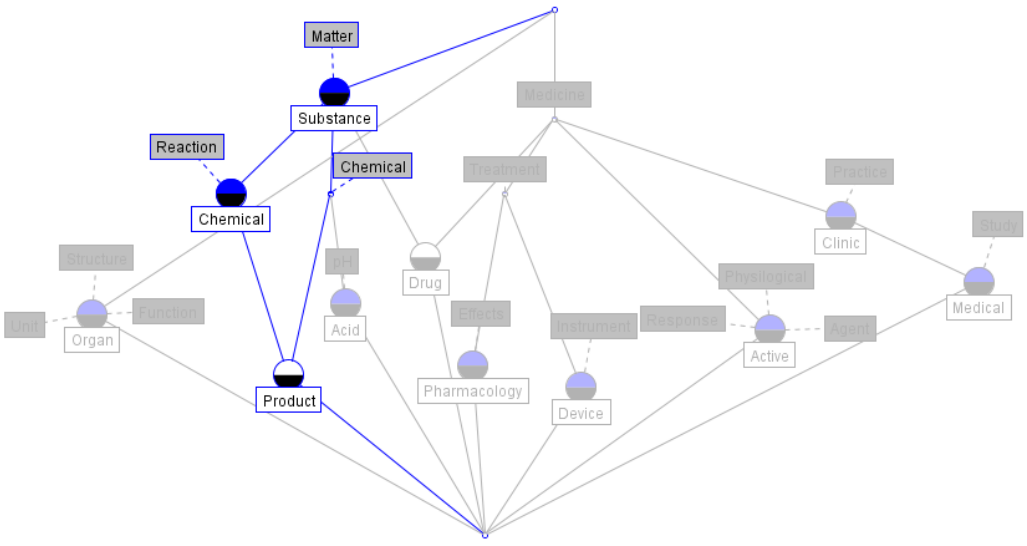


Figure 24: Lattice Path for Product.

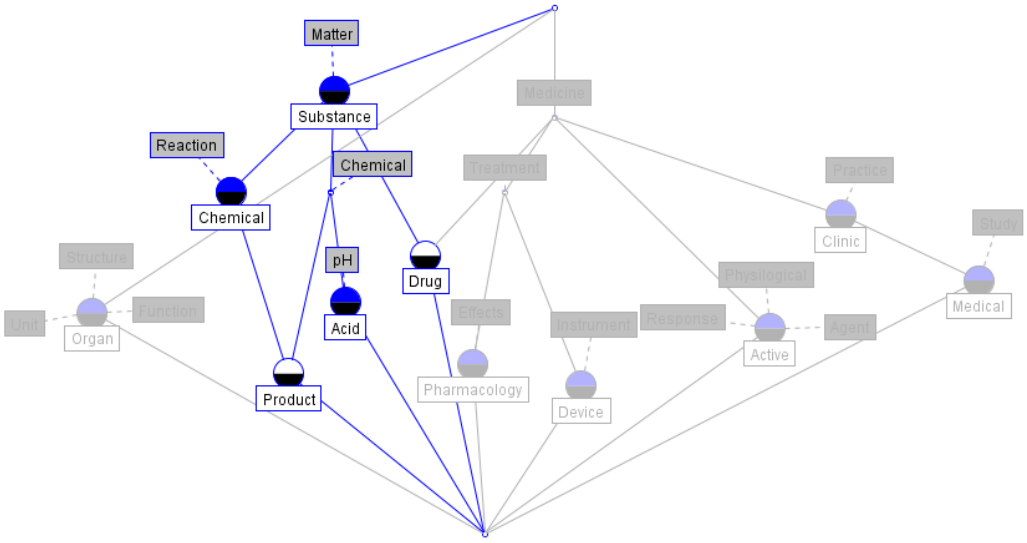


Figure 25: Lattice Path for Substance.

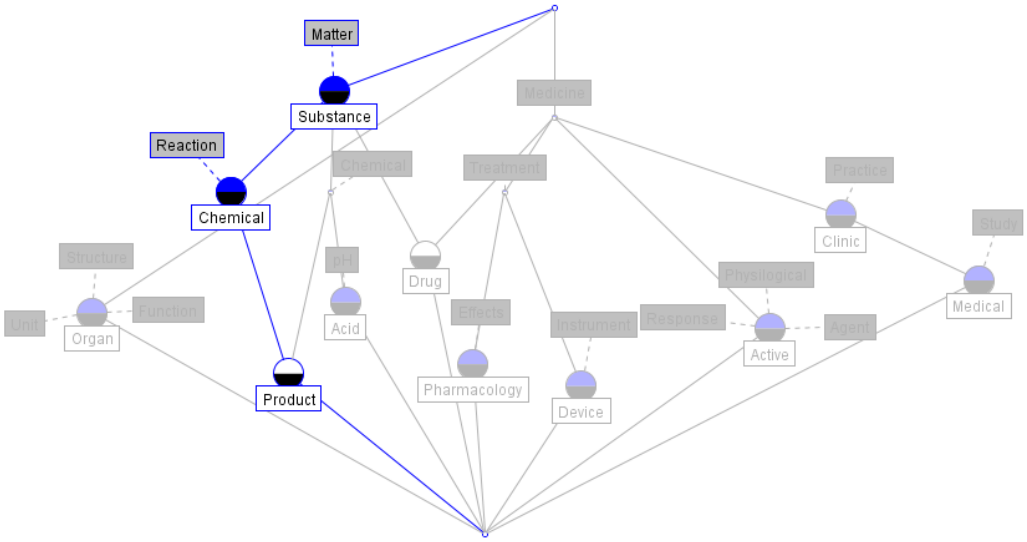


Figure 26: Lattice Path for Chemical.

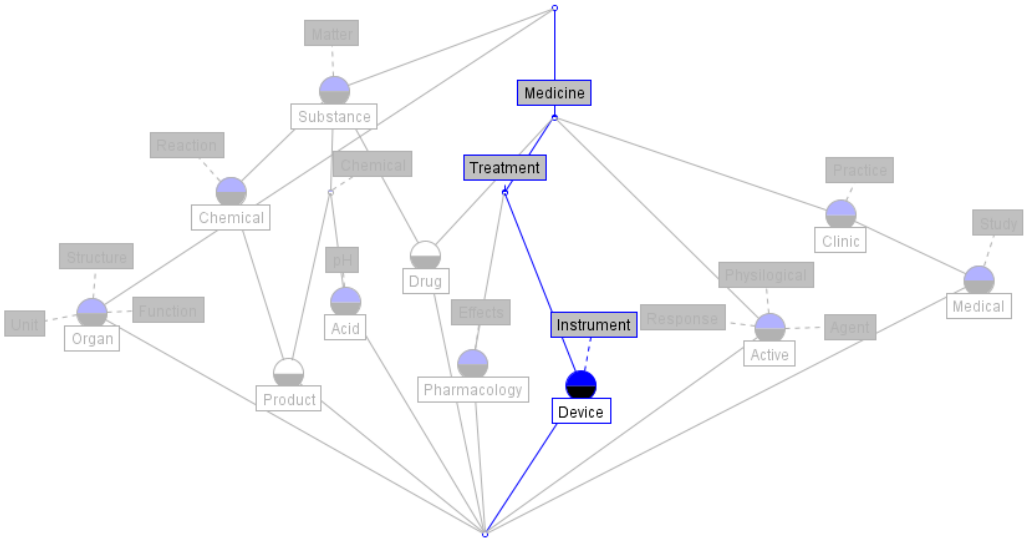


Figure 27: Lattice Path for Device.

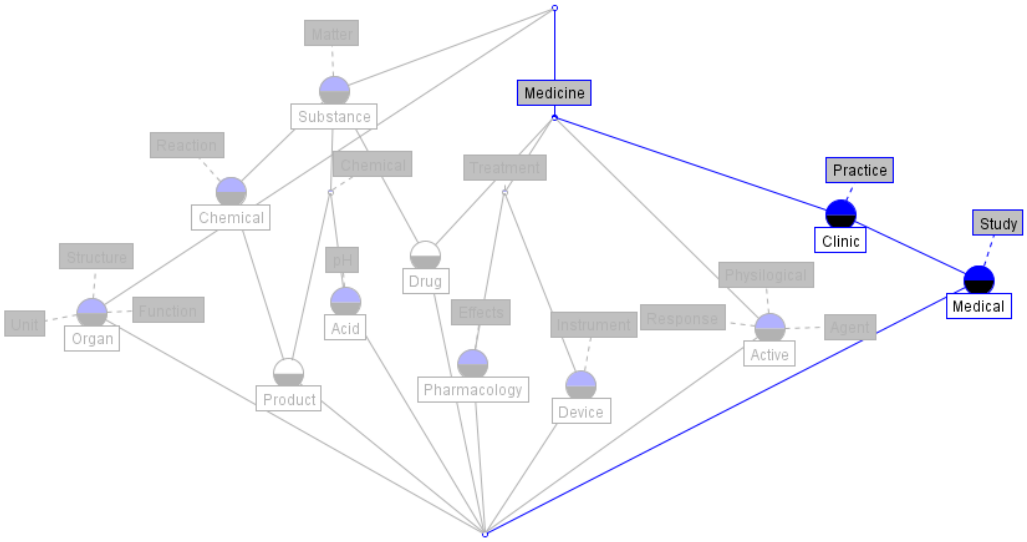


Figure 28: Lattice Path for Medical.

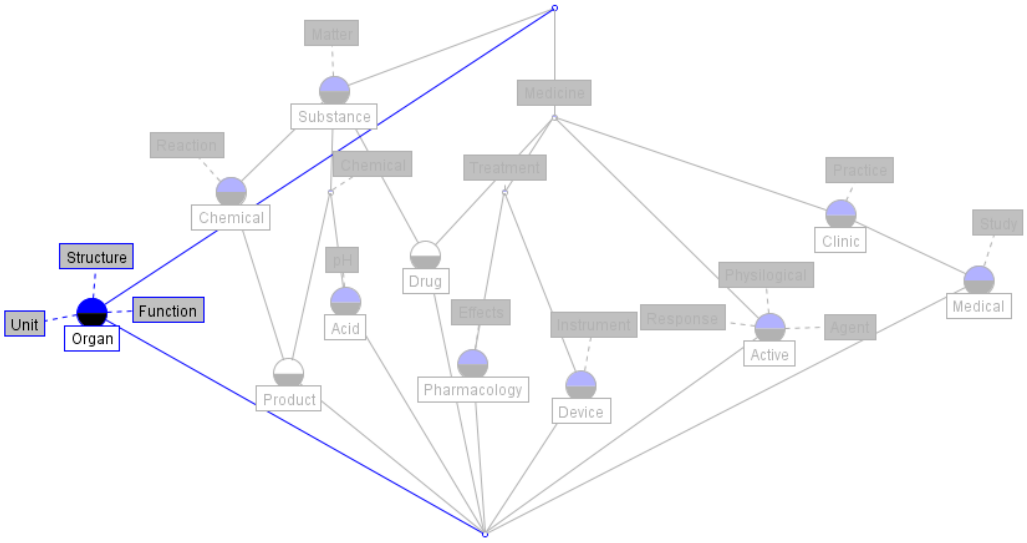


Figure 29: Lattice Path for Organ.

To summarize, the developed ontology is written in Web Ontology Language (OWL 2) which is a universal language in web semantics and thus meets semantic extendibility criteria. Therefore, semantic extendibility criteria are met in addition to modular extendibility.

CHAPTER 5

DISCUSSION

5.1 Overview of the Core Reference Ontology

This research proposed the development of a core reference primitive ontology for electronic medical records semantics interoperability. A core reference ontology provides the taxonomic and axiomatic scope structure of finer granularity than a foundational ontology for a core sub-discipline within a discipline's body of knowledge by integrating differing domain viewpoints. Likewise, the core reference ontology level provides the first opportunity to identify and incorporate cross domain latent composite concept keys necessary for the proper propagation of primitive concepts to domain and application level ontologies.

Currently, electronic medical records (EMR) cannot be exchanged among hospitals, clinics, laboratories, pharmacies, and insurance providers or be made available to patients. This research examined the interoperability problem amongst the medical terminologies and proposed an extraction method to contribute to the resolution of interoperability issues by identifying core reference primitive concepts and building operational axioms based on the correlation amongst them, which can be propagated to domain and application level ontologies in the future. This research identified primitive concepts for EMR core reference ontology (Figure 7) and their structure that specify a core reference hierarchical ontology and how those terms are axiomatically correlated with each other.

The development of this core reference ontology took a different approach than what Bodenreider and other researchers have done previously. Previous research used bottom-up manual approaches for identifying incomplete terms and missing terminology links among

medical terminologies. This core reference ontology applied a top-down, primitive concept identification approach to integrate the three dominant medical terminologies to establish interoperability at the core reference ontology level. According to Gomez-Perez (2004), the bottom up approach constructs a hierarchy using some clustering techniques; documents similar in content are associated with the same concept in the ontology, and then a concept for each cluster of documents relative to the same topic in the hierarchy is assigned using a bottom-up concept assignment mechanism. Conversely, for the top down approach, first the most general concepts of the ontology are selected. Then more specific concepts are added by classifying them in the already present structure. The top-down approach uses a downward knowledge acquisition process, which assures that the knowledge engineer considers all possible cases while avoiding redundant acquisition (Ganter and Stumme, 2003).

In this research, the top-down primitive identification approach ensured the identification of the manifest and latent dimensions within and across the three terminologies. Identified hierarchical latent categories are treatment, active, medication, and diagnosis. These terms were all buried within the terminologies and were ignored as a means to integrate. Identification of these top primitive latent categories ensures integration by establishing the latent connections amongst the terminologies. Subsequent constrained propagation of the core reference primitives through the domain and application level terms provides the potential to make these terminologies interoperable. Hence, the structural implications of identifying the manifest and latent ontological dimensions provides better potential to achieve interoperability than the current medical terminology development approach of trying to integrate existing terminologies from the bottom up.

5.2 Research Implications

Currently, the methods that are used to develop interoperable medical terminologies are:

- Structural methods which use the taxonomic structure of concept lattices.
- Semantic methods which use description logic-based concept definitions.
- Lexical methods which were based on term properties.
- Other evaluation methods included transforming the representation of a terminology to a different formalism and evaluating for compliance to that formalism, evaluating terminologies to specified principles, and mapping to other ontologies.

All methods (discussed in detail in Chapter 2) have limitations. For example: Application of structural-lexical methods to SNOMED CT extracted 6,801 non-lattice subgraphs that matched four primary lexical patterns. A random sample of 59 small subgraphs out of 2,046 amenable to visual inspection showed that all 59 contained errors as confirmed by terminology experts. The most frequent error was missing "is-a" relationships. (Bodenreider, 2018)

The core reference ontology method for EMR developed herein provides the basis that can contribute to overcoming these issues. This ontology used the top three terminologies and defined the primitives and semantic integration at the core reference level. The subsequent propagation of this core reference EMR primitive ontology to domain and application level EMR ontologies presents the potential to achieve maximum interoperability and to resolve non-lattice subgraphs, missing "is-a" relationships, incomplete mappings, and axiomatic relationships among them. This research has established the basis for hierarchical propagation of core reference primitive concepts to domain and application ontologies in patient electronic medical records.

5.3 Research Limitations

The primary limitation of this research was the inability to access SNOMED CT, RxNorm, and LOINC directly and having to use only their glossary textual definitions, normalized names and codes, and core definitions in the corpus. Since primary-foreign key relations were numerically encoded and not usable for this research purpose, some *a priori* specified axiomatic interrelationships among categories and terms might not have been fully discovered by this methodology.

Some other major risks and limitations of this research were set forth as problems needing resolution in section 3.3 and are summarized below.

1. The first limitation was the selection of the ontology development method that produces a hierarchy of primitive ontologies. Since ontology learning is a relatively new field, only two standards have been applied for evaluation of learned ontologies: (1) human expert evaluation and (2) comparing the learned ontology to a previously learned gold-standard ontology. Neither was available for this research.
2. The second limitation was related to the first. Specifically, what primitive breadth is necessary and sufficient to assure semantic translation among ontologies and terminologies with minimal human intervention?
3. The third limitation was identifying the limits of ontological semantics completeness such that incomplete or missing hierarchical branches can be identified.

To address these limitations, this research used a top-down strategy for building the patient electronic medical records core reference ontology to improve interoperability. The strategy integrates text mining and content analysis as the logical basis for identifying and extracting manifest and latent seed terms (primitive concepts- Figure 7) and hierarchical path

interrelationships within the SENSUS-like ontology Process 1 and 2 methods and verified the ontological properness by applying Welty and Guarino's (2001) criteria; normalization and modularity applying Rector's criteria; and completeness, closure, and cohesion using Formal Concept analysis (Figure 17 to Figure 28).

Another limitation that must be addressed in future research is that an ontology and its associated knowledge base are dynamic entities in that they must change with the addition of new knowledge. Biomedical terminologies which are the basis of this EMR core reference ontology, are dynamic with changes in term definitions, dropping terms, adding terms, and local extensions requiring constant monitoring and revisions maintain the static mappings up to date (Lau and Shakib, 2005). Without constant monitoring and automatic updating, static patient data may become non-interpretable and therefore non-interoperable. For example, standard vocabularies may retire or delete certain codes. If patient data is stored using the retired or deleted code, it will no longer be interoperable with other systems. Thus, automated monitoring and updating will be required to maintain interoperability of static data sets. There are tools and software available currently, but none have been tested in the core reference ontological EMR environment as this is a newly developed ontology. There are popular approaches like Protégé and CHAO that could be implemented to maintain the Ontology, *but are these approaches enough?* The answer to this question is out of the scope of this research but points to a path for future research.

Another point to note is that EMR interoperability is a major problem. Smith (1988) defines three criteria for a problem: (1) a gap between current and desired state, (2) difficulty in bridging that gap, and (3) someone must wish to bridge the gap. While it seems straightforward, in practice it is not. Solving a problem like EMR interoperability failure is complex not only

because there are no fully interoperable ontologies but also because of the necessity of having stakeholders involved in strategy implementation. Stakeholder involvement is necessary because (1) stakeholders have radically different world views and different frames of reference for understanding problems and (2) constraints and resources to solve the interoperability issues change over time; therefore, the interoperability problem may never have a complete solution.

CHAPTER 6

CONCLUSIONS

6.1 Primary Contributions of this Study

Electronic medical records were supposed to be beneficial for all. Electronic medical records were supposed to make medicine safer, bring higher-quality care, and empower patients all while also being economical. Electronic medical records were supposed to help researchers who would harness the big data to reveal the most effective treatments for disease and sharply reduce medical errors. Patients were supposed to get true portable health records which would enable them to share their medical histories with doctors and hospitals anywhere in the country. A recent study done by Kaiser Health News (KHN) and Fortune (Schulte and Fry, 2019), spoke with more than 100 physicians, patients, IT experts and administrators, health policy leaders, attorneys, top government officials and representatives at more than a half-dozen HER/EMR vendors, including the CEOs of two of the companies. The interviews reveal a tragic missed opportunity: rather than an electronic ecosystem of information, the nation's thousands of EMRs largely remain a sprawling, disconnected patchwork (Schulte and Fry, 2019). The systems cannot communicate with each other unless there is a standardized and seamless flow of information. Thus, having a fully interoperable system will have a major positive impact on healthcare. However, the lack of interoperability in healthcare systems and services has long been identified as one of the major challenges in healthcare, and prior work has been unable to mitigate it. As noted by Adler-Milstein (2017), after 30 years of monetary investment and research into the development of electronic medical record terminologies, the major technical issue still to be overcome is lack of semantics interoperability. This research reviewed prior approaches to

resolving medical terminology differences and identified the interoperability errors driving the interoperability problem. The primary contribution of this research is that it applied a top-down, primitive concept identification approach to EMR ontology development by integrating the three dominant medical terminologies to establish interoperability at the core reference ontology level, which is different than prior approaches.

This research is the first demonstration of the capability of a core reference, hierarchical primitive ontological architecture with integrated primitive concept ontology and concept attributes decomposition to integrate and resolve non-interoperable semantics among and extend coverage across existing clinical, drug, and hospital ontologies and terminologies. By using the methodology of this research and by propagating it to domain and application ontology levels, this developed and integrated core reference ontology has the potential to mitigate and improve the interoperability issues.

Other primary contributions of this study are summarized below.

- **Discipline:** Within ontology engineering, this research was the first demonstration of the ability of primitive concepts to integrate inconsistent terminologies.
- **Other Disciplines:** This research demonstrated the capability of hierarchical primitive ontological architectures to integrate and resolve non-interoperable semantics which can be extended directly to other disciplines to contribute to the resolution of non-interoperable semantics and knowledge.
- **Higher Education and Training:** EMR core reference ontology extends the theory and techniques for development of modular hierarchical primitive ontological architectures.
- **Broader Society:** This research contributed to interoperability and transferability of

electronic patient medical records; thus, it contributes to societal quality of health.

6.2 Widening the Scope

The scope of this research includes developing and designing a hierarchical core reference ontology in Electronic Medical Records. The developed ontology used the top three most used medical terminologies, named SNOMED CT, RxNorm, and LOINC, at the definition level. One extension of this research would be applying the primitive ontology methodology directly to these three databases as opposed to just applying it to definitions; this has the potential to provide a fully interoperable EMR system.

This scope may also be widened by extending the knowledge discovered in this research to all medical terminologies. The outcome of this EMR ontology is a human understandable theoretical basis for the ontology and a machine readable hierarchical taxonomic logic shareable across medical domains. This core reference primitive ontology can be propagated to domain and application level ontologies to improve medical record interoperability across all medical fields. The development of this EMR core reference ontology around which EMR machine intelligence knowledge can be encoded to form the basis for informed transition to artificially intelligent electronic medical records.

Another way the scope could be widened is by using the primitive concept ontology development methodology in non-medical ontologies where the same interoperable problems exist. The top-down, primitive concept identification approach has the potential to improve the underlying interoperability issues in non-medical fields as well.

6.3 Suggestions for Future Research

This core reference ontology is only the first version and needs to be updated frequently so that it does not become static. EMR is not a static field. Medical terminologies used in EMR are dynamic with changes in term definitions, dropping terms, adding terms, and local extensions requiring constant monitoring and revisions to maintain the static mappings up to date (Lau and Shakib, 2005). If patient data is stored using retired or deleted code it will no longer be interoperable with other systems. There are tools and software available currently, but none of them have been tested in the EMR environment as this core reference ontology applied a top-down, primitive concept identification approach to integrate the three dominant medical terminologies to establish interoperability at the core reference ontology level, which has not been used in EMR before. To keep the EMR primitive ontology interoperable, automated updating and maintenance methods will be needed. These methods must be developed and refined with future primitive ontology engineering research.

This core reference EMR primitive ontology must be propagated to domain and application level EMR ontologies to achieve maximum interoperability. Future research must specify the axiomatic ontology set theory necessary and sufficient for primitive propagation, identification of modular semantic subsets, and proper propagation of primitive and modular subsets with their interoperable axioms.

The primitive concepts identification process and methodologies can be extended to other applicable disciplines where interoperability problems exist. This research methodology could be used by ontology engineers in those disciplines even if they are not in the medical field.

BIBLIOGRAPHY

Adler-Milstein, J. (2017). Moving Past the EHR Interoperability Blame Game. *NEJM Catalyst*. https://catalyst.nejm.org/ehr-interoperability-blame-game/

AHIMA. (2018) AHIMA – Who We Are. http://www.ahima.org/about/aboutahima. Accessed November 14, 2018.

Australian Medicines Terminology (AMT). (2018) Australian Digital Health Agency.

Barbarito, F., Pinciroli, F., Mason, J., Marceglia, S., and Mazzola, L. (2012) Implementing standards for the interoperability among healthcare providers in the public regionalized Healthcare Information System of the Lombardy Region. *Journal of Biomedical Informatics*, 45(4), 736-745.

Bodenreider, O. (2010). Quality Assurance in Biomedical Terminologies and Ontologies. *A Report to the Board of Scientific Counselors, the Lister Hill National Center for Biomedical Communications, National Library of Medicine*, Bethesda, MD.

Bodenreider, O. (2016). Identifying Missing Hierarchical Relations in SNOMED CT from Logical Definitions Based on the Lexical Features of Concept Names. *Proceedings of the 6th International Conference on Biomedical Ontology (ICBO 2016) 2016* :(electronic proceedings: http://ceur-ws.org/Vol-1747/IT1601_ICBO2016.pdf).

Bodenreider, O. (2018). Evaluating the Quality and Interoperability of Biomedical Terminologies. *Technical Report to the LHNCBC Board of Scientific Counselors*, April 2018, https://lhncbc.nlm.nih.gov/publication/pub9754.

Cholan, R. and Bodenreider, O. (2018) Interoperability between Values Sets for Clinical Research and Healthcare: Mapping Value Sets between the Clinical Data Interchange

Standards Consortium (CDISC) and Meaningful Use. *AMIA 2018 Summits on Translational Science Proceedings*, 426-427.

Cimiano, P., Hotho, A., and Staab, S. (2005) Learning Concept Hierarchies form Text Corpora using Formal Concept Analysis. *Journal of Artificial Intelligence*, 24, 305-339.

Corcho, O., Fernandez-Lopez, M., and Gomez-Perez, A. (2003). Methodologies, Tools, and Languages for Building Ontologies. Where is Their Meeting Point? *Data & Knowledge Engineering*: *45*, 41-64.

Cotter, T., Mahumd, F., and Zahedi, Z. (2020). A Semantic Axiomatic Set Theory of Ontology Primitive Concepts. *ACM Transactions on Knowledge Discovery from Data* (in review).

Cristani, M. and Cuel, R. (2005). A Survey on Ontology Creating Methodologies. *International Journal on Semantic Web 7 Information Systems: 1*(2), 48-68.

Cui, L., Zhu, W., Tao, S., Case, J., Bodenreider, O., and Zhang, G. (2017). Mining non-lattice subgraphs for detecting hierarchical relations and concepts in SNOMED CT. *Journal of the American Medical Association*, 24(4), 788-798.

Dictionary of medicines and devices (dm+d). 2018. NHS Business Service Authority.

Esfeld, M. (2014). The primitive ontology of quantum physics: guidelines for an assessment of the proposals. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 47, 99-106.

Euzenat, J. and Shvaiko, P. (2013). Ontology *Matching*. New York: Springer.

Evans, R. (2016). Electronic Health Records: Then, Now, and the Future. *IMIA Yearbook of Medical Informatics*, 548-567.

Ganter, B. and Willie, R. (1999). *Formal Concept Analysis: Mathematical Foundations*. Heidelberg: Springer-Verlag Berlin

Ganter, B., & Stumme, G. (2003, July). Creation and merging of ontology top-levels. In International conference on Conceptual structures (pp. 131-145). Springer, Berlin, Heidelberg.

Ginsburg, P. (2005). Competition in Health Care: Its Evolution over the Past Decade. *Health Affairs*, 24(6), 1512-1522.

Gomez-Perez, A. (1996). Towards a Framework to Verify Knowledge Sharing Technology. Expert Systems with Applications, 519-529.

Gomez-Perez, A. (1999). Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases. Proceedings of the Banff Knowledge Acquisition for Knowledge-Based Systems Workshop KAW'99. Banff, Alberta, Canada: Ontology Engineering Group, 6.1.1 - 6.1.18.

Gomez-Perez, A. (2001). Evaluation of Ontologies. International Journal of Intelligent Systems, 16, 391-409.

Gómez-Pérez, A., Fernandez-Lopez, M., & Corcho, O. (2004). Ontological Engineering. London: Springer.

Gruber, T. R. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies: 43* (5-6), 907-928.

Gruber, T. R. (1993). A Translation Approach to Provide Ontology Specification. *Knowledge Acquisition*: *5*(2), 199-220.

Gur-Arie, M. (2013). The History of Healthcare Interoperability. *The Health IT Consultant*, 4(11), https://hitconsultant.net/2013/04/11/history-of-healthcare-interoperability/.

Hierholzer, W. (1992) Of Guides and Guidelines. *Infection Control and Hospital Epidemiology*, 13(6), 329-330.

Jones, D., Bench-Capon, T., and Visser, P. (1998). Methodologies for Ontology Development. *Proceedings of IT & KNOWS Conference of the 15th IFIP World Computer Conference,* Budapest: Chapman-Hall, 62-75.

Institute of Medicine. (1997). The *Computer-Based Patient Record: An Essential Technology for Health Car*e. Dick, R., Steen, E. and Detmer, D. eds., Washington, D.C.: National Academy Press.

Lewis, P. (2016) *Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics*. New York: Oxford University Press.

Lau, L. and Shakib, S. Towards Data Interoperability: Practical Issues in Terminology Implementation and Mapping. *Proceedings of the 2005 Health Informatics Conference*, Melbourne, Australia, July 31 to August 2, 2005.

Mahmud, F. (2018) *Human-Intelligence and Machine-Intelligence Decision Governance Formal Ontology*. Dissertation, Old Dominion University.

Manos, D. (2014) Obama: EHRs for Americans by 2014. *Healthcare IT News*, https://www.healthcareitnews.com/news/obama-ehrs-americans-2014.

Mascardi, V. and Paolo, R. (2007). A Comparison of Upper Ontologies. *Proceedings of the Workshop on Objects and Agents (WOA),* Genova, Italy, 55-64.

Obrst, L. (2010). Ontological Architectures. In Poli, R., Healy, M., and Kameas, A., eds., Theory and Applications of Ontology: Computer Applications, New York: Springer, 27-66.

Ochs, C., Geller, J., Perl, Y., Chen, Y., Xu, J., Min, H., Case, J., and Wei, Z. (2015) Scalable quality assurance for large SNOMED CT hierarchies using subject-based sub

taxonomies. *Journal of the American Medical Information Association*, 22(3): 507-518 [PubMed: 25336594].

Ochs, C., Perl, Y., Geller, J., Halper, M., Gu, H., Chen, T, and Elhanan, G. (2015) Scalability of abstraction- abstraction-network-based quality assurance to large SNOMED hierarchies. *AMIA Annual Symposium Proceedings*, 1071-1080 [PubMed: 24551393].

Rector, A. (2003) Modularization of Domain Ontologies Implemented in Description Logics and related formalisms including OWL. *K-CAP 2003 Proceedings of the 2nd International Conference on Knowledge Capture*, 121-128.

Regan, J. (1991) Computerized Information Exchange in Health Care. *Medical Journal of Australia*, Jan: 154(2): 140-144.

Roussey, C., Pinet, F., Ah Kang, M., and Corcho, O. (2011) Chapter 2: An introduction to ontologies and ontology engineering. In Ontologies in Urban Development Projects, Falquet, G., Metral, C., Teller, J., and Tweed, C., ed's, Springer London, 2011. 9-38.

Schrödinger, E. (1926). An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review*, 28(6), 1049-1070.

Schulte, F., & Fry, E. (2019, April 01). Death By 1,000 Clicks: Where Electronic Health Records Went Wrong. Retrieved from https://khn.org/news/death-by-a-thousand-clicks/

Seitner, J., Bizer, C., Eckert, K., Faralli, S., Meusel, R., Paulheim, H., and Aolo Ponzetto, S. (2016). A Large Database of Hypernymy Relations Extracted from the Web. *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*. Portorož, Slovenia.

Smith, G. (1988). Towards a heuristic theory of problem structuring. Management Science, 34(12), 1489–1506.

Stuckenschmidt, H., Parent, C., and Spaccapietra, S., (eds). (2009) Modular Ontologies: Concepts, Theories, and Techniques for Knowledge Modularization. Berlin: Springer-Verlag.

Tierney, W., Miller, M., Overhage, J., and McDonald, C. (1993) Physician inpatient order writing on Microcomputer Workstations: Effects on Resource Utilization. *Journal of the American Medical Association*, 269(3), 379-383.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. The knowledge engineering review, 11(2), 93-136.

Vieira, R. (2007) Professional *SQL Server 2005 Programming*. New York: Wiley.

Wachter, R. (2017) *The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age*. Columbus, OH: McGraw-Hill Education.

Welty, C. and Guarino, N. (2001). Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering*, 39(1), 51-74.

Wiedemann, L. (2010). Fundamentals for Building a Master Patient Index/Enterprise Master Patient Index (Updated September 2010). *Journal of AHIMA*. https://engage.ahima.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=ca53ccdc-60bb-4320-a014-8652894a588e.

Zheng, L., Yumak, H., Chen, L., Ochs, C., Geller, J., Kapusnik-Uner, J., and Perl, Y. (2017) Quality assurance of chemical ingredient classification for the National Drug File - Reference Terminology. *Journal of Biomedical Informatics*, 73, 30-42 [PubMed: 28723580].

APPENDIX A

DETAILED R CODE

```
> #Package installation
> install.packages("tm")
> library(tm)
> install.packages("SnowballC")
> library(SnowballC)
> install.packages("ggplot2")
> library(ggplot2)
> install.packages("cluster")
> library(cluster)
> install.packages("fpc")
> library(fpc)

> #Create corpus
> cname <- file.path("C:", "Corpus_LRS_txt")
> cname
> docs <- VCorpus(DirSource(cname))
> docs <- tm_map(docs, content_transformer(tolower))

> #Strip digits/numbers
> docs <- tm_map(docs, removeNumbers)
```

> **#Remove punctuation**

> docs <- tm_map(docs, removePunctuation)

> **#Remove stopwords using the standard list in tm**

> docs <- tm_map(docs, removeWords, stopwords("english"))

> **#Stem document**

> docs <- tm_map(docs, stemDocument)

> **#Document-term matrix**

> dtm <- DocumentTermMatrix(docs)

> tdm <- TermDocumentMatrix(docs)

> dtm

> freq <- colSums(as.matrix(dtm))

> ord <- order(freq)

> freq <- sort(colSums(as.matrix(dtm)), decreasing=TRUE)

> head(freq, 25)

> **#Remove custom English words**

> docs <- tm_map(docs, removeWords, "rxnorm")

> docs <- tm_map(docs, removeWords, "mthspl")

> docs <- tm_map(docs, removeWords, "nddf")

> docs <- tm_map(docs, removeWords, "mgml")

```
> docs <- tm_map(docs, removeWords, "snomedctus")
> docs <- tm_map(docs, removeWords, "find")
> docs <- tm_map(docs, removeWords, "mmsl")
> docs <- tm_map(docs, removeWords, "hpx")
> docs <- tm_map(docs, removeWords, "first")
> docs <- tm_map(docs, removeWords, "however")
> docs <- tm_map(docs, removeWords, "eng")
> docs <- tm_map(docs, removeWords, "random")
> docs <- tm_map(docs, removeWords, "use")
> docs <- tm_map(docs, removeWords, "add")

> #Document-term matrix
> dtm <- DocumentTermMatrix(docs)
> tdm <- TermDocumentMatrix(docs)
> dtm
> freq <- colSums(as.matrix(dtm))
> ord <- order(freq)
> freq <- sort(colSums(as.matrix(dtm)), decreasing=TRUE)
> head(freq, 25)
> docs <- tm_map(docs, removeWords, "mmx")
> dtm <- DocumentTermMatrix(docs)
> tdm <- TermDocumentMatrix(docs)
> dtm
```

```
> freq <- colSums(as.matrix(dtm))
> ord <- order(freq)
> freq <- sort(colSums(as.matrix(dtm)), decreasing=TRUE)
> head(freq, 25)
> wf <- data.frame(word=names(freq), freq=freq)
> head(wf)

> #Cluster diagram
#
> p <- ggplot(subset(wf, freq>49000), aes(x = reorder(word, -freq), y = freq)) +
        geom_bar(stat = "identity") +
       theme(axis.text.x=element_text(angle=45, hjust=1))
> p
#
> dtmss5 <- removeSparseTerms(dtm, 0.5) *** Change the sparsity value for 0.5, 010, 0.15,
0.20, 0.25, 0.30, 0.35, and 0.45
> d5 <- dist(t(dtmss5), method="euclidian")
> fit <- hclust(d=d5, method="complete")
> plot(fit, hang=1, main = "Cluster Dendogram - 5% Sparsity") ***Change the naming
convention based on sparsity value

> #CLUSPLOT
> groups <- cutree(fit, k = 7)
```

> rect.hclust(fit, k = 7, border = "red") *** Change the value for means (K) to 3, 4, 5, 6, 7, 8, and 9

#

> d5_7 <- dist(t(dtmss5), method="euclidian")

> kfit <- kmeans(d5_7,7) *** Change the value for 3, 4, 5, 6, 7, 8, and 9

> clusplot(as.matrix(d5_7), kfit$cluster, color=T, shade=T, labels=2, lines=0, main = "CLUSPLOT - 5% Sparsity, k = 7 means") *** Change the naming convention based on the value of K and sparsity

> **#Association of terms (***** Change the frequencies from 0.99 to 0.80)

#

> findAssocs(dtm, c("eng"), corlimit = 0.999999999)

> findAssocs(dtm, c("oral"), corlimit = 0.99)

> findAssocs(dtm, c("drug"), corlimit = 0.99)

> findAssocs(dtm, c("clinic"), corlimit = 0.99)

> findAssocs(dtm, c("activ"), corlimit = 0.99)

> findAssocs(dtm, c("tablet"), corlimit = 0.99)

> findAssocs(dtm, c("eng"), corlimit = 0.80)

> findAssocs(dtm, c("drug"), corlimit = 0.80)

> findAssocs(dtm, c("clinic"), corlimit = 0.80)

> findAssocs(dtm, c("activ"), corlimit = 0.80)

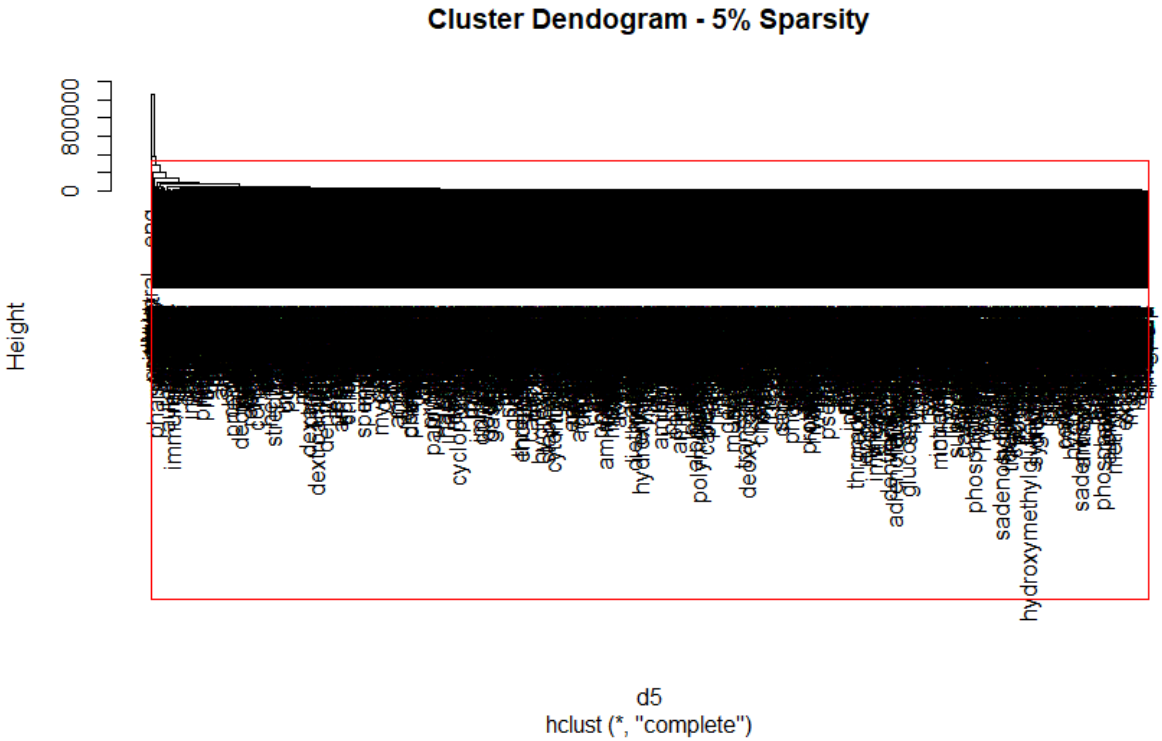
> findAssocs(dtm, c("product"), corlimit = 0.80)

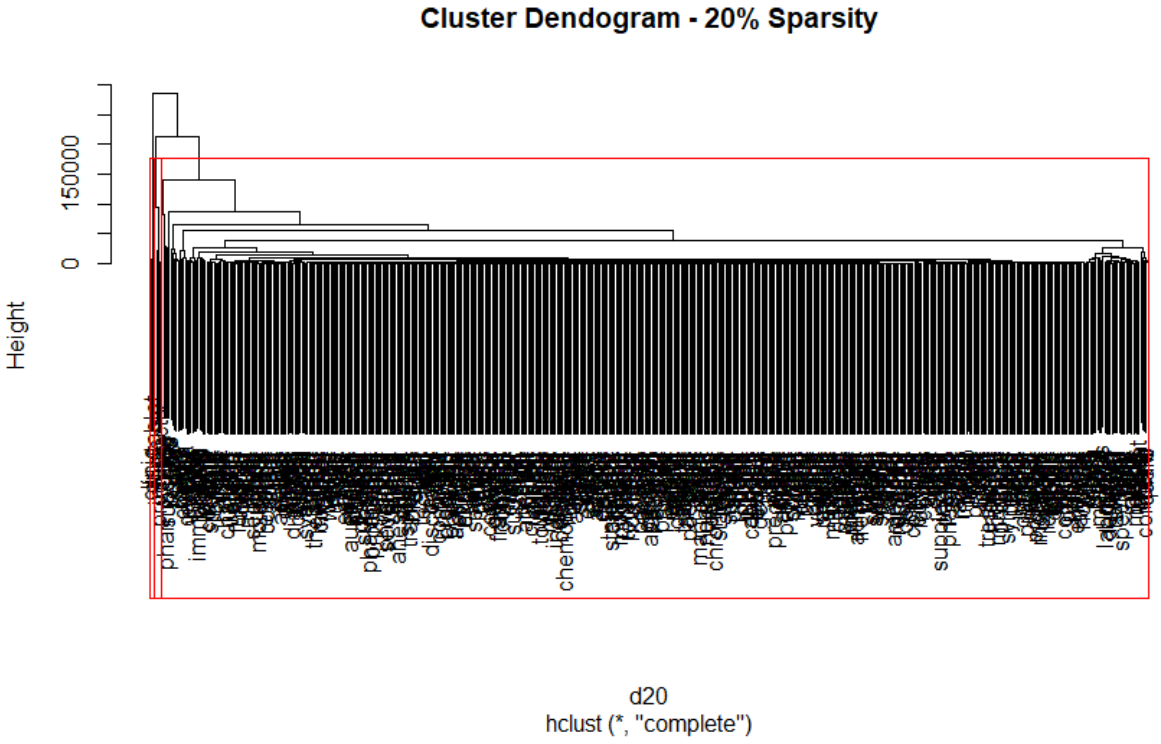
> findAssocs(dtm, c("pharmacolog"), corlimit = 0.80)

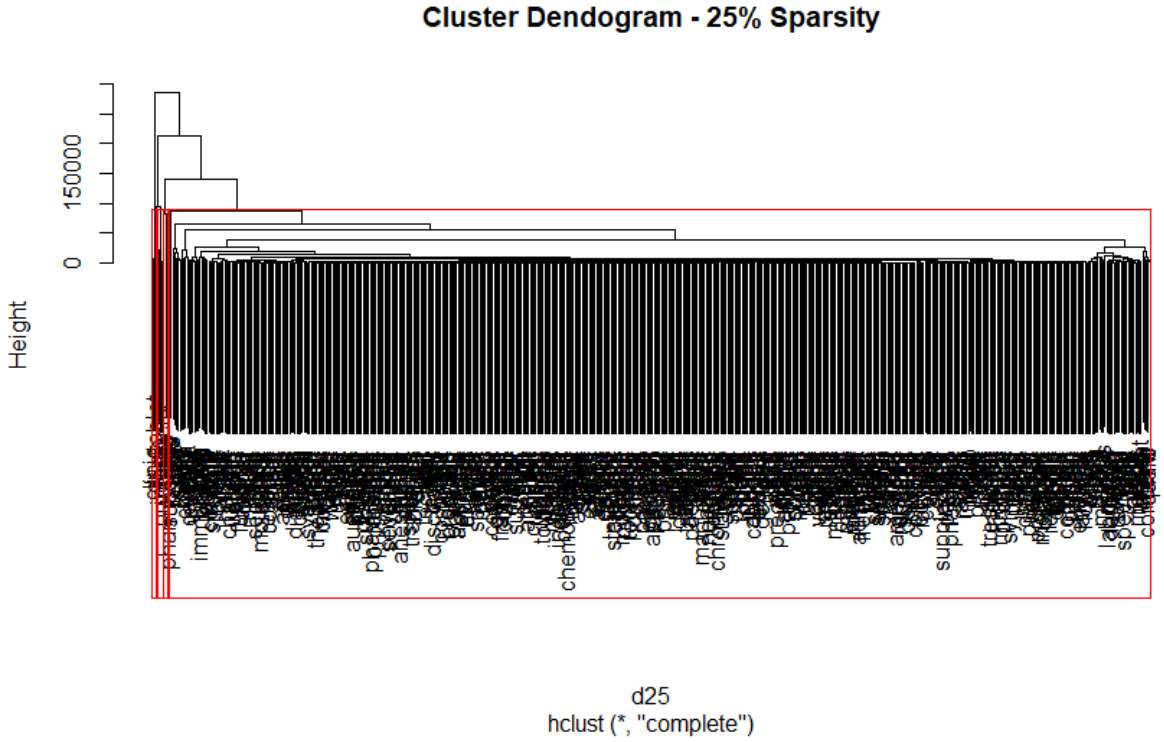
```
> findAssocs(dtm, c("substanc"), corlimit = 0.80)

> findAssocs(dtm, c("acid"), corlimit = 0.80)

> findAssocs(dtm, c("devic"), corlimit = 0.80)

> findAssocs(dtm, c("medic"), corlimit = 0.80)

> findAssocs(dtm, c("chemic"), corlimit = 0.80)

> findAssocs(dtm, c("organ"), corlimit = 0.80)

> findAssocs(dtm, c("cell"), corlimit = 0.80)

> findAssocs(dtm, c("eng"), corlimit = 0.80)
```

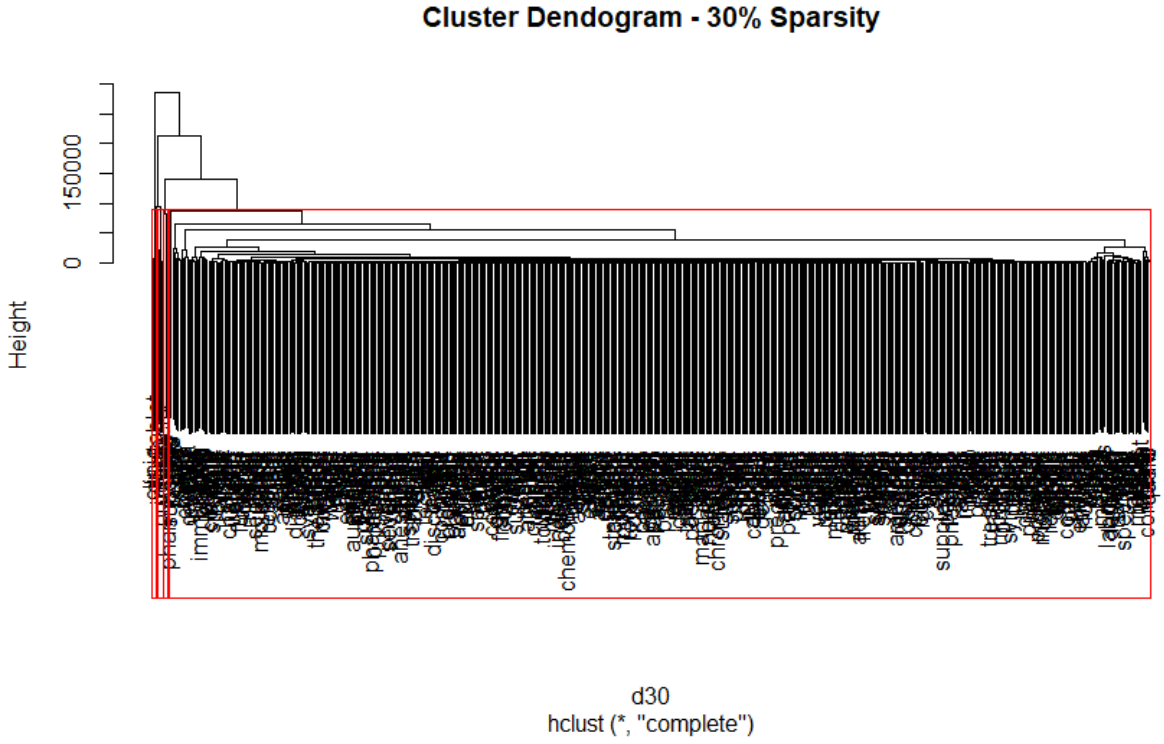
APPENDIX B

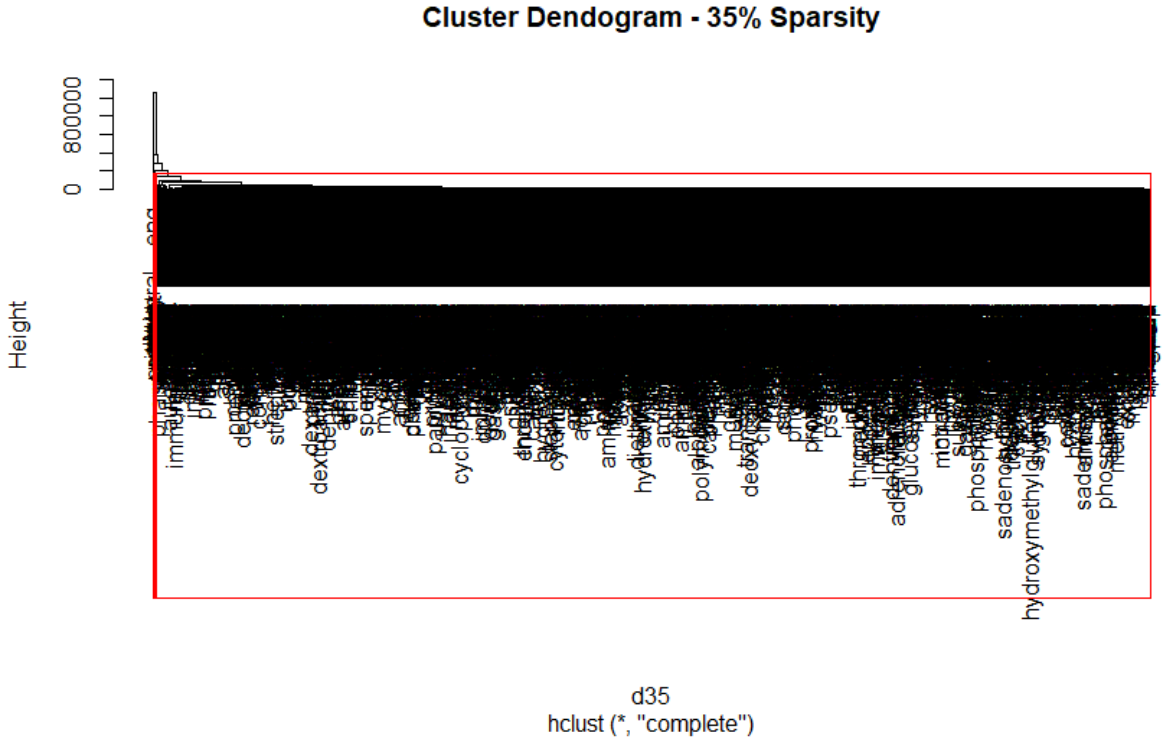
ADDITIONAL DENDOGRAM FIGURES

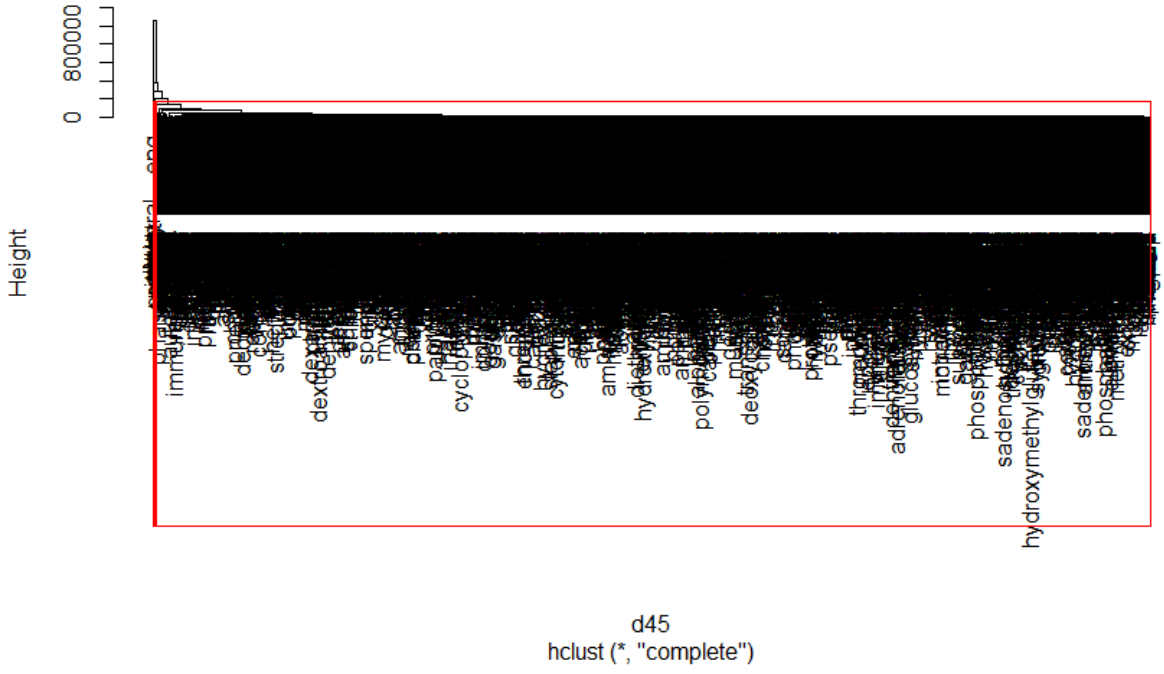






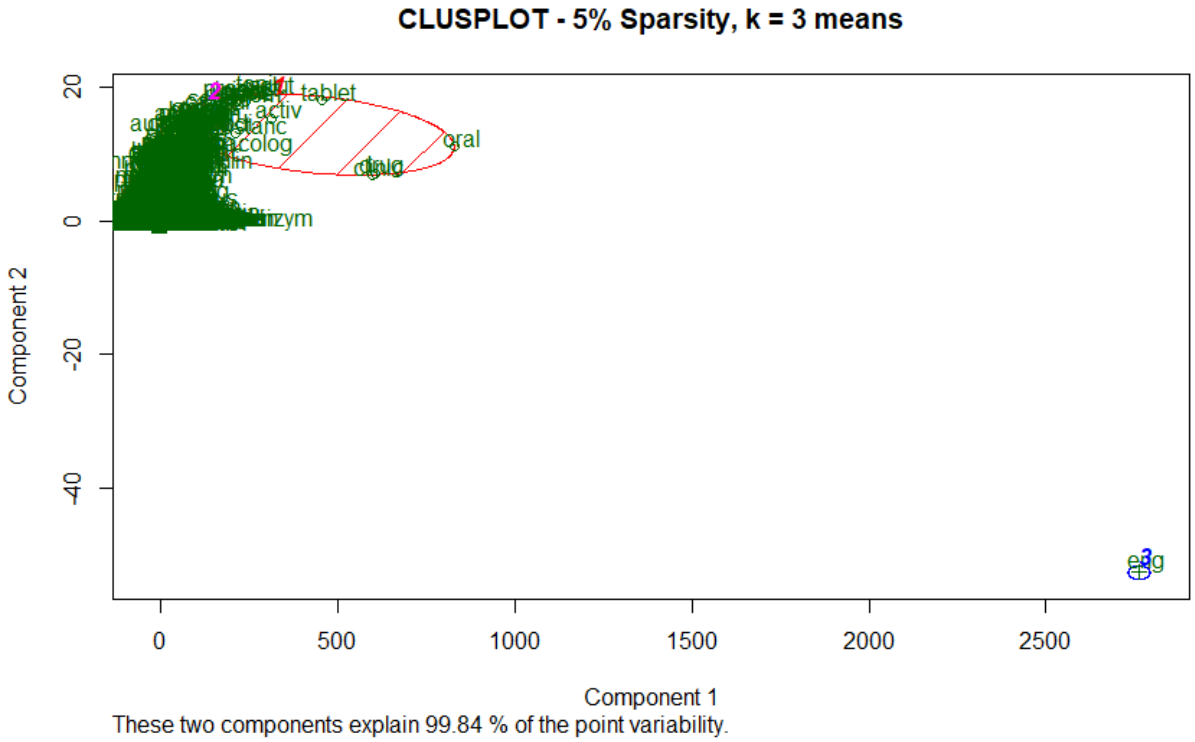


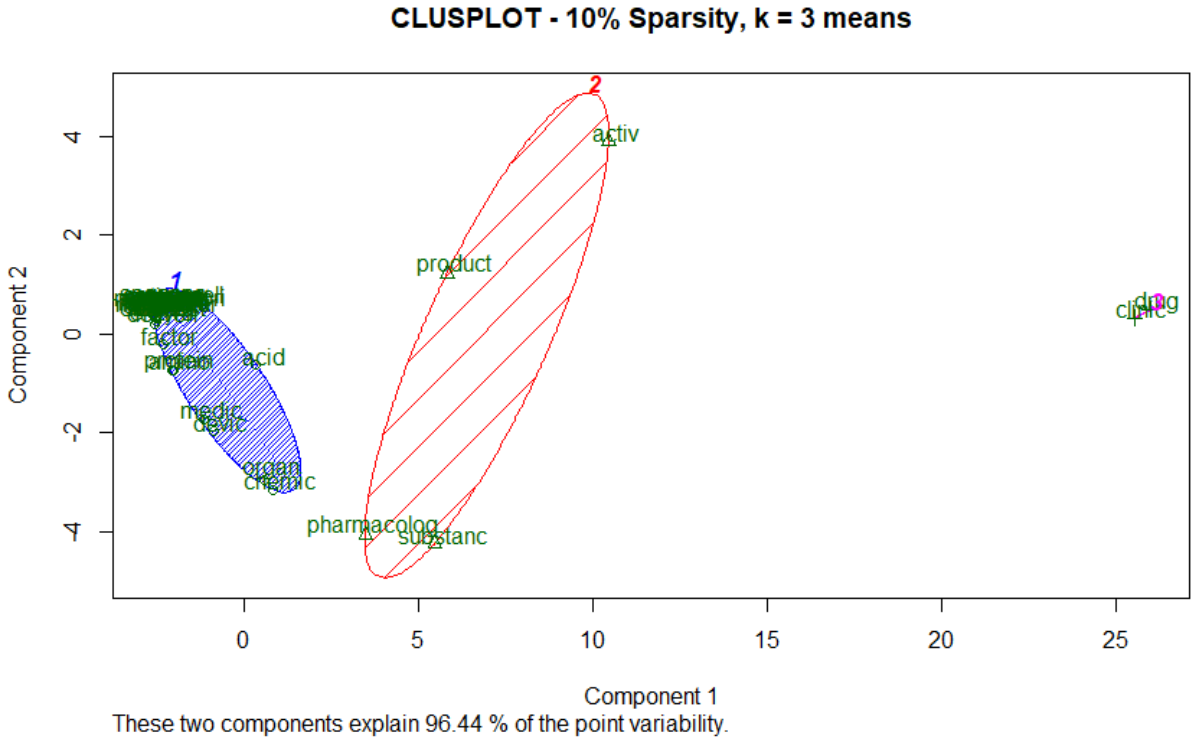


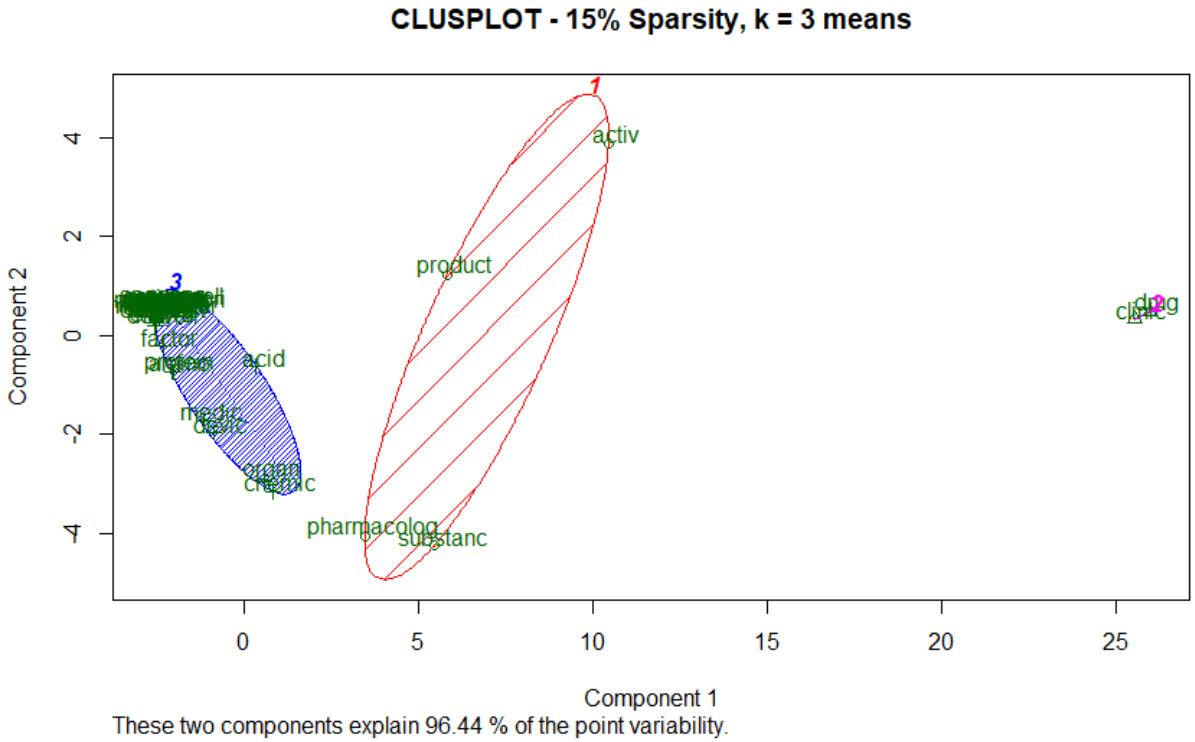


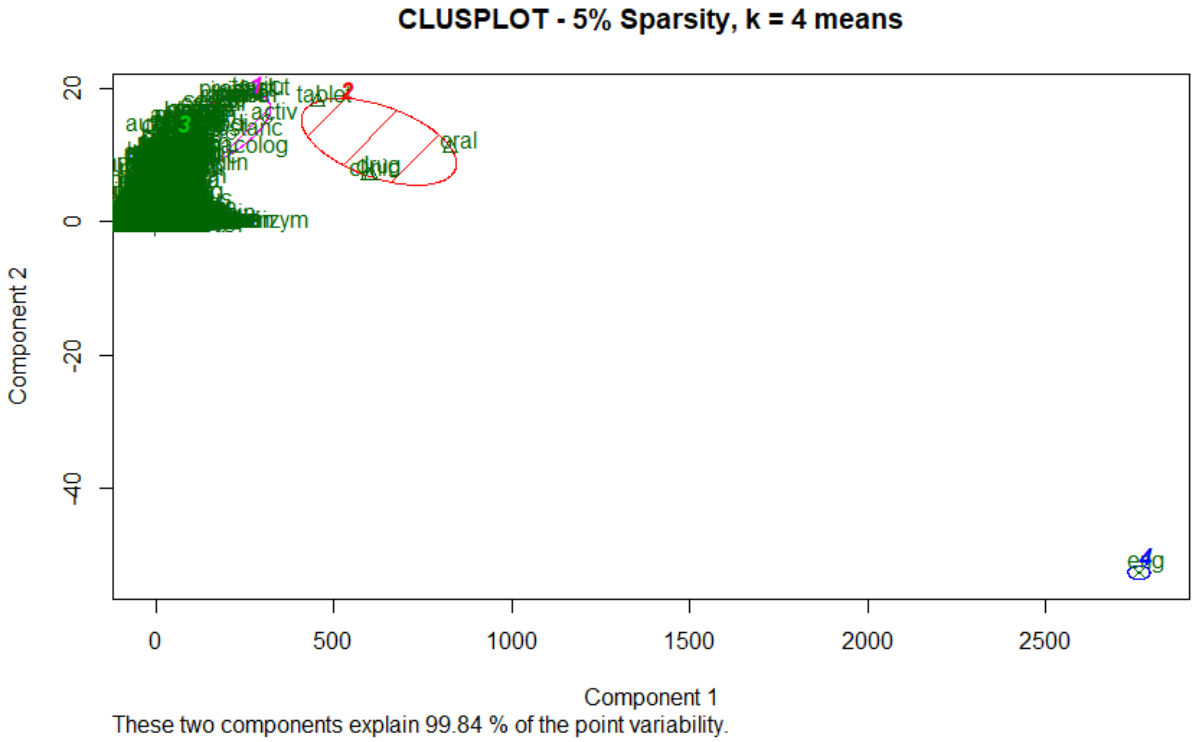
APPENDIX C

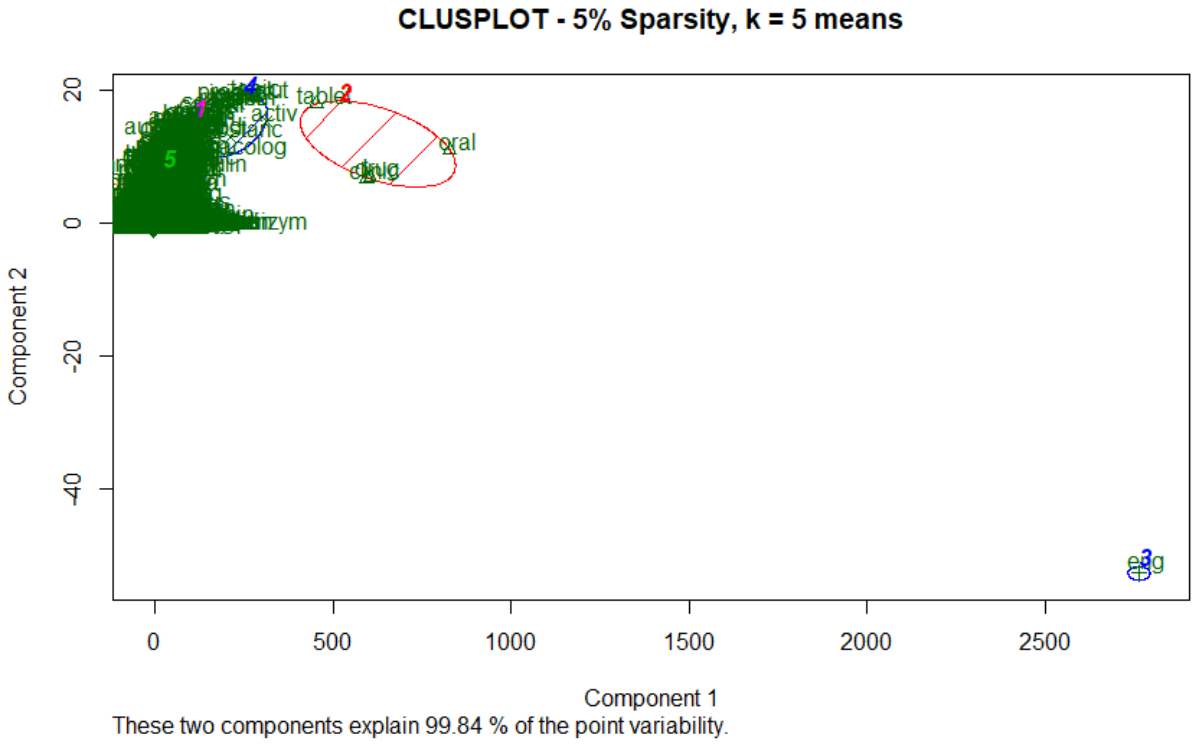
ADDITIONAL CLUSPLOTS

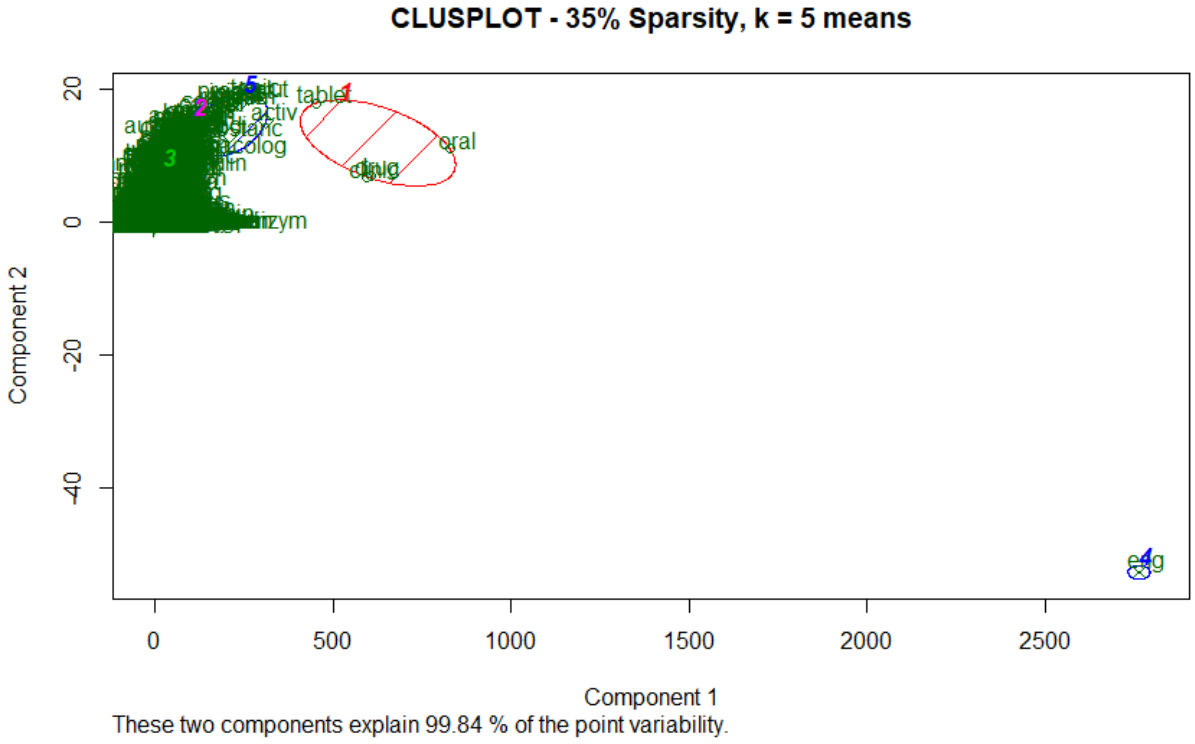


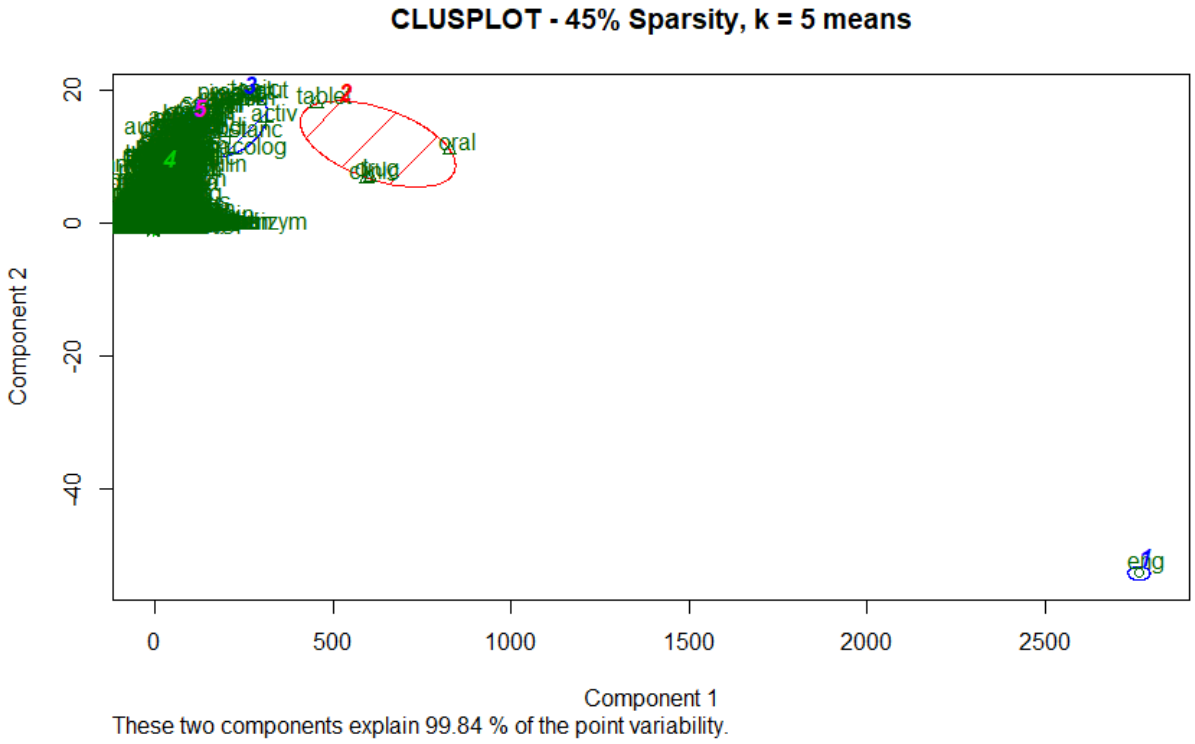


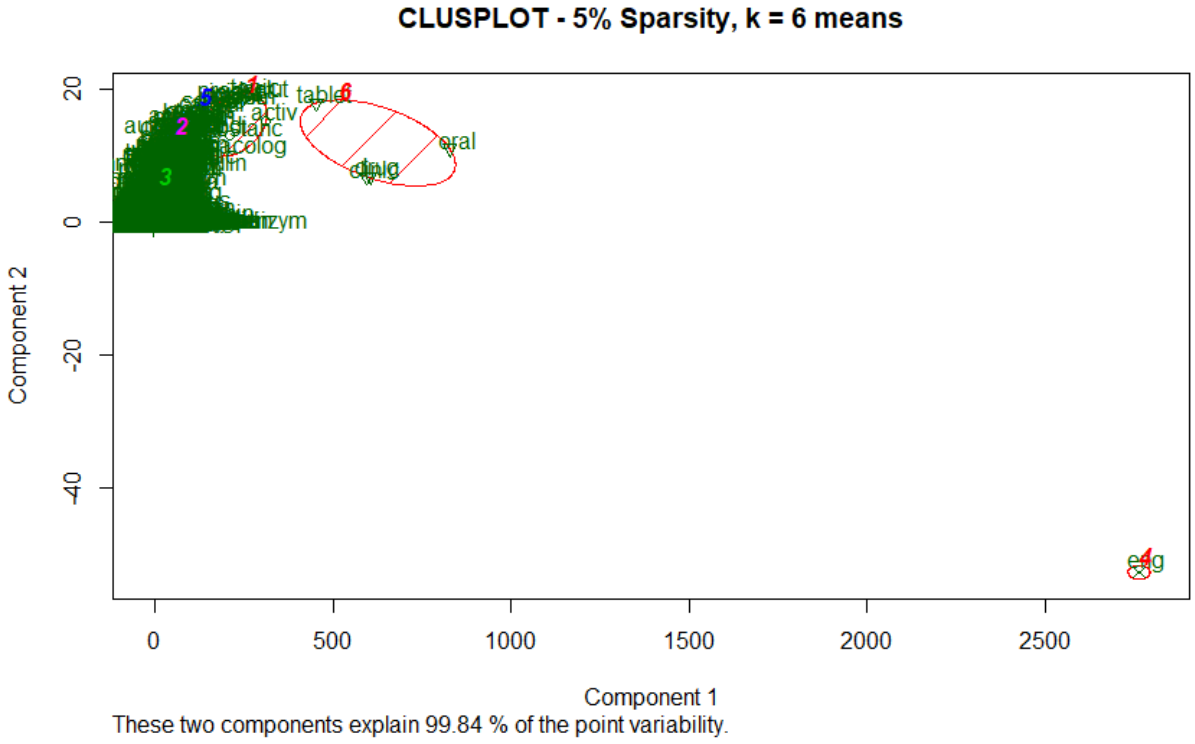


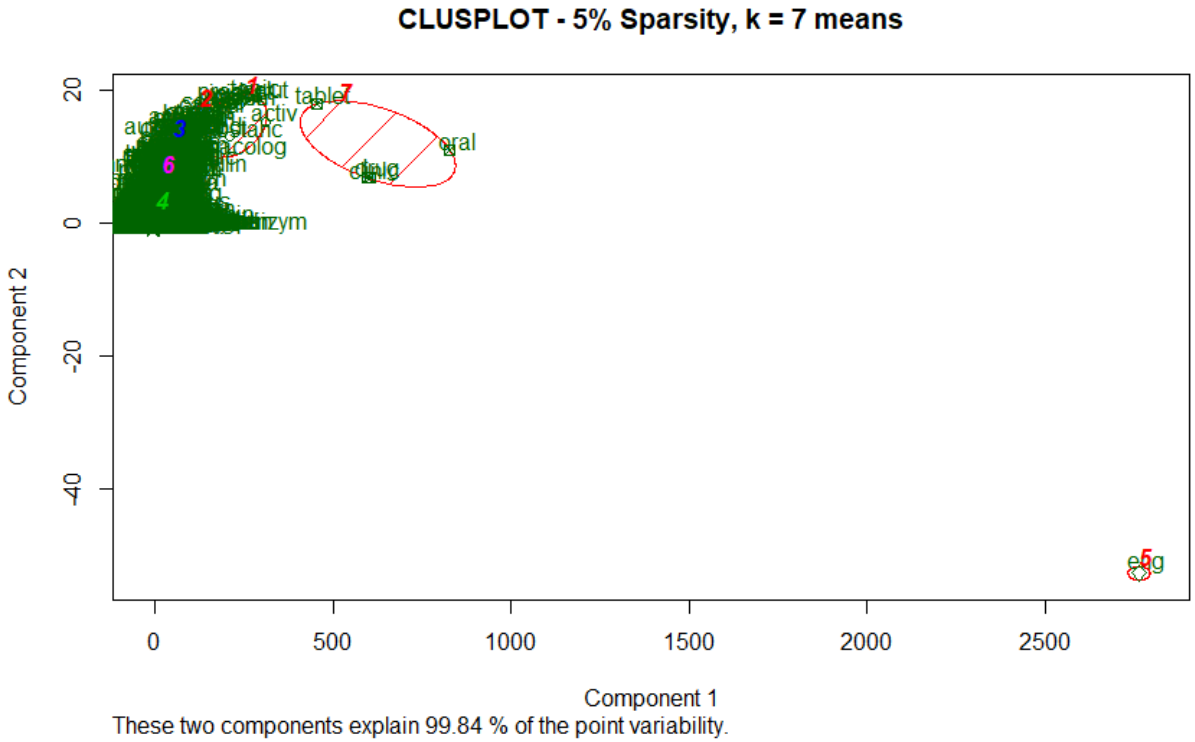


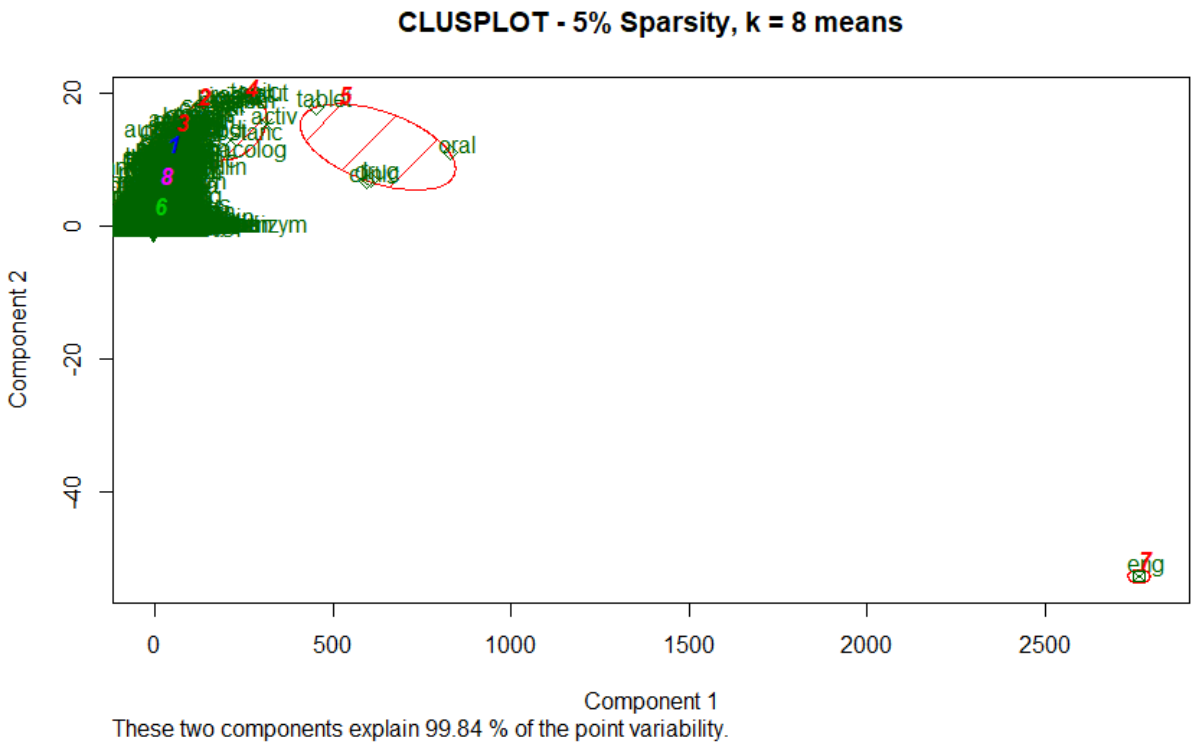




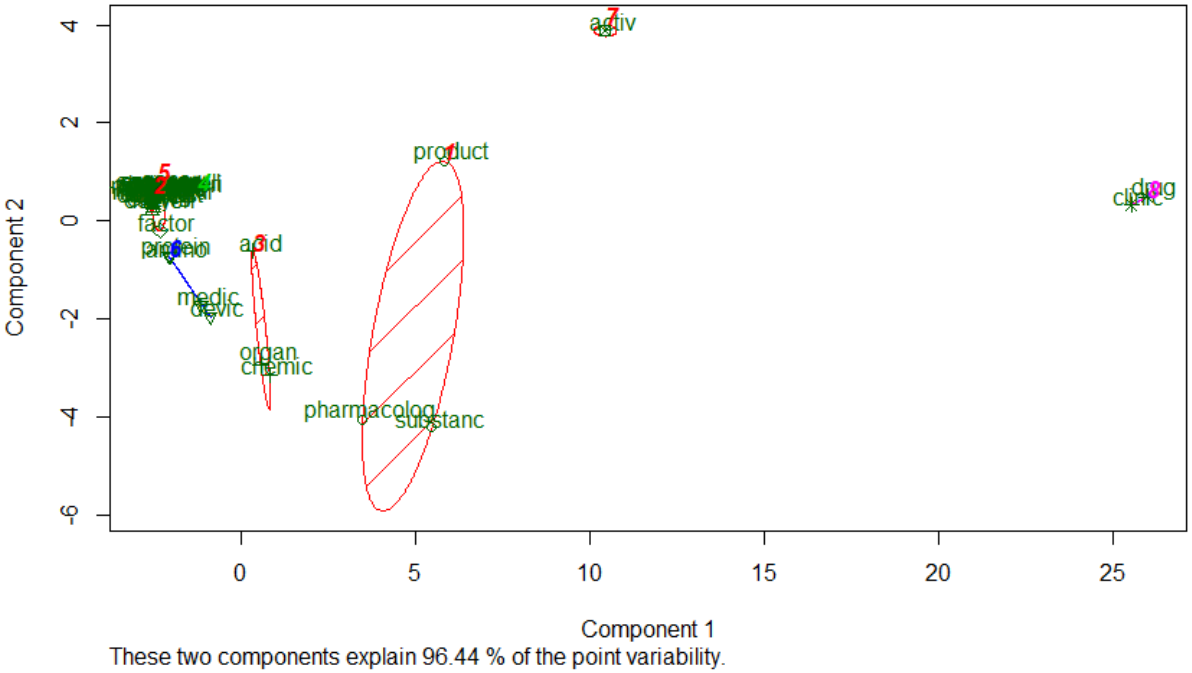






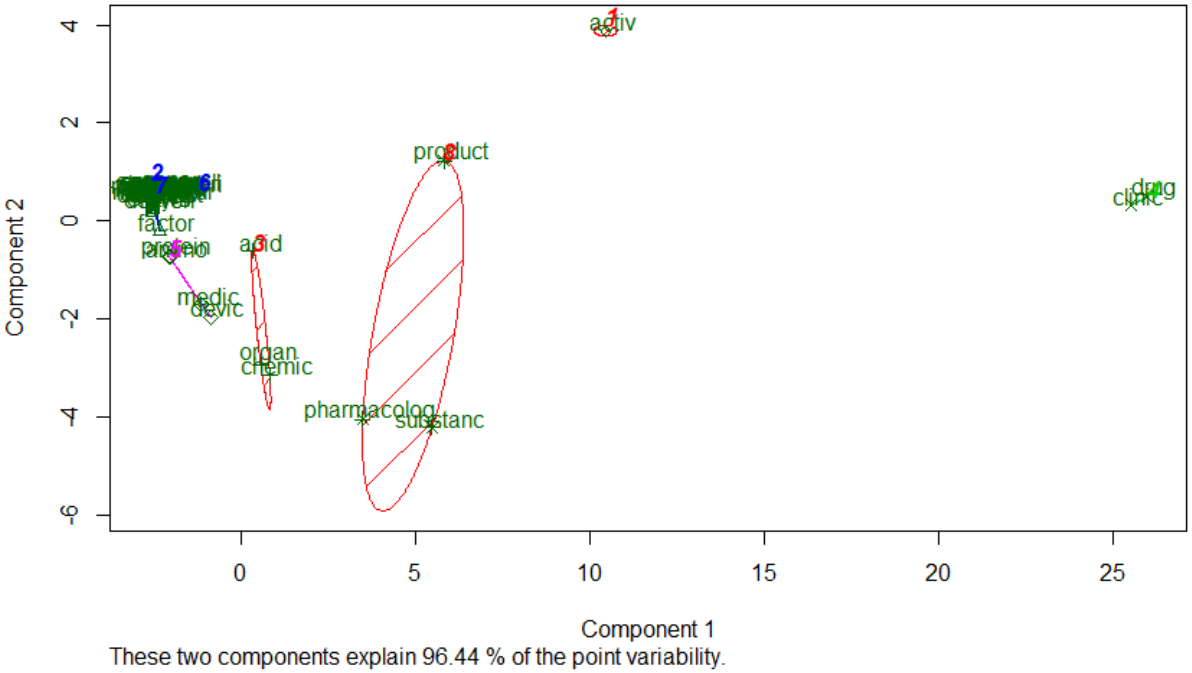


CLUSPLOT - 10% Sparsity, k = 8 means



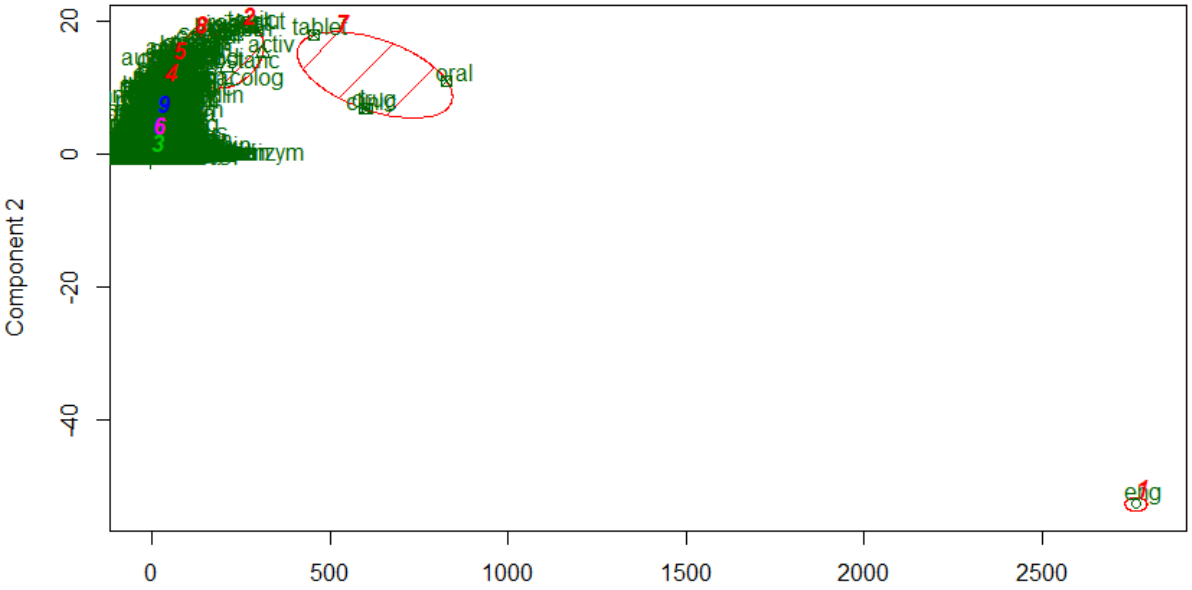
These two components explain 96.44 % of the point variability.

CLUSPLOT - 15% Sparsity, k = 8 means



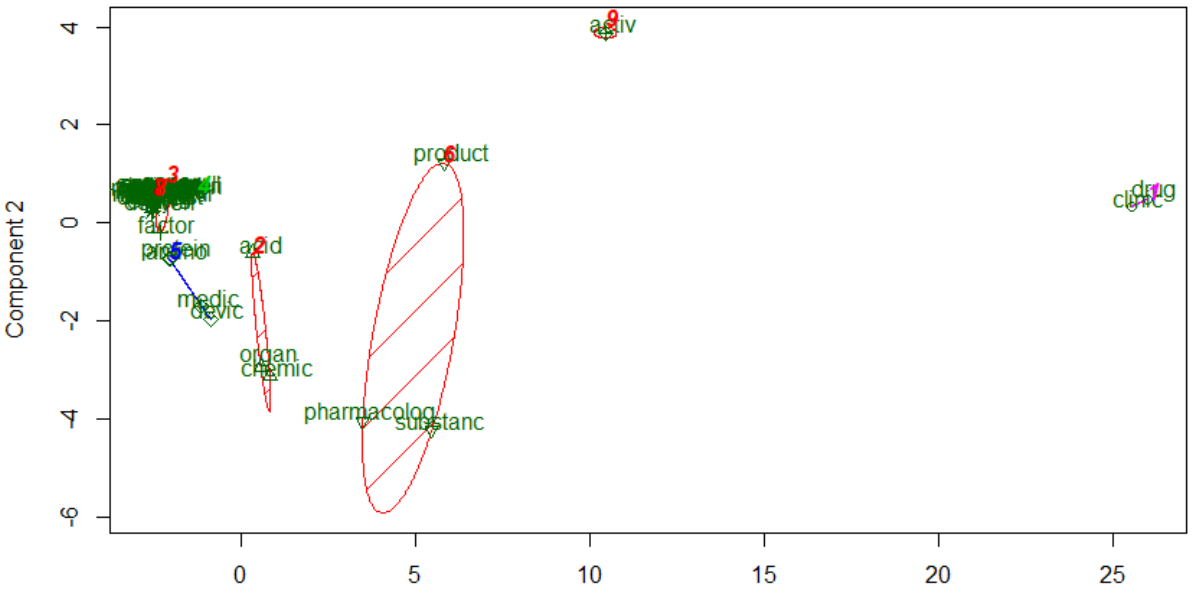
These two components explain 96.44 % of the point variability.

CLUSPLOT - 5% Sparsity, k = 9 means

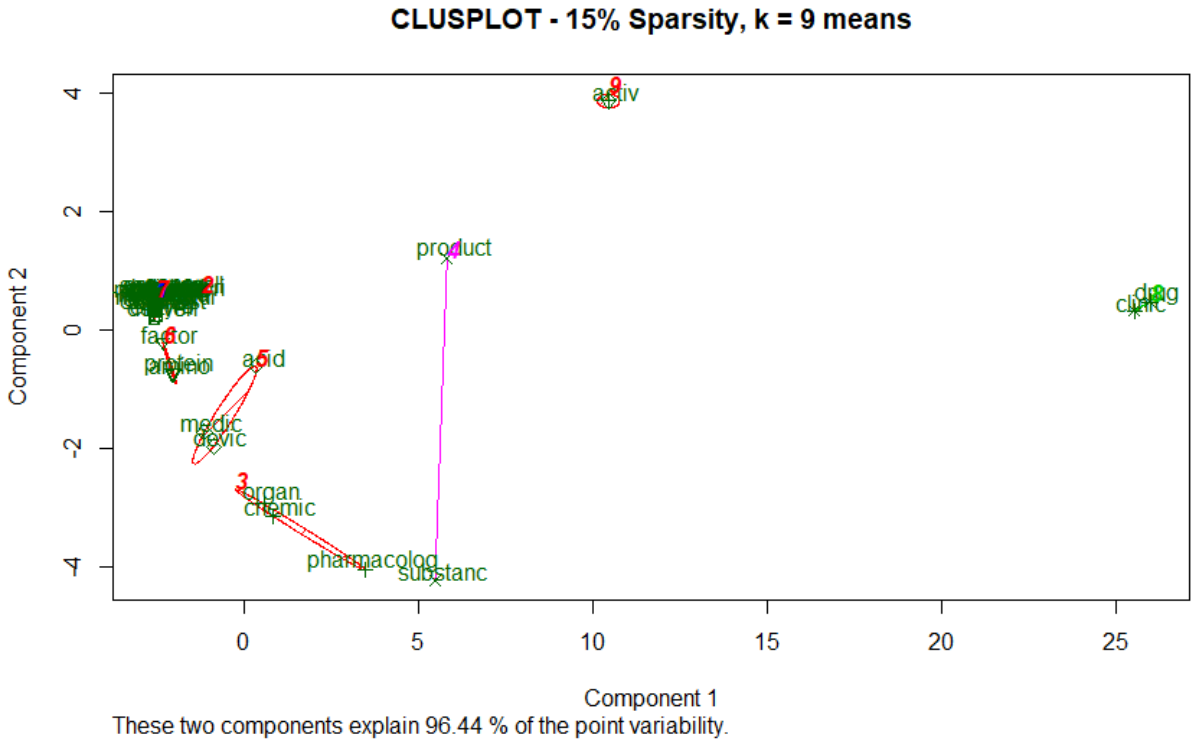


These two components explain 99.84 % of the point variability.

CLUSPLOT - 10% Sparsity, k = 9 means



These two components explain 96.44 % of the point variability.



APPENDIX D

EMR CORE REFERENCE ONTOLOGY ENCODING

Title: 'EMR Core Reference Ontology design'.

Author: 'Ziniya Zahedi'.

Namespace: 'http://ontorion.com/namespace'.

Comment: 'Primitive concept definitions'.

Every clinic is a primitive-concept.

Every drug is a primitive-concept.

Every active is a primitive-concept.

Every acid is a primitive-concept.

Every product is a primitive-concept.

Every pharmacology is a primitive-concept.

Every substance is a primitive-concept.

Every device is a primitive-concept.

Every medical is a primitive-concept.

Every chemical is a primitive-concept.

Every organ is a primitive-concept.

Comment: 'Primitive concepts existential attribute specifications'.

Every medicine is a clinic.

Every practice is a clinic.

Every specialist is a clinic.

Every medicine is a drug.

Every matter is a drug.

Every agent is an active.

Every medicine is an active.

Every physiological is an active.

Every response is an active.

Every chemical is an acid.

Every matter is an acid.

Every ph is an acid.

Every chemical is a product.

Every reaction is a product.

Every matter is a product.

Every effects is a pharmacology.

Every medicine is a pharmacology.

Every treatment is a pharmacology.

Every matter is a substance.

Every instrument is a device.

Every medicine is a device.

Every treatment is a device.

Every medicine is a medical.

Every practice is a medical.

Every study is a medical.

Every matter is a chemical.

Every reaction is a chemical.

Every function is an organ.

Every structure is an organ.

Every unit is an organ.

Comment: 'Primitive concepts state modification attribute specifications'.

Every clinic has-profession equal-to *'medicine'*.

Every clinic has-profession equal-to *'Practice'*.

Every clinic has-curing equal-to *'medicine'*.

Every clinic has-curing equal-to *'Practice'*.

Every clinic has-generalist equal-to *'medicine'*.

Every clinic has-generalist equal-to *'Practice'*.

Every clinic has-specialist equal-to *'medicine'*.

Every clinic has-specialist equal-to *'Practice'*.

Every drug has-profession equal-to *'medicine'*.

Every drug has-profession equal-to *'Matter'*.

Every drug has-curing equal-to *'medicine'*.

Every drug has-curing equal-to *'Matter'*.

Every drug has-pharmaceutic equal-to *'medicine'*.

Every drug has-pharmaceutic equal-to *'Matter'*.

Every active has-causal equal-to *'agent'*.

Every active has-causal equal-to *'medicine'*.

Every active has-causal equal-to *'physiological'*.

Every active has-causal equal-to 'response'.

Every active has-profession equal-to *'agent'*.

Every active has-profession equal-to *'medicine'*.

Every active has-profession equal-to *'physiological'*.

Every active has-profession equal-to *'response'*.

Every active has-curing equal-to *'agent'*.

Every active has-curing equal-to *'medicine'*.

Every active has-curing equal-to *'physiological'*.

Every active has-curing equal-to *'response'*.

Every active has-body equal-to *'agent'*.

Every active has-body equal-to *'medicine'*.

Every active has-body equal-to *'physiological'*.

Every active has-body equal-to *'response'*.

Every active has-pathology equal-to *'agent'*.

Every active has-pathology equal-to *'medicine'*.

Every active has-pathology equal-to *'physiological'*.

Every active has-pathology equal-to *'response'*.

Every acid has-chemic equal-to *'chemical'*.

Every acid has-chemic equal-to *'matter'*.

Every acid has-chemic equal-to *'pH'*.

Every acid has-pharmaceutic equal-to *'chemical'*.

Every acid has-pharmaceutic equal-to *'matter'*.

Every acid has-pharmaceutic equal-to *'pH'*.

Every acid has-0 to 7 potential hydrogen equal-to *'chemical'*.

Every acid has-0 to 7 potential hydrogen equal-to *'matter'*.

Every acid has-0 to 7 potential hydrogen equal-to *'pH'*.

Every product has-chemic equal-to *'chemical'*.

Every product has-chemic equal-to *'reaction'*.

Every product has-chemic equal-to *'matter'*.

Every product has-decomposition equal-to *'chemical'*.

Every product has-decomposition equal-to *'reaction'*.

Every product has-decomposition equal-to *'matter'*.

Every product has-synthesis equal-to *'chemical'*.

Every product has-synthesis equal-to *'reaction'*.

Every product has-synthesis equal-to *'matter'*.

Every product has-pharmaceutic equal-to *'chemical'*.

Every product has-pharmaceutic equal-to *'reaction'*.

Every product has-pharmaceutic equal-to *'matter'*.

Every pharmacology has-result equal-to *'effects'*.

Every pharmacology has-result equal-to *'medicine'*.

Every pharmacology has-result equal-to *'treatment'*.

Every pharmacology has-profession equal-to *'effects'*.

Every pharmacology has-profession equal-to *'medicine'*.

Every pharmacology has-profession equal-to *'treatment'*.

Every pharmacology has-curing equal-to *'effects'*.

Every pharmacology has-curing equal-to *'medicine'*.

Every pharmacology has-curing equal-to *'treatment'*.

Every pharmacology has-therapy equal-to *'effects'*.

Every pharmacology has-therapy equal-to *'medicine'*.

Every pharmacology has-therapy equal-to *'treatment'*.

Every substance has-pharmaceutic equal-to *'matter'*.

Every device has-tool equal-to *'instrument'*.

Every device has-tool equal-to *'medicine'*.

Every device has-tool equal-to *'treatment'*.

Every device has-profession equal-to *'instrument'*.

Every device has-profession equal-to *'medicine'*.

Every device has-profession equal-to *'treatment'*.

Every device has-curing equal-to *'instrument'*.

Every device has-curing equal-to *'medicine'*.

Every device has-curing equal-to *'treatment'*.

Every device has-diagnosis equal-to *'instrument'*.

Every device has-diagnosis equal-to *'medicine'*.

Every device has-diagnosis equal-to *'treatment'*.

Every device has-prognosis equal-to *'instrument'*.

Every device has-prognosis equal-to *'medicine'*.

Every device has-prognosis equal-to *'treatment'*.

Every medical has-profession equal-to *'medicine'*.

Every medical has-profession equal-to *'practice'*.

Every medical has-profession equal-to *'study'*.

Every medical has-curing equal-to *'medicine'*.

Every medical has-curing equal-to *'practice'*.

Every medical has-curing equal-to *'study'*.

Every medical has-generalist equal-to *'medicine'*.

Every medical has-generalist equal-to *'practice'*.

Every medical has-generalist equal-to *'study'*.

Every medical has-specialist equal-to *'medicine'*.

Every medical has-specialist equal-to *'practice'*.

Every medical has-specialist equal-to *'study'*.

Every medical has-understanding equal-to *'medicine'*.

Every medical has-understanding equal-to *'practice'*.

Every medical has-understanding equal-to *'study'*.

Every chemical has-pharmaceutic equal-to *'matter'*.

Every chemical has-pharmaceutic equal-to *'reaction'*.

Every chemical has-decomposition equal-to *'matter'*.

Every chemical has-decomposition equal-to *'reaction'*.

Every chemical has-synthesis equal-to *'matter'*.

Every chemical has-synthesis equal-to *'reaction'*.

Every organ has-transformation equal-to *'function'*.

Every organ has-transformation equal-to *'structure'*.

Every organ has-transformation equal-to *'unit'*.

Every organ has-composition equal-to *'function'*.

Every organ has-composition equal-to *'structure'*.

Every organ has-composition equal-to *'unit'*.

Every organ has-element equal-to *'function'*.

Every organ has-element equal-to *'structure'*.

Every organ has-element equal-to *'unit'*.

Every organ has-element equal-to *'function'*.

Every organ has-element equal-to *'structure'*.

Every organ has-element equal-to *'unit'*.

Comment: 'Primitive axioms specifications'.

Every clinic is-strongly-correlated-with a drug.

Every clinic is-strongly-correlated-with a pharmacology.

Every clinic is-strongly-correlated-with a substance.

Every clinic is-strongly-correlated-with a device.

Every clinic is-strongly-correlated-with a medical.

Every clinic is-strongly-correlated-with a chemical.

Every clinic is-strongly-correlated-with an organ.

Every drug is-strongly-correlated-with a clinic.

Every drug is-strongly-correlated-with a pharmacology.

Every drug is-strongly-correlated-with a substance.

Every drug is-strongly-correlated-with a device.

Every drug is-strongly-correlated-with a medical.

Every drug is-strongly-correlated-with a chemical.

Every drug is-strongly-correlated-with an organ.

Every pharmacology is-strongly-correlated-with a clinic.

Every pharmacology is-strongly-correlated-with a drug.

Every pharmacology is-strongly-correlated-with a substance.

Every pharmacology is-strongly-correlated-with a device.

Every pharmacology is-strongly-correlated-with a medical.

Every pharmacology is-strongly-correlated-with a chemical.

Every pharmacology is-strongly-correlated-with an organ.

Every substance is-strongly-correlated-with a clinic.

Every substance is-strongly-correlated-with a drug.

Every substance is-strongly-correlated-with a pharmacology.

Every substance is-strongly-correlated-with a device.

Every substance is-strongly-correlated-with a medical.

Every substance is-strongly-correlated-with a chemical.

Every substance is-strongly-correlated-with an organ.

Every device is-strongly-correlated-with a clinic.

Every device is-strongly-correlated-with a drug.

Every device is-strongly-correlated-with a pharmacology.

Every device is-strongly-correlated-with a substance.

Every device is-strongly-correlated-with a medical.

Every device is-strongly-correlated-with a chemical.

Every device is-strongly-correlated-with an organ.

Every medical is-strongly-correlated-with a clinic.

Every medical is-strongly-correlated-with a drug.

Every medical is-strongly-correlated-with a pharmacology.

Every medical is-strongly-correlated-with a substance.

Every medical is-strongly-correlated-with a device.

Every medical is-strongly-correlated-with a chemical.

Every medical is-strongly-correlated-with an organ.

Every chemical is-strongly-correlated-with a clinic.

Every chemical is-strongly-correlated-with a drug.

Every chemical is-strongly-correlated-with a pharmacology.

Every chemical is-strongly-correlated-with a substance.

Every chemical is-strongly-correlated-with a device.

Every chemical is-strongly-correlated-with a medical.

Every chemical is-strongly-correlated-with an organ.

Every organ is-strongly-correlated-with a clinic.

Every organ is-strongly-correlated-with a drug.

Every organ is-strongly-correlated-with a pharmacology.

Every organ is-strongly-correlated-with a substance.

Every organ is-strongly-correlated-with a device.

Every organ is-strongly-correlated-with a medical.

Every organ is-strongly-correlated-with a chemical.

VITA

Ziniya Zahedi

Engineering Management & Systems Engineering

2101 Engineering Systems Building

Norfolk, VA 23529

Ms. Ziniya Zahedi (MEng) has received her Bachelor of Science in Business Administration (Major: Marketing) in 2012 and Master in Engineering Management in 2015 from Old Dominion University.

Currently, she is working as a Business Operations Analyst at Georgetown University Law Center. She is also managing two startup businesses on the side, one in Web Design and Digital Media (Trinyan) and the other in Research Analytics and Data Visualization (Lezolve).

Previously, she worked as a Faculty Administrator at Old Dominion University and as a Database Analyst at Eastern Virginia Medical School. Her fields of expertise are engineering management, systems engineering, healthcare, data analytics, marketing research, economics, artificial intelligence, and operations management. She is passionate about research and analytics in different domains. Ms. Zahedi has a long list of conference papers and publications, and some of her previous work has been published in notable scholarly journals.