The Relationship Between User Expertise and Structural Ontology Characteristics

A Thesis

Submitted to the Faculty

of

Drexel University

by

Ilya Michael Waldstein

in partial fulfillment of the requirements for the degree

of

Doctor of Philosophy

November 2013
Acknowledgements

I would like to thank my distinguished committee for their patience, understanding, and invaluable advice for my dissertation.

Advisor:

Dr. Rosina Weber

Committee:

Dr. Neal Handly
Dr. Larry Kerschberg
Dr. Xia Lin
Dr. Jung-ran Park
Dr. Michelle Rogers

I would especially like to thank Dr. Weber for her countless hours of guidance, advocacy, and support without which I would not have achieved this goal.

I would also like to thank the Department of Graduate Studies at Drexel University for their consideration and support.
Table of Contents

List of Tables ................................................................................................................................. vii
List of Figures ................................................................................................................................. viii
Abstract ............................................................................................................................................ ix

CHAPTER 1: Introduction .................................................................................................................. 1
  1.1 Motivation ................................................................................................................................... 4
  1.1.1 Problem Statement ................................................................................................................ 7
  1.2 Research Questions .................................................................................................................... 9
  1.2.1 Overall Research Goals ........................................................................................................ 10
  1.3 Organization of the Dissertation ............................................................................................... 10

CHAPTER 2: Literature Review ......................................................................................................... 11
  2.1 Ontologies .................................................................................................................................. 11
  2.2 Lightweight Ontologies ............................................................................................................. 13
  2.3 Knowledge Representation ......................................................................................................... 14
  2.4 Ontologies as a Form of Knowledge Representation ................................................................. 15
  2.5 Ontology Characteristics .......................................................................................................... 16
  2.5.1 Criteria .................................................................................................................................. 17
  2.5.2 Measures .............................................................................................................................. 18
  2.6 Ontology Engineering ................................................................................................................ 19
  2.7 Evaluation of Knowledge Representation Formalisms ............................................................ 22
  2.8 Ontology Evaluation ................................................................................................................ 24
  2.8.1 Evaluation Areas ................................................................................................................... 24
  2.8.2 Evaluation Approaches ....................................................................................................... 25
  2.8.2.1 Gold-standard approach ................................................................................................. 26
  2.8.2.2 Data-based approach .................................................................................................... 27
  2.8.2.3 Criteria-based approach ............................................................................................... 28
  2.8.2.4 Assessment by humans ................................................................................................. 29
  2.8.2.5 Framework ..................................................................................................................... 30
  2.8.2.6 Task-based approach .................................................................................................... 31
  2.9 Ontology Application .............................................................................................................. 32
  2.9.1 Ontology Supported Tasks ................................................................................................. 32
  2.9.2 Ontology Domains .............................................................................................................. 34
  2.10 Context ....................................................................................................................................... 35
CHAPTER 3: Methodology

3.1 Elements of the Methodology

3.1.1 Context Parameters

3.1.1.1 User

3.1.1.2 Application Task

3.1.1.3 Domain

3.1.1.4 Time Context Parameter

3.1.2 Ontology Characteristics

3.1.2.1 Structural Characteristics

3.1.2.2 Non-structural Characteristics

3.1.3 Objective Quality Metrics

3.2 Methodology Steps

3.2.1 Phase 1

3.2.1.1 Select

3.2.1.2 Measure Initial Performance

3.2.2 Phase 2

3.2.2.1 Change Values

3.2.2.2 Measure Resultant Performance

3.2.2.3 Analyze

3.2.2.4 Save and Apply

3.3 Changing Ontology Characteristics in the Methodology

3.4 General Hypothesis in the Methodology

CHAPTER 4: Studies and Results

4.1 Study Description: Applying the Methodology in a Healthcare Setting

4.2 Elements of the Methodology in the Studies

4.2.1 Context Parameters in the Studies
Appendix B: Ontology Resources ................................................................. 139
Appendix C: Sample Guideline from the National Guideline Clearinghouse .... 142
Appendix D: Sample MeSH Ontology ............................................................ 157
Appendix E: Complete Screenshots of Website .......................................... 160
Appendix F: Complete Text for All Scenarios .............................................. 167
Appendix G: Post Task Survey Answers .................................................... 168
Author Vita ................................................................................................. 179
List of Tables

Table 2-1: Criteria used in ontology engineering and evaluation approaches ........................................ 17
Table 2-2: Measures used in ontology engineering and evaluation approaches .................................... 19
Table 2-3: Evaluation approaches ............................................................................................................. 26
Table 3-1: Phases of the methodology ...................................................................................................... 51
Table 4-1: Study context parameters ........................................................................................................ 58
Table 4-2: Number of participants per subgroup ..................................................................................... 58
Table 4-3: Results for Question 2, participant self-assessment of familiarity with scenario topic 64
Table 4-4: Results for Question 3, keyword selection ............................................................................. 66
Table 4-5: Results for Question 4, participant interpretation of scenario topic ....................................... 69
Table 4-6: MeSH Ontology categories before and after pruning ............................................................. 79
Table 4-7: Characteristics selected for testing using the methodology ..................................................... 80
Table 4-8: Summary calculations for the selected structural characteristics ........................................... 90
Table 4-9: Hypotheses testing the impact of a change to structural ontology characteristics on novices and experts performing a search application task in the medical domain. ......................... 93
Table 4-10: Results for novice and expert resident physician and nursing student groups with significant changes in performance in bold ......................................................................................... 97
Table 4-11: Results for novice resident physicians .................................................................................. 97
Table 4-12: Results for expert resident physicians .................................................................................... 99
Table 4-13: Results for novice vs. expert resident physician task performance ...................................... 100
Table 4-14: Results for novice nursing students ....................................................................................... 101
Table 4-15: Results for expert nursing students ....................................................................................... 102
Table 4-16: Results for novice vs. expert nursing student task performance ........................................ 103
List of Figures

Figure 1-1: Impact of context parameters on ontology engineering ............................................. 4
Figure 1-2: Interrelationship between design and quality ............................................................. 7
Figure 2-1: Sample wine ontology adapted from Noy and McGuinness (2001) ............................... 12
Figure 2-2: Range of ontology types ........................................................................................... 14
Figure 2-3: Concepts of information seeking behavior ................................................................. 40
Figure 2-4: Semantic Web Stack .................................................................................................. 44
Figure 3-1: Change to elements of the methodology between two phases ................................. 46
Figure 3-2: Excerpt of the unchanged MeSH ontology ................................................................. 49
Figure 3-3: General application of the methodology ................................................................. 51
Figure 4-1: Methodology applied in the studies ......................................................................... 57
Figure 4-2: Example scenario ..................................................................................................... 71
Figure 4-3: Excerpt of the scenario step on the Task webpages ................................................. 73
Figure 4-4: Excerpt of the hierarchy search step on the Task webpages .................................... 74
Figure 4-5: Excerpt of the guideline selection step on the Task webpages ............................... 75
Figure 4-6: Sample of the MeSH pneumonia related hierarchy .................................................. 78
Figure 4-7: Excerpt of the unchanged MeSH ontology ............................................................... 85
Figure 4-8: Excerpt of the changed MeSH ontology ................................................................. 86
Figure 4-9: Summary of task performance results for resident physician and nursing student studies .......................................................................................................................... 111
Figure 5-1: Research questions answered by methodology ....................................................... 114
Figure 5-2: Guideline components ............................................................................................. 118
Abstract

The Relationship Between User Expertise and Structural Ontology Characteristics
Ilya Michael Waldstein, MSIS
Rosina O. Weber, Ph.D.

Ontologies are commonly used to support application tasks such as natural language processing, knowledge management, learning, browsing, and search. Literature recommends considering specific context during ontology design, and highlights that a different context is responsible for problems in ontology reuse. However, there is still no clear direction for ontology design in engineering for a variety of different contexts expressed through contextual parameters, e.g., what kind of design will better serve specific users performing a specific application task in a specific domain? There is an inbuilt connection between ontology design during ontology engineering and ontology quality. For design to determine quality, we need to first assess quality to inform design.

Ontologies are described through characteristics. Context is described through parameters. These studies investigate the relationship between ontology characteristics and context parameters in order to support guidance for ontology design for a specific context. We compare the impact of a change in ontology characteristics in a lightweight ontology on the performance of novices versus experts completing a search application task in the medical domain. We propose a methodology that associates ontology design through ontology characteristics with context parameters in order to evaluate their combined performance. We use Path Selected and Guidelines Selected as the objective performance metrics. The results show a significant negative difference in performance
for a medical search application task completed by novice residents and nursing students due to a change made to the ontology’s structural characteristics that impact expertise and no significant difference for expert residents and nursing students.
CHAPTER 1: Introduction

All forms of knowledge require context to be understood. Knowledge is the application of information within a context. Experts apply knowledge by utilizing rules about a context. Knowledge can be seen as the behavior of an intelligent agent in its environment (Guarino, 1995). Knowledge representation is a surrogate that represents a certain view of a domain (Davis et al., 1993). Ontology is one way to represent knowledge. Ontology is an explicit specification of a conceptualization (Gruber, 1995). A conceptualization is an abstract, simplified view of the world represented for some purpose (Gruber, 1995). Ontologies are an integral part of the Semantic Web. They act as the knowledge representation of a domain and enable knowledge sharing (Chandrasekaran et al., 1999). Ontologies are an important layer that supports data by structuring and defining content.

The way to describe, to manipulate, to evaluate, to improve, and to understand ontology is through its characteristics. Ontology design, engineering, and evaluation literature describes multiple types of ontology characteristics. Ontology characteristics can be structural, conceptual, or user defined (Yu et al., 2009). Structural characteristics are physical dimensions of the ontology schema. They relate to the ontology itself. Conceptual characteristics describe the ontological or philosophical dimensions of ontology concepts. They relate to the knowledge base. User defined characteristics describe hurdles that the ontology must meet in order to satisfy an arbitrary dimension selected and defined by the ontology user.

Characteristics are expressed through criteria or measures. Criteria are hurdles set by an author of an evaluation approach that have to be satisfied for the ontology to be
considered correct within the scope of that approach. An example of a user defined characteristic that is a criterion is conciseness, which is part of a criteria-based approach proposed by Gómez-Pérez. In the approach, concept definitions are compared against definitions of other concepts and if there are no duplicate definitions, similar definitions between concepts, and the original definition cannot be additionally inferred based on other definitions or axioms, the ontology is considered concise (Gomez-Perez, 2001).

Measures are calculable metrics based on some type of ontology parameter. An example of a measure is depth. Depth is one of the structural dimensions of an ontology in a framework proposed by Gangemi et al. Depth is defined as “a graph property (a topological property) related to the cardinality of paths in a graph, where the arcs considered are only is-a arcs (Gangemi et al., 2006).”

Ontologies can range in their specification from lightweight to heavyweight. While there are no concrete definitions, lightweight ontologies can consist of terms that have only minimal specification of the term meaning whereas heavyweight ontologies consist of rigorously formalized logical theories (Uschold & Grunninger, 2004). Typically, lightweight ontologies consist of a simple taxonomy (Uschold & Grunninger, 2004) where there is a trade-off between their expressivity and usability (Brewster & O’Hara, 2007). Literature illustrates the relevance of lightweight ontologies. Only ten percent of currently implemented ontologies by an ontology vendor involve inferential requirements, and research shows that over-engineered systems tend to not be applied (Brewster & O’Hara, 2004). This research from 2004 and the evident growth of lightweight ontologies on the Web both suggest that lightweight ontologies are the largest proportion of ontologies used today.
Both lightweight and heavyweight ontologies are ubiquitous today. For example, heavyweight ontologies are used in Apple’s Siri, a widely available commercial product. Siri is a virtual personal assistant that uses heavyweight active ontologies that support specific domains and tasks such as locating restaurants and making reservations, purchasing tickets for sporting events, or providing stock information, as well as more general tasks such as searching the web. The Internet contains many hierarchical structures that can be considered lightweight ontologies, such as web directories (e.g. Yahoo!, Open Directory Project), category structures (e.g. Wikipedia, National Guideline Clearinghouse), or online thesauri (e.g. Art & Architecture Thesaurus).

Ontologies are engineered, evaluated, and used across a wide range of domains and tasks. Ontologies are already employed in a number of domains including museums (Collins et al., 2005), legal libraries (Gilardoni et al., 2005), international affairs (Rodrigo et al., 2005), military battle spaces (Song et al., 2005), syndicated Web content (Abel et al., 2005), and electronic newspapers (Maidel et al., 2008). Ontologies also support a variety of application tasks such as classification (Cheng et al., 2003), annotation (Vargas-Vera et al., 2002), search (Rodrigo et al., 2005) (Gilardoni et al., 2005) (Kerschberg et al., 2004), mediation (Muthaiyah & Kerschberg, 2007), browsing (Collins et al., 2005), and personalization (Chen, 2008). Ontologies require a context. Literature defines context in terms of context types in general (Dey & Abowd, 1999) and context parameters specifically. In this work context is defined by parameters such as user, application task, domain, and time to be explained in detail in Section 2.10 Context. Each context parameter has an impact on the engineering, evaluation, and use of an ontology as shown in Figure 1-1. This impact is related to the specific parameter. For
example, in a domain of fruit farming an ontology used for the application task of selecting pesticides would not be the same as the ontology for selecting fruits for cooking (Chandrasekaran et al., 1999).

1.1 Motivation

Many current ontology engineering and evaluation approaches are focused on constructing an ontology that correctly represents a domain. Different techniques are used depending on the properties of the domain content (i.e. documents, images, sound, tacit knowledge). The techniques can produce an ontology that is a good representation of the domain. However, when the ontology is used to support a user performing a task within that domain, the combination of the ontology and context fails to produce acceptable performance results. The poor performance by itself cannot be used as an indicator of the specific changes in the ontology to improve the combined performance.
Even task-based evaluation approaches that use competency questions to determine performance cannot currently point to specific changes for ontology engineers. The reason for this disconnect is that, with respect to context, ontology evaluation is akin to black-box software testing that examines the functionality of an ontology without peering into its internal structure. The surrounding context is not described and linked to the characteristics of the ontology.

Ontology engineering literature recommends taking context into account when engineering ontologies. Tatir et al. (2010) recommend considering the specific goals and actions of users when designing ontologies for systems that support tasks. Furthermore, they state that ontology engineering and evaluation must support changes in conceptualization which can result from a changing view of the world and from the change in usage perspective. For example, a university is very different from the faculty perspective versus the student perspective, and the adoption of a specific perspective impacts the ontology design (Tatir et al., 2010). Noy and McGuinness (2001) state that ontology engineering needs to incorporate information about the task in order to properly represent the domain. Tempich et al. (2006) describe the DILIGENT process that focuses on ontology engineering in real settings and includes local adaptation and local update steps, with the inclusion of the users of the ontology, performed cyclically to incorporate the changing user needs over time.

Context also plays an important part in the reuse of ontologies. Russ et al. encountered problems during the merging of the INSPECT and Aircraft ontologies into a single JFACC ontology. For example, in the INSPECT ontology the aircraft classification was determined by mission while the Aircraft ontology considered other
properties such as the technical aspects of airplanes (Russ et al., 1999). In this case the modeled concepts presented problems for reuse because they were application task specific and differed in the two ontologies.

Ontology related research literature describes context as being important, but does not specify how or when to include context parameters in ontology engineering or evaluation. Additionally, the engineering approaches mentioned before do not make a direct connection between ontology characteristics and context parameters as they impact ontology quality, nor do they describe how to use contextual parameters during design.

Since the inception of ontology engineering and evaluation there has been a call for more research into this area. Brewster et al. (2004) called for research to test a number of different ontologies to determine which ontology was the most appropriate for an application, and determine what aspects of the ontology, the application, and the combination made it the most appropriate. They emphasized the need for methods and tools to evaluate ontologies, and to develop techniques that could determine how appropriate a particular ontology is to a domain (Brewster et al., 2004). Despite that need there is still no structured methodology to test ontology performance taking characteristics and context parameters into account, or trying to link the two in a meaningful way to guide ontology engineering and evaluation.
1.1.1 Problem Statement

Currently there is no scientific guidance that provides ontology engineers with a way to determine what kind of characteristics an ontology should have for a specific context, i.e., what characteristics it should have for a specific type of user and domain to be adequate to support a specific application task. Both ontology engineering and evaluation approaches are presently context independent.

The overall objective of this research is to develop a methodology that will allow for the testing of changes to ontology characteristics across different contexts.

Rationale

The proposed methodology tests ontology characteristics against a context parameters and will provide the foundational research to test the relationship between the two. The methodology will also provide a reproducible process for testing different combinations of characteristics and parameters.

This research aims to provide guidance for engineering ontologies for a context.

Figure 1-2: Interrelationship between design and quality
Specifically, the research provides guidance about the characteristics that may support quality for an ontology in a given context. Note that the same ontology characteristics used to describe an ontology when examining its quality are used for guiding design during engineering. There is an inherent connection between ontology design and quality as depicted in Figure 1-2. For design to determine quality, we need to first assess quality to inform design. It stands to reason that design is correlated with quality, that design requires consideration of the context of use, and that quality also requires consideration of the context of use.

Ontology is a form of knowledge representation, thus its use in context has different impact on different users trying to accomplish their goals. Typically, ontologies are built by groups of experts who use specific terms and rules from the domain based on their domain knowledge. The ontology is then used in the domain to support the execution of a task by all users problems can occur. Novice users do not know the terms used by experts, and cannot utilize the ontology to its full capability. Furthermore, if changes are made that impact the ontology’s ability to support the task, the effect can be different for different users. Experts can use their knowledge to offset a negative change in the ontology whereas novices cannot. Novices will be adversely affected in the performance of an application task supported by the changed ontology.

Developing a methodology and finding a relationship between ontology characteristics and context parameters will help to find a better way to engineer and evaluate ontologies. These relationships would shed light as to how to guide the construction and evaluation of ontologies, i.e., how effective are characteristics to indicate quality of an ontology in support of a given application task for a given user type.
and domain. Significant results would allow for the development of recommendations based on statistically accurate metrics. Additionally, the methodology developed as part of this research can be utilized by ontology engineers to serve as a repeatable procedure that would help to incorporate context into engineering and evaluation. We would have a methodology that allows us to provide scientific guidance for ontology engineering in context.

1.2 Research Questions

In this research, characteristics are defined as criteria (e.g. competency, correctness) and measures (e.g. depth, breadth, fanoutness). They are described in detail in Chapter 2 in the Ontology Characteristics section.

A context parameter is characterized by the user, the application task, and the domain. The context varies as at least one of the parameters varies. The user can vary based on different levels of understanding and/or experience about the domain. The application task can vary based on the actions of the user. The domain can vary based on different fields of study, underlying set of objects, or area of activity.

The research questions frame the overall research goals and are:

- **R1:** Is there a relationship (either positive or negative) between characteristics and performance for a given context?
- **R2:** Can ontology characteristics be used as indicators of ontology quality?
- **R3:** Can we develop guidelines for ontology engineering and evaluation taking context into account?
1.2.1 Overall Research Goals

This research has three overall goals:

- The first goal is to investigate the potential relationship between ontology characteristics expressed by criteria and measures with different contexts expressed by parameters of user type, application task, and domain to provide guidelines for their use in ontology engineering and evaluation.
- The second goal is to test a structured methodology for linking ontology characteristics to context parameters.
- The third goal is to confirm the role of lightweight ontologies as knowledge representation formalisms that support users in performing expert tasks.

1.3 Organization of the Dissertation

This dissertation is organized into five chapters. Chapter 1 has provided the introduction to this research, framing the problem and motivation, stating the research questions and the overall goals. Chapter 2 provides a literature review and necessary background to understand the methodology and study. Chapter 3 describes the methodology used to link characteristics and parameters, and to test their combined performance using an application task. Chapter 4 describes the implementation of our methodology in the study and presents the study results. Chapter 5 explains how the implementation of the methodology in the study and the results answer the research questions. Chapter also lists the contributions of this research, potential limitations, and future work.
CHAPTER 2: Literature Review

This chapter will provide the necessary theoretical and applied background to understand the methodology presented in Chapter 3. The chapter explains the history, provides a definition, and explains ontology differentiation. The use of ontologies as knowledge representation is also explained.

2.1 Ontologies

Ontology, a term that originates in philosophy, is the study of existence or the nature of being and of the kinds of entities, both abstract and concrete. Ontology is expressly concerned with categories and their relationships. “The two sources of ontological categories are observation and reasoning. Observation provides knowledge of the physical world, and reasoning makes sense of observation by generating a framework of abstractions called metaphysics (Sowa, 2000).” Ontologies express theories about content – about the sorts of objects, properties of objects, and possible relations between objects – for a specified domain of knowledge. Ontologies formally express the terms that describe knowledge about the specified domain (Chandrasekaran et al., 1999).

The ontology term was later co-opted by computer and information scientists to represent an analogous concept, that of expressing information about a domain for the purpose of facilitating communication and understanding between organizations and software agents of those organizations. Ontology has come to mean the conceptualization of the terms in the domain vocabulary and the body of knowledge describing some domain (Chandrasekaran et al., 1999).
A more technical definition for an ontology is offered by Blomqvist (2005) where it is defined as “a hierarchically structured set of concepts describing a specific domain of knowledge that can be used to create a knowledge base. An ontology contains concepts, a subsumption hierarchy, arbitrary relations between concepts, and axioms. It may also contain other constraints and functions (Blomqvist, 2005).” However, the most common definition is that used by Gruber, and defined as an “explicit specification of a conceptualization (Gruber, 1993).”

Figure 2-1 shows a sample wine ontology. Wine is the most general concept in the hierarchy, White, Rosé, and Red wines are general top level concepts, and Pauillac and Margaux are the most specific bottom level concepts.
2.2 Lightweight Ontologies

Ontologies can be classified based on the degree of formality and expressivity of the language used to specify them (Uschold & Grunninger, 2004). Figure 2-2 adapted from (Uschold & Grunninger, 2004) and (Giunchiglia & Zaihrayeu, 2007), shows the full range of specification structures and their interpretation on the ontology specification spectrum. As we move from informal lightweight ontologies to formal heavyweight ontologies the ontology structure becomes more expressive and formalized (Uschold & Grunninger, 2004). More expressivity and formalization leads to less ambiguity and better support for automated reasoning. Expressivity is the amount of definition of the term in the ontology. It is the hierarchy, other relationships, and especially axioms that provide for constraints and associations between the terms. Formalization refers to the ontology being encoded in a logic-based language suitable for automating reasoning such as versions of OWL-DL. Lightweight ontologies have minimal expressivity and may not be coded in a formal language. Typically, lightweight ontologies consist of a simple taxonomy (Uschold & Grunninger, 2004). Informal ontologies include controlled vocabularies, taxonomies, thesauri, business catalogues, faceted classifications, web directories, and user classifications as shown in Figure 2-2.
An example of a lightweight ontology is the Medical Subject Headings (MeSH) taxonomy utilized by the National Guideline Clearinghouse (NGC) to structure medical guidelines on the Web.

2.3 Knowledge Representation

Knowledge representation is a surrogate used by an entity to determine consequences by thinking following a theory of intelligent reasoning in a medium for pragmatically efficient computation that requires an ontological commitment implemented in some representation technology using a mode of human expression (Davis et al., 1993). The definition has several implications for this research. First, because knowledge representation is a surrogate, it cannot be completely accurate because it depends on the specified attributes versus those omitted (Brewster & O’Hara, 2007). Additionally, the eventual use of the surrogate must be taken into account to properly correspond to the real world concept. This correspondence becomes the surrogate’s semantic meaning (Davis et al., 1993). Second, because knowledge representation requires an ontological
commitment that provides a certain view of the world. For example, ontologies in the healthcare field can have a view from the perspective on anatomy, organisms, diseases, chemicals and drugs, or diagnostic and therapeutic techniques. These are all valid as ontological commitments for healthcare, but present very different alternatives. Finally, because knowledge representation involves a mode of human expression that tells other people (or machines) about the world it acts as a channel for expression and communication. Its function as a medium for communication becomes vitally important because knowledge representation can make some tasks possible but difficult to perform. In this case real-world users do not know if they misunderstand the knowledge representation, if they cannot apply it correctly to complete the task, or if the representation itself does not express the required knowledge to achieve their intended goal (Davis et al., 1993).

2.4 Ontologies as a Form of Knowledge Representation

Ontology is one way to represent knowledge. Ontology is an explicit specification of a conceptualization (Gruber, 1995). A conceptualization is an abstract, simplified view of the world represented for some purpose (Gruber, 1995). An ontology is a surrogate where external objects are represented in order to support reasoning. Ontology is also a set of ontological commitments in that the concepts selected and their inter-relationships provide a particular perspective about the world (Brewster & O’Hara, 2007). In other words, ontological commitments are agreements to use a shared vocabulary in a coherent and consistent manner (Gruber, 1995). As in other knowledge representation, ontology is a medium for human expression used for communication and knowledge sharing (Gruber, 1995) that uses human language to represent the world (Brewster & O’Hara,
In this way, ontologies are eminently suitable for representing taxonomic information (Brewster & O’Hara, 2007), and there are numerous applications now utilizing this strength to support different tasks.

2.5 Ontology Characteristics

The way to describe, to manipulate, to evaluate, to improve, and to understand ontology is through its characteristics. Ontology engineering and evaluation literature describes multiple types of ontology characteristics. Ontology characteristics can be structural, conceptual, or user defined (Yu et al., 2009). Structural characteristics are based on the physical properties of the ontology such as number of concept nodes, number of concept relationships, ratio of populated concept class instances, or average concept class population. Structural characteristics are physical dimensions of the ontology schema and relate to the ontology itself. Conceptual characteristics are based on semantic properties of the ontology such as concept is-a relationship structure, inconsistency of concept definitions, repetitive concept definitions, or ontological commitment of concepts. Conceptual characteristics describe the ontological or philosophical properties of ontology concepts and relate to the knowledge base. User defined characteristics depend on a definition created by the user that is external to the ontology itself such as relevance, authority, or history of the ontology. User defined characteristics describe hurdles that the ontology must meet in order to satisfy an arbitrary dimension selected and defined by the ontology user.

Characteristics are expressed through criteria or measures. The type of metric used determines the type of characteristic. Criteria are characteristics based on meeting a set standard. Measures are characteristics based on a calculation of a metric.
2.5.1 Criteria

Criteria are defined by the author within the approach. They do not necessarily have to have a calculation associated with them. Table 2-1 lists different examples of criteria used in ontology engineering and evaluation approaches. An example of a criterion is **Conciseness**, which is part of a criteria-based approach proposed by Gómez-Pérez. “An ontology is concise if it does not store any unnecessary or useless definitions, if explicit redundancies do not exist between definitions, and redundancies cannot be inferred using other definitions and axioms (Gomez-Perez, 2001).”

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burton-Jones et al. (2003)</td>
<td>Lawfullness</td>
</tr>
<tr>
<td></td>
<td>Richness</td>
</tr>
<tr>
<td></td>
<td>Interpretability</td>
</tr>
<tr>
<td></td>
<td>Consistency</td>
</tr>
<tr>
<td></td>
<td>Clarity</td>
</tr>
<tr>
<td></td>
<td>Comprehensiveness</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Relevance</td>
</tr>
<tr>
<td></td>
<td>Authority</td>
</tr>
<tr>
<td></td>
<td>History</td>
</tr>
<tr>
<td>Gruber (1993)</td>
<td>Clarity</td>
</tr>
<tr>
<td></td>
<td>Coherence</td>
</tr>
<tr>
<td></td>
<td>Extendibility</td>
</tr>
<tr>
<td></td>
<td>Minimal ontological commitment</td>
</tr>
<tr>
<td>Gruninger and Fox (1995)</td>
<td>Competency</td>
</tr>
<tr>
<td>Gomez-Perez (2001)</td>
<td>Consistency</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Conciseness</td>
</tr>
<tr>
<td></td>
<td>Expandability</td>
</tr>
<tr>
<td></td>
<td>Sensitiveness</td>
</tr>
<tr>
<td>Guarino (1998)</td>
<td>Corectness (Identity and Dependence)</td>
</tr>
<tr>
<td>Guarino and Welty (2002)</td>
<td>Corectness (Essence, Rigidity, Identity, Unity)</td>
</tr>
</tbody>
</table>

Several criteria have the same name in different approaches; however they are not the same. For example, Burton-Jones et al. (2003) define Clarity as “the average number of
word senses” in WordNet which would make the context of the term clearly defined. Their example is “if an ontology claims that class ‘Chair’ has the property ‘Salary,’ an agent must know that this describes academics, not furniture (Burton-Jones et al., 2003).” However, Gruber (1993) defines Clarity as an ontology that “effectively communicates the intended meaning of defined terms” that are objective and “independent of social or computational context.” Thus, the two definitions are specific to the approach and author.

2.5.2 Measures

Characteristics that can be calculated are called measures. There are several different types of measures related to the underlying characteristic they are based on and their intended use. The structural, coupling, and instance type measures are based on structural characteristics. The schema measures are based on ontological or philosophical characteristics. The consistency, completeness, conciseness, and coverage measures are based on user defined characteristics. Table 2-2 shows various measures used in ontology engineering and evaluation approaches. An example of a measure is depth. Depth is one of the structural dimensions of an ontology in a framework proposed by Gangemi et al. Depth is defined as “a graph property (a topological property) related to the cardinality of paths in a graph, where the arcs considered are only is-a arcs. This measure only applies to digraphs (directed graphs) (Gangemi et al., 2006).” Appendix A: Candidate Criteria and Measures describes the implementation of additional measures found in ontology engineering and evaluation.
<table>
<thead>
<tr>
<th>Researcher</th>
<th>Measure Type</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gangemi et al. (2006)</td>
<td>Structural</td>
<td>Depth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Breadth</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tangledness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fanoutness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Density</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modularity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Differential specifica</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Meta-logical adequacy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degree distribution</td>
</tr>
<tr>
<td>Tatir et al. (2007)</td>
<td>Schema</td>
<td>Relationship richness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attribute richness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inheritance richness</td>
</tr>
<tr>
<td></td>
<td>Instance</td>
<td>Class richness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average population</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cohesion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Importance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fullness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Connectivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Readability</td>
</tr>
<tr>
<td>Orme et al. (2006)</td>
<td>Coupling</td>
<td>Number of external classes (NEC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reference to external classes (REC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Referenced includes (RI)</td>
</tr>
<tr>
<td>Gomez-Perez (1996)</td>
<td>Consistency</td>
<td>Circularity error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partition error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Semantic inconsistency error</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
<td>Incomplete concept classification</td>
</tr>
<tr>
<td></td>
<td>Consiseness</td>
<td>Grammatical redundancy error</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identical formal definition</td>
</tr>
<tr>
<td>Brewster et al. (2004)</td>
<td>Coverage</td>
<td>Number of overlapping concepts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vector space similarity measure</td>
</tr>
</tbody>
</table>

### 2.6 Ontology Engineering

Ontology engineering and implementation are important to the development and success of the Semantic Web. Noy and McGuinness describe an iterative manual process for engineering ontologies consisting of seven steps: 1) determine domain and scope of the ontology, 2) consider reusing existing ontologies, 3) enumerate important terms in the ontology from the domain, 4) define the classes and the class hierarchy, 5) define the properties of classes-slot, 6) define the facets of the slots, 7) create instances (Noy &
McGuinness, 2001) (Cristani & Cuel, 2005). This type of process was used by ontology engineers and domain experts, alone or in teams, to engineer ontologies. However, manual ontology-engineering practices insufficiently address at least five fundamental aspects of building and committing to ontologies (Hepp, 2007):

1. **Ontology engineering lag versus conceptual dynamics.**
   Can we build ontologies fast enough to reflect quickly evolving domains?

2. **Resource consumption.**
   Does the gain in automation that the ontology provides justify the resources needed to develop it? From another perspective, do the technical problems that the ontology can help us solve outweigh the problems we must master to create it?

3. **Communication between creators and users.**
   Can the individuals who consider using an ontology to annotate data or express queries easily grasp the meaning of all the elements as intended by the ontology creators?

4. **Incentive conflicts and network externalities.**
   Is the incentive structure for relevant actors in the process compatible with the required contributions? For example, are those who must dedicate time and resources benefiting from the ontologies? Moreover, ontologies exhibit positive network effects, such that their perceived utility increases with the number of people who commit to them. This implies that convincing individuals to invest effort into building or using ontologies is particularly difficult while the user base associated with it is small or nonexistent.
5. Intellectual property rights.

For many applications, we need ontologies that represent existing standards. However, standards are often subject to intellectual property rights. Establishing the legal framework for deriving ontologies from relevant standards is thus nontrivial.

Reducing the effort required to engineer the ontology and using engineering methods such as collaboratively evolving an ontology while performing daily work would address some of the shortcomings outlined by Hepp (Liu & Gruen, 2008). Additionally, researchers have started to look at the technology required to make automated and automatic ontology engineering a reality (Blomqvist, 2005) (Cristani & Cuel, 2005). Semi-automated and automated methods could potentially solve the issues mentioned above. For example, automation would help with engineering lag and resource consumption by extracting terms and relationships, and organizing the terms into a hierarchy.

Several semi-automated methods have already been developed. They include a method based on extracting ontology information from a schema of HTML web forms and creating instances from the data located on the HTML web forms pages (Astrova & Stantic, 2005), along with a similar approach that overcomes the lack of identification of inheritance relationships (Benslimane et al., 2006). An approach that incorporates Swoogle and ontology parameters, and merges existing ontology sections into a new ontology that is to be engineered was also implemented (Astrova & Stantic, 2005). Fortuna et al. developed a method based on Latent Semantic Indexing (LSI), K-Means clustering, keyword extraction using centroid vectors and Support Vector Machine
(SVM) to extract terms and relationships for ontology engineering (Fortuna et al., 2006). Karoui et al. use an unsupervised hierarchical clustering algorithm called COCE which uses the K-means partitioning algorithm and structural context from HTML pages to perform incremental clustering to produce, divide, and then refine semantic clusters as it works to build a hierarchy for the ontology (Karoui et al., 2006). Finally, a fuzzy domain ontology discovery method utilizing contextual information and a Balanced Mutual Information (BMI) method was used to engineer an ontology based on Web documents (Lau et al., 2007).

A fully automated method utilizing a pattern ranking and selection approach based on the highest coverage match between a library of ontology patterns and the extracted terms and relations was used to engineer an ontology for the automotive supplier domain (Blomqvist, 2005).

Most of the methods above rely on the underlying domain content without incorporating additional knowledge about context such as the user type or the final application task where the ontology will be applied. Because many of these techniques are automated or automatic, the choice of characteristics when the ontology is engineered is a reflection of the chosen automation method without the incorporation of additional context parameters.

2.7 Evaluation of Knowledge Representation Formalisms

Ontology evaluation has borrowed some of its evaluation techniques and methodologies from prior and current evaluation of similar knowledge representation formalisms and structures. Eckert et al. assess thesaurus suitability for the task of document thesaurus-based indexing and annotation by visualizing the difference of
Information Content (IC) and Intrinsic Information Content (IIC) in a visualization thesaurus analysis tool (Eckert et al., 2007). The equation specifying the difference is similar to the Kullback-Leibler Divergence used in information theory which has also been applied to the evaluation of ontologies (Eckert et al., 2007). With respect to thesauri, manual thesauri are evaluated in terms of soundness, coverage of classification, and thesaurus item selection; and automatically constructed thesauri are evaluated in terms of improving information retrieval via query expansion (Jing & Croft, retrieved 2009). Similar evaluations are performed for ontologies or used for ontology enrichment (Faatz & Steinmetz, 2004). In an overview of thesaurus construction and evaluation, Owens and Cochrane state that evaluation should compare the thesaurus to the four principal purposes of the ANSI/NISO guidelines of translation, consistency, indication of relationships, and retrieval (Owens & Cochrane, 2004). Evaluation methods mentioned also include the use in an IR system to improve relevance, card sorts to improve hierarchy, and focus groups to improve vocabulary terms. Comparative studies, similar to a gold-standard approach for ontology evaluation, are also performed between two thesauri to determine the best audience for the thesaurus (Owens & Cochrane, 2004). An overview of the evaluation of knowledge-based systems (KBS) in general and legal knowledge-bases systems (LKBS) in particular was presented by Hall and Zeleznikow and shows that frameworks presented acknowledge including users and impact on environment; call for expert participation to interpret the symbolic encoding of domain knowledge; and suggest splitting up the system into a software component which is verified and a model component which is validated (Hall & Zeleznikow, 2001). The evaluation of knowledge-based systems shows the difficulty in evaluating a system where
the application is a model of knowledge, or a conceptualization of knowledge when referring to ontologies, and the difficulty of separating the evaluation of the model versus the evaluation of the system in its environment.

2.8 Ontology Evaluation

Ontology evaluation refers to checking that the ontology correctly conceptualizes the real world, that the content has been correctly implemented, and that the ontology performs correctly in the real world (Gomez Perez, 2001). The “real world” refers to context, or more specifically to the parameters of identity (users), location (domain), activity (application task), and time which describe it in a measurable way.

2.8.1 Evaluation Areas

Evaluation can be separated into four general areas: verification, validation, assessment, and benchmarking (Gomez-Perez, 1996). Verification, which can be seen as an internal process, is used to determine if the ontology meets specifications or some conditions imposed at the start of development. Validation, which can be seen as an external process, is used to determine if the ontology accomplishes the functionality specified and involves acceptance of fitness for purpose with end users. Assessment is used to determine the usability and usefulness of the ontology within the software environment or application when they are used, reused, or shared. Benchmarking is used to determine the technical capability of the ontology using various performance metrics and provides a snapshot of the performance of the ontology in relation to some established standard.
2.8.2 Evaluation Approaches

Evaluation approaches are designed by their authors to provide a structure to evaluate an ontology in an evaluation area. Table 2-3 lists various evaluation approaches. When an ontology is evaluated, it is actually the characteristics that are being compared in order to determine ontology “correctness” according to the approach. As stated previously, characteristics are divided into criteria and measures based on the type of metric used. Each approach defines what it means to be correct according to specific criteria or measures. For example, in OntoClean each class is assigned several of the four metaproperties of essence, rigidity, identity and unity that suitably describe its behavior (Guarino & Welty, 2002). An essential property of hammers is hardness, but it is not an essential property of sponges, as a sponge can be hard when dry and soft when wet. A definition of hardness as an essential property of sponges would then violate the ontological notion of the essence metaproperty. An analysis is performed in this manner based on the ontological notions associated with the properties for each class. When all of the classes do not violate the ontological notions, then the ontology is considered correct. Thus essence, rigidity, identity and unity are criteria that determine correctness of an ontology in OntoClean.
Table 2-3: Evaluation approaches

<table>
<thead>
<tr>
<th>Evaluation Approach</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>gold standard</strong></td>
<td>target ontology is compared against a benchmark ontology</td>
</tr>
<tr>
<td>- Maedche &amp; Staab (2002)</td>
<td></td>
</tr>
<tr>
<td><strong>data-based</strong></td>
<td>target ontology is compared against the documents describing the original domain</td>
</tr>
<tr>
<td>- Brewster et al. (2004)</td>
<td></td>
</tr>
<tr>
<td><strong>criteria-based</strong></td>
<td>target ontology is compared against a set of predefined dimensions that were defined to represent quality – or another evaluation criterion</td>
</tr>
<tr>
<td>- Gomez-Perez (1996)</td>
<td></td>
</tr>
<tr>
<td>- Gomez-Perez (2001)</td>
<td></td>
</tr>
<tr>
<td>- Gilbert &amp; Williams (2008)</td>
<td></td>
</tr>
<tr>
<td><strong>assessment by humans</strong></td>
<td>target ontology is compared against a predefined set of criteria that reflect the needs of a user community as defined by those criteria</td>
</tr>
<tr>
<td>- Xu &amp; Ma (2008)</td>
<td></td>
</tr>
<tr>
<td><strong>framework</strong></td>
<td>target ontology is evaluated following a predefined process and set of criteria, and helps to produce a selection of an ontology for a specific user purpose</td>
</tr>
<tr>
<td>- Fernandez et al. (2006)</td>
<td></td>
</tr>
<tr>
<td>- Strasunskas &amp; Tomassen, (2007)</td>
<td></td>
</tr>
<tr>
<td>- Strasunskas &amp; Tomassen, (2008)</td>
<td></td>
</tr>
<tr>
<td>- Gangemi et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>- Gilbert &amp; Williams (2008)</td>
<td></td>
</tr>
<tr>
<td><strong>task-based</strong></td>
<td>target ontology is evaluated based on its competency in completing tasks</td>
</tr>
<tr>
<td>- Strasunskas &amp; Tomassen, (2007)</td>
<td></td>
</tr>
<tr>
<td>- Strasunskas &amp; Tomassen, (2008)</td>
<td></td>
</tr>
<tr>
<td>- Gilbert &amp; Williams (2008)</td>
<td></td>
</tr>
<tr>
<td>- Yu et al. (2007)</td>
<td></td>
</tr>
</tbody>
</table>

2.8.2.1 Gold-standard approach

In gold-standard evaluation a newly created or newly found ontology is compared against an existing ontology that is considered to be the benchmark. The reason for an existing ontology to be considered a benchmark is because it is widely accepted by a community; because it was developed by a recognized authority in a domain; or because it already performs well for some particular purpose. The comparison between the new ontology and the existing benchmark ontology is performed by measuring the similarity
between the two ontologies. The new ontology is considered good if it is highly similar to the existing benchmark ontology.

(Maedche & Staab, 2002) propose to measure this similarity in two ways: a lexical comparison and a conceptual comparison. In the lexical comparison they measure the similarity of the actual concept strings (e.g. TopHotel and Top_Hotel). The fewer changes it takes to change one string to match the other, the higher their similarity. In the conceptual comparison they measure the taxonomy overlap and the relation overlap. The more the structure of the concept hierarchy and other relationships of one ontology matches the structure of the other, the higher their similarity.

(Brank et. al., 2006) propose to measure this similarity by using the instances of the ontology concepts. The more the assignment of the instances (actual objects) for a concept (class) matches between the two ontologies, the more similar the two ontologies are. The new ontology is considered good if it is highly similar to the existing benchmark ontology.

2.8.2.2 Data-based approach

Brewster et al look at the fit between an ontology and the domain of knowledge it represents by defining a conditional probability measure based on identity of keywords, number of possible query expansion terms, and a mapping of ontology to the terms in the domain texts. Structural fit is defined as the overlap of ontology concepts against a text corpus. Two approaches are proposed, one using a vector space representation – for both the corpus and the ontologies, comparing the similarity by the closeness of the vectors. The second using a more sophisticated architecture with a keyword component where keywords are found by Latent Semantic Analysis; a query expansion component to find
hyponyms in text using WordNet; and a mapping component that maps the ontology to the corpus. A count of how many concept terms in the ontology match those lexical items that have been marked up using the method above is then made and used as a measure of fit (Brewster et al., 2004).

2.8.2.3 Criteria-based approach

Gruninger and Fox judge and ontology based on a criterion of competency. Problems give rise to questions that an ontology must be able to answer. This is the competency of the ontology. They recommend defining the competency questions in a stratified manner so that answers to lower level questions are required to answer higher level questions. A planning and scheduling competency question example is: “What sequence of activities must be completed to achieve some goal? At what times must these activities be initiated and terminated? (Gruninger & Fox, 1995)” Differences between ontologies are reflected in the way they are able to solve the formal competency questions. Formal competency questions are informal competency questions represented through predicates in first order logic. Formal competency questions are used to evaluate the completeness of sets of axioms of the ontology for any axiomatization. An axiomatization is the act of defining axioms that specify the definition of terms and constraints on their interpretation. If two axiomatizations can answer the same set of competency questions, they are equivalent (Gruninger & Fox, 1995).

Gomez-Perez verifies an ontology based on three criteria: verification of the architecture based on soundness, verification of the lexicon and syntax based on correctness, and verification of the content based on consistency, completeness, conciseness, expandability, and sensitiveness. Architecture is verified based on the
structure of an ontology developed following the principles of design of the environment in which the ontology is included. Lexicon and syntax are verified through a tool; with a scanner to check the lexical structure of expressions, and a parser to detect that the syntactic structure is correct. Content is verified based on consistency which refers to whether it is possible to obtain contradictory conclusions from valid input data; completeness which really shows incompleteness of individual definitions in the ontology, and missing definitions from the ontology which makes the entire ontology incomplete; conciseness which refers to whether all the information gathered in the ontology is useful and precise; expandability which refers to the required effort to add new definitions or the effort to add new information to a definition without altering the set of well-defined properties; and sensitiveness which refers to how small changes in a definition alter the set of well-defined properties that are already guaranteed (Gómez-Pérez, 2001).

2.8.2.4 Assessment by humans

Xu and Ma propose a prototype democratic ranking Web system, and separate the reviewers into groups of domain experts, ontology researchers, and common users. Their system uses the assessment by humans approach that helps users subjectively evaluate ontologies uploaded by other users (Xu & Ma, 2008). Because the utility of ontologies depends on the task and is subjective, the review should include a wider community, rather than a possibly biased opinion of several hand-selected reviewers, which makes the review a democratic process (Musen, 2007) (Noy et al., 2005).
2.8.2.5 Framework

Guarino and Welty propose a formal OntoClean methodology that evaluates correctness of the ontology based on essence, rigidity, identity and unity. An essential property of an entity is a property that must hold true for an example of the entity. Rigidity, related to essence, is a property that must hold true for all the instances of an entity. Identity refers to the property of recognizable distinct individual entities, while unity refers to the property of a recognizable group of parts that form an individual entity. The “OntoClean method is based on general ontological notions drawn from philosophical ontology, expressed using metaproperties that describe the classes, properties, and relations (Guarino & Welty, 2002).” This methodology was automated by a tool called AEON (Automatic Evaluation of Ontologies) developed by Völker, Vrandecic, and Sure. AEON automatically tags ontology concepts with OntoClean metaproperties of essence, rigidity (which they call dependence), identity and unity. The tool uses lexico-syntactic patterns mined from the Web to obtain evidence of the four metaproperties above and tags the ontology concepts with one of the metaproperties along with positive or negative support based on the set of patterns (Völker et al., 2005).

Sleeman and Reul use a definition based approach in CleanONTO where definitions are paths from the concept to the root node of the ontology. CleanONTO proceeds through three phases: First, each of the concepts which occur in the unclean ontology was looked up in the online version of WordNet by an investigator who then reported each of the paths. Second, an analysis of the ontology is made and all links which are inconsistent according to a specified definition are broken. Third, orphan nodes (nodes without any
relationship connections in the ontology) and subtrees are placed back onto the tree so that the consistency rules are not violated (Sleeman & Reul, 2006).

2.8.2.6 Task-based approach

Task-based evaluation is based on the performance of the ontology when it is used to support a specific task. A recent research direction for evaluation has looked at equating characteristics to measures that can predict performance in a task (Yu et al., 2007). The researchers looked at tangledness, breadth, depth, and fanout within the browsing task. Browsing was defined as “being able to locate information using the category structure for articles for a given information need (Yu et al., 2007).” Users were given three tasks with different levels of specificity from broad to narrow as to the required information needed to answer during the task, and were measured as to the duration and correctness of finding information for the specified task.

Porzel and Malaka propose a three level evaluation framework tested by looking at concept tagging, specifically using error rates proposed in (Jurafsky & Martin, 1991) to evaluate how well the ontology performs (Porzel & Malaka, 2004) (Porzel & Malaka, 2005).

Strasunskas and Tomassen use precision and recall in a three level evaluation framework, structured similarly to the work done by Porzel and Malaka (2004) to evaluate an ontology in the Web Search task (Strasunskas & Tomassen, 2008).
2.9 Ontology Application

2.9.1 Ontology Supported Tasks

Ontologies have been used to support many application tasks such as classification (Cheng et al., 2003), annotation (Vargas-Vera et al., 2002), search (Rodrigo et al., 2005) (Gilardoni et al., 2005) (Kerschberg et al., 2004), mediation (Muthaiyah & Kerschberg, 2007), browsing (Collins et al., 2005), and personalization (Chen, 2008).

Classification based on semantics was implemented by Cheng et al. by using a signature ontology to determine user context, and then using a category ontology to classify documents into categories based on the signature. They see the ontology as a context model of some application domain that defines the meaning of the vocabulary of the domain according to the user. Thus, classification using an ontology would help to properly sort documents into the correct conceptual areas based on the user perspective (Cheng et al., 2003).

Annotation of web resources with semantic markup was performed using MnM, a tool that provides automated and semi-automated support for this task (Vargas-Vera et al., 2002). The MnM tool marks up Web page text with a set of tags defined in the ontology in order to facilitate semantic retrieval and personalization services for a Web based newsletter that would otherwise be unavailable using a standard Web presentation.

Ontology enabled support for the search task was implemented by several researchers. A Semantic search engine was built for the Royal Institute Elcano (Real Instituto Elcano) in Spain that retrieves and presents semantically marked up documents using the ontology hierarchy and relationships, and includes links that allow users to navigate between the documents and the ontology (Rodrigo et al., 2005). LKMS (Legal
Knowledge Management System) is a collaborative web-based system that supports document management in law firms and uses a legal ontology to help search for legal documents by interpreting natural language queries pertinent to a legal case (Gilardoni et al., 2005). The Knowledge Sifter framework and application allows for ontology-guided semantic search and ranking of relevant image data from heterogeneous databases (Kerschberg et al., 2004). Ontology supported search enables systems to not only provide better results by retrieving more relevant resources due to correct semantic definition of terms, but also allows for the documents to be connected to additional resources based on the semantic relationships in the ontology. For example, the LKMS system hyperlinks case document annotations to relevant information about a particular article of law or useful information about juridical concepts. The links are derived from the semantic layer on the fly, not explicitly defined beforehand (Gilardoni et al., 2005).

Ontology was also used for mediation of security policies between virtual organizations (VO) and real organizations (RO) where a mapping between the two ontologies defined a Security Policy Domain Model (SPDM) and allowed the VO and RO to share one common security policy (Muthaiyah & Kerschberg, 2007). Here, the ontology is used to mitigate interoperability problems due to heterogeneity in security policy data among various virtual and real organizations. Semantic mapping using ontologies allows mapping of dynamic systems as compared to direct mapping of data elements in heterogeneous static systems of the past (Muthaiyah & Kerschberg, 2007).

The Bletchley Park Text system was built to support the browsing of digital collections and to promote free-choice learning for the Bletchley Park museum using an ontology of collection related concepts to help search collections and present customized
content based on user’s selection of concepts (Collins et al., 2005). Personalization of content for e-learning, developed by Chen, utilizes an ontology structure that acts as a structured knowledge representation scheme to illustrate conceptual relationships between course materials (Chen, 2008). Both systems above utilize the ontology for its conceptualization — similar to a concept map — to structure the presentation of content and links to additional resources in such a way as to make the relationships between resources clear and transparent to the user.

2.9.2 Ontology Domains

The domains where ontologies are used are also as varied as the application tasks. MediaCaddy is a recommendation engine that uses an ontology to support user research and exploration of meta-content (content about content e.g. entertainment news, articles, reviews, interviews, trailers, etc.) leading them to find desired content (e.g. music, movies, TV programs, etc.) in entertainment domain content portals (Garg et al., 2005). Collins et al. built a Bletchley Park Text system that allows a museum visitor to extend their museum experience by providing a visitor with online access to archived resources tailored to their interests. The knowledge level description of museum resources that supports the Bletchley Park Text system was created using three ontologies: the CIDOC Conceptual Reference Model (CRM), a Story and Narrative ontology, and a Bletchley Park domain ontology (Collins et al., 2005). Filtering implemented for personalization of dynamic user interfaces based on context and the information the system has about a user to simplify Web services was developed for a mobile device. Khushraj and Lassila extend the OWL-S ontology used for semantic annotation of Web service descriptions to control the mobile phone UI display (Khushraj & Lassila, 2005). Maidel et al.
implemented an ontology-based method for ranking the relevancy of items in the electronic newspaper domain for a personalized electronic newspaper project called ePaper. ePaper uses the IPTC NewsCodes ontology hierarchy to provide content filtering of news items. An ontology based profile containing concepts was built for news items and the user, and similarity between the two profiles is determined by the distance in the hierarchy of the concepts in the two profiles (Maidel et al., 2008). García and Sicilia use ontologies to aid in information seeking by supporting search tactics such as super, sub, relate, contrary, etc. described by Marcia J. Bates (1990) and implement a system design that allows for a concrete stepwise refinement of the search from general to specific, yet allows for the combination of search terms using arbitrary relations (García & Sicilia, 2003).

2.10 Context

Context has always been an important aspect of information and software systems development, and is just as appropriate for the Semantic Web. Ontologies have to incorporate contextual information in engineering and evaluation. Context can be defined by location, identities of nearby people and objects, and changes to those objects (Schilit & Theimer, 1994); the time of day (Brown, Bovey, & Chen, 1997); environment, identity (Ryan, Pascoe, & Morse, 1997); orientation, and date (Dey, 1998). In general, the above list who, where, what, and when – or identity, location, activity, and time – which are the most important aspects of context for designers and are called context types (Dey & Abowd, 1999). Context types are the most general definition of context and are not measurable. User (identity), application task (activity), and domain (location) are context parameters that represent the context types. They are more specific and used in
our method descriptively because they can link context to ontology characteristics. Context variables vary the context parameters, are the most specific, measurable and used in our method to distinguish different configurations of context. For example, the user expertise variable (i.e., novice and expert) has an impact on both ontology design and use (Vargas-Vera et al., 2002). It plays a role in the interaction of the user with the system, especially if the structure of the ontology is in some way visible to the user through the application, as a hierarchy (Vargas-Vera et al., 2002).

2.11 Context and Ontologies

2.11.1 Changing and Adapting Ontologies

The domain, application task, domain experts, and ontology engineers impact the conceptualization (Russ et al., 1999). For example, within the scope of the DARPA Joint Forces Air Component Commander (JFACC) program the authors discovered differences between two ontologies for the air campaign planning domain. The Aircraft ontology classified airplane type by the missions the aircraft could perform while the INSPECT ontology did not follow the same classification structure (Russ et al., 1999). Ontology differences in the security policy domain were shown in the authentication policy structure where the virtual organization ontology had “UserPassword” and the real organization ontology had “Token”, while other classifications in the ontology trees remained the same (Muthaiyah & Kerschberg, 2007). It is difficult for ontology engineers in most cases to adapt an ontology to another domain. The examples above illustrate the issue where an ontological commitment has been made and all other relationships and axioms reinforce the concept definitions.
2.11.2 Ontology Reuse and Evolution

The engineering and application of multiple ontologies across the varied application tasks and domains has made it necessary to provide systems that list, find, and to some extent rate existing ontologies. Researchers have begun to develop such systems. There are a number of currently available Web sites that provide a listing or a keyword enabled search of registered or submitted ontologies: Swoogle, OntoSearch, Protégé Ontology Library, OBO Foundry, etc. (see Appendix B: Ontology Resources for examples). However, issues of ontology reuse have cropped up because searches of these databases or review of the links are based on keywords. A Swoogle search for the keyword “anatomy” produced 59 hits and the user was left to examine the 59 XML ontology files to determine how well each ontology matched the user’s domain conceptualization (Lewen et al., 2006).

Knowledge Zone is a web-based ontology repository that utilizes the Open Rating System approach extended with topic-specific trust to provide a personalized ontology ranking to users based on peer-reviews (Lewen et al., 2006). AKTiveRank is a prototype system that ranks ontologies based on the analysis of their structural characteristics expressed by four measures: the semantic similarity measure (SSM), density measure (DEM), class match measure (CMM), and the betweenness measure (BEM) and orders the ontologies with respect to a query submitted by the user (Alani & Brewster, 2006). However, all of the Web sites and tools mentioned above do not incorporate specific guidance that provides ontology engineers with a way to determine what kind of characteristics an ontology should have for a specific context, rather focusing on keywords, metadata, and user feedback.
Russ et al. also highlight some of the challenges of ontology reuse. They encountered problems of reuse related to the differences in representation of a publicly available ontology of time (Hama et al., 1992) due to differences between Ontolingua (the time ontology’s representation language) and Loom (the representation language used by Russ et al.) because of a mismatch of modeling styles and inference engine bias (Russ et al., 1999). Additionally, the merging of two aircraft ontologies (Aircraft ontology and INSPECT ontology) into one JFACC ontology revealed problems where the aircraft ontologies overlapped. There were philosophical, naming, and structural differences that caused merging issues. For example, in the Aircraft ontology the classification of fighter or bomber was determined by the missions the aircraft could perform. In the INSPECT ontology the classification of fighter or bomber was not determined by mission. The modeling of domain objects as concepts or instances present problems for reuse because they can be domain or application task specific. Kalfoglou and Schorlemmer present an overview of ontology mapping, and possible solutions to mapping and merging issues through heuristics and other techniques (Kalfoglou & Schorlemmer, 2003). Kotis et al. use Latent Semantic Indexing as part of a six step algorithm to create intermediate ontologies and use WordNet to map between the concepts before the merge (Kotis et al., 2006).

Another aspect of ontology use relates to ontology evolution. Ontology evolution is the adaptation of an ontology to changing requirements, ontology instances, and patterns of use of an ontology-based application (Stojavovic et al., 2002). Haase et al. introduced a collaborative filtering recommender system that assists users in managing the evolution of personal ontology as changing interests and data change the domain in the Bibster
semantics-based peer-to-peer bibliography sharing application (Haase et al., 2005). CRAFT (Collaborative Reasoning and Analysis Framework and Toolkit) allows non-knowledge engineering expert users to collaboratively evolve an ontology while performing their daily work (application tasks) (Liu & Gruen, 2008). In their study Liu and Gruen allowed users to extend an ontology, including classes, properties, links and subclasses, within the scope of their work process. The evolved ontologies were evaluated by using relationship richness and inheritance richness, among other evaluation measures; and the results showed horizontal ontologies with high inheritance richness and tangledness. This demonstrated that organic user controlled evolution of ontologies tends to create basic level classes without abstract superclasses or more specific subclasses (Liu & Gruen, 2008).

2.12 Information Seeking

An information need arises from an inadequate state of knowledge of a user about a topic or situation (Belkin et. al., 1982). The information need leads the user to engage in information seeking behavior by placing demands on information systems or other resources to satisfy that need. Figure 2-3 shows the concepts related to information seeking behavior (Wilson, 2006). If we look at the information search process divided into the levels of need, it consists of visceral, conscious, formalized, and compromised levels (Taylor, 1962)( Kuhlthau, 1991). The visceral level is described as the “actual, but unexpressed need for information”; the conscious level is described as the “conscious within-brain description of the need”; the formalized level is described as a “formal statement of the question”; and the compromised level is described as the “question as presented to the information system (Taylor, 1962). “The representation of an
information need, or the query, is an expression of what the user does not know but has to define to an information retrieval system to fulfill that need (Taylor, 1962; Belkin et al., 1982). Search strategies are based, in general, on a comparison between terms in a query and terms in the stored documents, thus a document is relevant to an information need if it contains at least one sentence which is relevant to that need (van Rijsbergen, 1979). Figure 2-3 shows the concepts related to information seeking behavior (Wilson, 2006).

2.12.1 Search

The experiment utilizes the search application task. Search is described as directed entry of keywords related to an information need into a system that returns a list of results from which the users can decide to select one to several results, or refine the search query again (Baeza-Yates & Ribeiro-Neto, 1999).
Variables in the information system influence the formation of the question, and thus influence the way the user interacts with the system. These system organization variables are divided into five groups: general aspects, system input, internal organization, question input, and output (Taylor, 1962). Internal organization and question input, in particular, are important to this research. Internal organization has to do with “classification, indexing, subject-headings, and similar access schemes”; while question input has to do with the “part human operators play in the system” where the success of information retrieval and query formation will depend on the translation of the information need into a query (Taylor, 1962).

2.12.2 Information Retrieval

The information retrieval process begins when a user submits a query into the information system (Frakes, 1992). At the system level, a query – or search string – is a formal statement of an information need, and does not uniquely identify a single object in the collection of objects in the information system. Objects matching the query, with different degrees of relevance, will be returned and ranked by the information system according to a score that defines how well the returned objects match the query. This process can be iterated if the user refines the query (Frakes, 1992). In this research, the query is submitted to the guidelines database from the National Guideline Clearinghouse (NGC) where the guidelines are indexed by a lightweight ontology based on the pruned Medical Subject Headings (MeSH) hierarchy focusing on Internal Medicine. The guideline database returns several guidelines from the collection. The relevant documents are known a priori and should be returned by the system given the correct
selection of search terms from the lightweight ontology based on the assigned information need.

2.12.3 Search and Expertise

Expertise has an impact on the search process, especially on query formulation and search strategy. Experts in a domain use and recognize more technical terms than novices (Liu & Wacholder, 2008; White et al., 2009). In their study on using MeSH terms for biomedical search, Liu and Wacholder (2008) found that experts achieved better precision than novice users by using additional MeSH terms. The search topics presented to the expert and novice participants in their study were specific to the Genomics domain. Terms from MeSH, which is a specialized medical taxonomy, were made available to participants to augment their general query terms. Experts achieved better precision than novices because they were able to recognize and select the correct domain-specific MeSH terms to add to general query. Likewise, White et al. (2009) found that experts created longer queries that contained more vocabulary from the domain-specific lexicon and that the queries were more technically-sophisticated.

Additionally, Liu and Wacholder (2008) found that domain experts spent significantly more time searching for a specific topic than domain novices. White et al. (2009) found similar results where domain experts consistently visited more pages per session, spent more time on the topic, and issued more queries. The extra time was used by experts to carefully consider more content and examine more possible branches when searching for results in their domain. In general both studies found that domain expertise contributed to better use of content and more effective use of technical terms in retrieval,
especially when the topics required specialized knowledge. The use of technical terms in retrieval by experts increased precision leading to better search results.

2.13 Semantic Web

Ontologies play an important part in the Semantic Web. Figure 2-4 shows the Semantic Web Stack (Semantic Web - XML2000, 2008) that depicts the required architecture, composed of many technologies that must all be implemented in order to make the Semantic Web successful. The technologies are a metadata based infrastructure that extends the current Web to make its content meaningful for computers. Ontology forms an important layer that supports software agent access to formally structured and defined collections of information that will support automated understanding by defining the concepts and their relationships for a domain for which the specific Web page or Web information resource has been built. The Semantic Web will provide enhanced capability where autonomous software agents can carry out more sophisticated tasks for their owners utilizing semantically defined Web page content and services.
The next chapter describes the proposed methodology to investigate the relationship between ontology characteristics and context parameters. First, the methodology is explained in terms of the three main elements required to determine the quality of the application task for the combination of ontology characteristics and context parameters. Next, the methodology is explained in terms of the steps across the two phases, with examples for each step where appropriate. The possible changes to the ontology characteristics, which are an important aspect of the methodology, are described using an example of measures from 2.5.2 Measures. Finally, the general hypothesis tested by each implementation of the methodology is introduced.
CHAPTER 3: Methodology

The methodology associates ontology characteristics with context parameters in order to evaluate their combined performance in context. It is implemented using the performance of an ontology-supported application task as its overall quality metric. The quality metric should be objective and selected to measure the performance based on the application task.

The methodology changes selected characteristics of an ontology to observe the impact on the final quality of the performed task. Within the scope of the methodology, changes to the ontology have to impact the selected ontology characteristics based on the context parameter being evaluated. The method is intended to work with any ontology characteristics. The change is made between the two phases of the methodology: Phase 1 – unchanged ontology before the changes are made (BC) and Phase 2 – changed ontology after the changes are made (AC).

3.1 Elements of the Methodology

The elements of the methodology are a specific context $c_i$ in a set of contexts $C$, selected characteristics $r_j$ in a set of characteristics $R$, and an application task with performance $p$. The methodology can be summarized as asking:

For a context $c_i$, does a variation in the value of one or more characteristics $r_j$ lead to a variation in performance from $p_{BC}$ to $p_{AC}$?

Performance $p$ should be measured by an objective quality metric. The assumption is that, if the variations are statistically significant, the relationship
between the examined characteristics $r_j$ can be used to guide ontology engineering for the examined context $c_i$. Figure 3-1 shows the elements of the methodology.

![Figure 3-1: Change to elements of the methodology between two phases]

3.1.1  Context Parameters

The context parameters describe context ($c$) which is operationalized as a user performing a specific application task within a domain.

3.1.1.1  User

The user parameter describes the actor in the context. The user parameter is defined by variables that can classify the users for a desired differentiation. For example, in the healthcare context the resident physicians – actors – are defined by their range of
expertise, making expertise and its values the defining variable for the user parameter. The expertise variable can have the values of expert and novice. Each value can be clearly defined to make the desired differentiation for the parameter.

### 3.1.1.2 Application Task

The application task parameter describes the action performed by the actor in the context. For example, in the healthcare context resident physicians can browse medical topics, search (requiring information retrieval or query expansion) for medical information, or classify medical content. The application task itself is not differentiated by a variable. The application task is measured by a metric with performance $p$.

### 3.1.1.3 Domain

The domain parameter describes the location where the actor performs the action. The domain does not have to represent a physical location, but can be any a set of objects or area of activity defined by a boundary. The ontology used to support an application task is a conceptual representation of the domain. The domain is not differentiated by a variable. The domain is differentiated by the boundary definition.

### 3.1.1.4 Time Context Parameter

The remaining context parameter is time. The time parameter describes the temporal location of the action. This research does not incorporate the measurement of performance over time. However, the time context parameter can be added to the methodology to complete the context description.
3.1.2 Ontology Characteristics

As stated previously in Section 2.5 ontology characteristics describe an ontology. Characteristics \((r)\) are expressed through criteria and measures whose values change based on the change to the ontology. The values are required in the methodology for determining the eventual changes to be made to the ontology based on performance \(p\) of the application task.

Currently, the candidate set of criteria includes approximately 22 different criteria proposed by six different approaches as shown in Table 2-1. The candidate set of measures includes approximately 30 different measures proposed by five different approaches as shown in Table 2-2. All of the measures are operationalized, or able to be calculated for an ontology.

In general characteristics can be seen as structural and non-structural. Structural characteristics are based on the physical dimensions of the ontology while non-structural characteristics are the conceptual and user defined characteristics. Both types of characteristics were introduced in Section 2.5 Ontology Characteristics.

3.1.2.1 Structural Characteristics

Structural characteristics are based on the physical dimensions of the ontology schema, and are thus internal to the ontology. Structural characteristics have several important aspects. They are common to many of the specification structures and ontologies on the ontology specification spectrum. They are easier to calculate than non-structural characteristics because they can be determined from the ontology schema which is easily accessible. They are easier to understand because they are based on visible and objective properties.
### Figure 3-2: Excerpt of the unchanged MeSH ontology

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial Infections and Mycoses</td>
<td>Bacterial Infections</td>
<td>Gram-Negative Bacterial Infections</td>
<td>Bordetella Infections</td>
<td>Whooping Cough</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Enterobacteriaceae Infections</td>
<td>Escherichia coli Infections</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Granuloma Inguinale</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salmonella Infections</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Salmonella Food Poisoning</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tick-Borne Diseases</td>
<td>Tularemia</td>
</tr>
</tbody>
</table>

Figure 3-2 is an excerpt of the MeSH category structure. Examples for the structural characteristics of depth, breadth, and fanoutness are calculated based on Figure 3-2.

Depth is calculated as the average of the sum of all is-a paths starting at the top node and terminating in a leaf node. Breadth is calculated as the sum of all is-a edges per level divided by the number of levels. Fanoutness is calculated as the total sum of the average of is-a edges per node divided by the number of levels.

- **Depth**: \( 4 + 4 + 4 + 5 + 4 = 21/5 = 4.2 \)
- **Breadth**: \( 1 + 1 + 3 + 5 + 1 = 11/5 = 2.2 \)
- **Fanoutness**: \( (1/1 + 1/1 + 3/1 + (1 + 3 + 1)/3 + 1/1)/5 = 1.5 \)

#### 3.1.2.2 Non-structural Characteristics

Non-structural characteristics are based on the ontological properties, or external user specified hurdles the ontology must meet. The example provided previously for the ontological dimension was that of essence in the OntoClean approach (Guarino & Welty, 2002). Hardness is an essential property of hammers, whereas it is not an essential property of sponges because a sponge can be hard when dry and soft when wet. A definition of hardness as an essential property of sponges would then violate the
ontological notion of the essence metaproperty. An example of a user specified hurdle is Completeness specified by Gomez-Perez (1996). Completeness is defined as an ontology having the required conceptualization explicitly set out or able to be inferred through definitions and axioms, and the concept definition as being correct through a comparison with competency questions or external resources in the real world.

Non-structural characteristics were not examined in this research. The methodology, however, does apply to non-structural characteristics as well. An examination of non-structural characteristics will be implemented in future work.

3.1.3 Objective Quality Metrics

The combined performance of the ontology characteristics and context parameters in the application task should be measured by objective quality metrics. The metrics vary depending on the application task and selected parameter. The metrics should be precise, accurate, reliable, and valid.

3.2 Methodology Steps

For a given context, with ontology characteristics and context parameters defined, the methodology is implemented in two phases. Table 3-1 describes the two phases of the methodology performed for a context.
Table 3-1: Phases of the methodology

<table>
<thead>
<tr>
<th>Methodology Phase</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Measure the performance of the application task with an <em>unchanged ontology</em> as a baseline.</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Measure the performance of the application task with a <em>changed ontology</em> to determine if there is a difference.</td>
</tr>
</tbody>
</table>

Figure 3-3 shows the details of the elements for the two phases of the methodology performed for a context. In Phase 1 the context, characteristics, and metrics are defined and performance measured. In Phase 2 the ontology characteristics are changed and performance is measured and analyzed. The following sections will detail the exact steps in each of the two phases of the methodology.

Figure 3-3: General application of the methodology
3.2.1 Phase 1

Measure the performance of the application task with an unchanged ontology as a baseline.

3.2.1.1 Select

The Select step populates and documents the elements in the methodology. The step populates the elements of the methodology by defining the initial values for $c$ and $r$ (see Section 3.1 Elements of the Methodology). It does this by defining the characteristics to use, and determines a range of values.

- **First, define the selected context parameter to the variable level and document the remaining parameters that define context.** For example, the identity context type is expressed by the User parameter, which is defined by the expertise variable set to the values of novice or expert.

- **Second, select the ontology characteristic(s) for performance measurement.** For example, structural characteristics expressed as measures of depth or breadth which are integral to the search application task where the ontology is a visualized hierarchy the users navigate.

- **Third, select change to the characteristic(s).** The change should be based on the selected context parameter. For example, in this research the selected context parameter is users defined by the expertise variable. The application task is search supported by a visualization of the ontology structure. The change to the characteristics was selected that should have an impact on both the expertise variable and the structural hierarchical characteristics. The selected change
makes it difficult for users to complete the search task depending on their level of expertise and alters the selected ontology characteristics at the same time.

- **Forth, select metrics to use for performance validation.** The metrics depend on the nature of the application task. For example, in this research the application task is searching for medical guidelines supported by a visualization of the ontology hierarchy. We selected two measures that, if performed correctly, would result in a successful completion of the task. The number of users to test is determined by a sample size estimation formula based on prior percent (%) change in similar characteristics changes to calculate statistical significance, or start with an estimate based on the size of the final user population.

### 3.2.1.2 Measure Initial Performance

The initial performance \( p_{BC} \) of the application task supported by the ontology **before the change** is measured.

### 3.2.2 Phase 2

Measure the performance of the application task with a changed ontology to determine if there is a difference.

#### 3.2.2.1 Change Values

The ontology characteristic(s) values are changed based on the method in the third Select step.

#### 3.2.2.2 Measure Resultant Performance

The resultant performance \( p_{AC} \) of the application task supported by the ontology **after the change** is measured.
3.2.2.3 **Analyze**

The difference in results for performance $p$ for the application task for the selected characteristics $r_j$ in the defined context $c_i$ is analyzed. A difference in performance indicates that the change to the ontology characteristics values had an impact. The difference should be tested for statistical significance.

3.2.2.4 **Save and Apply**

If the results show a statistical significance, the specific combination of context $c_i$ and selected characteristics $r_j$ can be documented for future reference and used as a guideline for ontology engineering and evaluation. The approaches that utilize the ontology characteristic(s) that show a statistical difference can be selected for an ontology engineering and evaluation process in the same or a similar context.

3.3 **Changing Ontology Characteristics in the Methodology**

Within the scope of this research, changes to the ontology have to impact the ontology characteristics used in the engineering and evaluation approaches. In principle, the method works with any criteria or measure. Because context is operationalized as a type of user performing a specific application task within a domain, when the ontology is changed, the changes have to be made so that they affect the user, task, or both user and task context parameters. For example, changing the depth measure of the ontology at a level that users are not required to traverse to perform a search application task will not allow for proper validation of the depth measure. The focus on a particular domain by the ontology cannot be changed as this would create a completely different conceptualization and resulting ontology. Only ontology characteristics that impact the context parameters should be selected and changed. If ontology characteristics are
changed that have no impact on the user or application task, no measurements can be made, and the difference in the performance $p$ for the application task and the selected characteristics $r_j$ cannot be validated for the defined context $c_i$. The quality of the change cannot be determined a priori. The quality will only be determined after performance $p$ is measured and tested for statistical significance. However, the quality can be estimated if it meets the above suggestion of having an impact the user and task.

Some changes to the selected characteristics $r_j$ can only be made in aggregate. For example, structural characteristics are necessarily interrelated for the same measure type (e.g. depth and breadth); therefore a change in one structural characteristic will force a change in another characteristic. It is important to verify if changing a value for one characteristic will impact the computation of another characteristic. Changes to characteristics across measure types, shown in Table 2-2 (e.g. structural, schema, and instance), will not impact computation of other characteristics.

3.4 General Hypothesis in the Methodology

The methodology developed for this research is framed in such a way that we test the general hypothesis:

$H$: A change in the selected ontology characteristics $r_j$ has an impact on the performance $p$ of the application task for defined context $c_i$.

The general hypothesis is formulated based on an expected impact of the change on the performance of the application task.

Chapter 4 will describe in detail how to apply the methodology presented in Chapter 3 in a specific context. The example is an implementation of the methodology in the healthcare context.
CHAPTER 4: Studies and Results

This chapter describes the application of the methodology explained in Chapter 3 in two studies with four participant groups in the healthcare context. The chapter outlines the procedure for the studies at a high level, presents the elements of the methodology as applied to the studies in the context, and describes how the methodology steps were employed in the implementation of the studies. The chapter also describes changes to the selected ontology characteristics and how these changes affect the category structure. Consequently, changes to the ontology category structure impact performance of a search task dependent on the expertise of resident physicians and nursing students. Finally, the chapter presents hypotheses and results of two studies conducted with resident physicians and nursing students.

4.1 Study Description: Applying the Methodology in a Healthcare Setting

The experiment varies the user context parameter while maintaining a constant application task and domain context parameters in order to measure the combined performance for selected ontology characteristics. The user (identity) context parameter varies based on expertise. The application task (activity) context parameter is search. The domain context parameter is medicine, specifically medical guidelines from the National Guideline Clearinghouse specified by a lightweight ontology based on the pruned Medical Subject Headings (MeSH) hierarchy focusing on Internal Medicine. For the characteristics to be tested, two types of users perform the application task.
The experiment is conducted in two phases as described in Table 3-1. Phase 1 measures performance of the application task supported by an existing unchanged ontology. Phase 2 measures performance of the application task supported by a changed ontology, the characteristics of which are altered to have an impact on the user context parameter (see 4.4 Changing Ontology Characteristics in the Applied Methodology).

Figure 4-1 shows the methodology applied in the studies.

Each phase of the experiment consists of a series of studies repeated for each context utilizing the characteristics selected. Table 4-1 shows the four different context variations based on the participant group and changed user parameter based on the expertise variable.
4.2 Elements of the Methodology in the Studies

This section describes the elements of the methodology introduced in Section 3.1 as they were implemented in the two studies.

4.2.1 Context Parameters in the Studies

4.2.1.1 User

The methodology was applied in two studies. The studies include two groups in the healthcare field with distinct levels of expertise. The first study consisted of participating resident physicians in Internal Medicine programs in the United States. The second study consisted of participating nursing students in the Drexel University College of Nursing and Health Professions. Table 4-2 shows the total participation per subgroup.

<table>
<thead>
<tr>
<th>Participant Subgroup</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study 1 – Resident Physicians</strong></td>
<td></td>
</tr>
<tr>
<td>Novice resident physician</td>
<td>13</td>
</tr>
<tr>
<td>Expert resident physician</td>
<td>16</td>
</tr>
<tr>
<td><strong>Study 2 – Nursing Students</strong></td>
<td></td>
</tr>
<tr>
<td>Novice nursing student</td>
<td>10</td>
</tr>
<tr>
<td>Expert nursing student</td>
<td>17</td>
</tr>
</tbody>
</table>
The user context parameter was differentiated by expertise. The first participant group was resident physicians in internal medicine residency programs. The resident physicians were divided into two subgroups of novice and expert. A novice resident is a physician within the first or second year of a residency program. An expert is a resident in the third or final year of a residency program, or a chief resident (resident in an additional fourth year of a residency program).

The second participant group was nursing students enrolled in undergraduate and graduate programs at the Drexel University College of Nursing and Health Professions. The nursing students were divided into two subgroups of novice and expert. A novice nurse is a nursing student within one of the tracks of the Accelerated (ACE) B.S.N or B.S.N Co-op. program. An expert nurse is a nursing student within one of the tracks of the M.S.N program.

4.2.1.1.1 User Context Parameter Defined by the Expertise Variable

Expertise was selected as the variable for the user context parameter based on research in the healthcare field with resident physician and nursing populations. The expertise variable was represented by the values of *novice* and *expert*. The definition of expertise and the resulting values differ in the two groups selected for the studies. This section describes research specifically related to expertise for resident physicians and nursing students. In general the studies show that duration of training, especially long term deliberate practice under supervised instruction, enhances the expertise of both physicians and nurses. This section also describes the results from the post-task expertise validation surveys administered to the two groups after they completed Phase1 and Phase 2 in the studies.
4.2.1.1.1 Resident Physician Expertise

Research looking at physician understanding of updated guidelines for the diagnosis and management of asthma by the National Heart, Lung, and Blood Institute (NHLBI) determined that asthma specialty physicians scored significantly higher overall than primary care physicians, but they did not score significantly higher than asthma specialty fellows; and residents scored lower than physicians at all other levels of training (Doerschug et al., 1999). The study also found that residents had the lowest scores, but verified that resident scores improved with duration of training; suggesting that residents are appropriately learning the topic during their training (Doerschug et al., 1999). This study exemplified that continued supervised training improves expertise in a domain, specifically for residents progressing through the residency program.

The resident physicians in our study were progressing through their residency and were distributed across the three years of the Internal Medicine program. Residents in the first and second year of residency were in the early stages of supervised training and thus considered novices. Residents in the third year were in the final stages of supervised training, thus having developed the necessary skills for unsupervised practice, and considered experts.

4.2.1.1.2 Nursing Student Expertise

Research looking at the assessment of nursing competence and expertise in nursing utilizing the expert-performance approach found that diagnostic expertise improved with deliberate practice, i.e. extended supervised training with feedback; and that graduate training and specialized training can be used to determine expertise (Ericsson et al., 2007). This study exemplified that nursing expertise can be differentiated between
undergraduate and graduate programs. In addition to the research, we verified expertise
difference with faculty from the nursing school. The program supervisors confirmed that
most undergraduate students had no work experience while most graduate students had
work experience. Consequently, nursing students enrolled in undergraduate programs
were considered novices while nursing students enrolled in graduate programs were
considered experts.

4.2.1.1.2 Confirmation of Expertise

All participants in the studies completed two post-task surveys to confirm their
expertise. The post-task surveys were used to confirm the division of the resident
physician and nursing student groups into the novice and expert subgroups using the
expertise variable. The first survey was administered after Phase 1 and the second survey
was administered after Phase 2. The participants were asked survey Question 1 through
Question 4 for Phase 1 of the studies and survey Question 2 through Question 4 for Phase
2 of the studies. The survey administered after Phase 1 inquired about participant
experience with searching for guidelines in the National Guideline Clearinghouse (NGC)
and confirmed their expertise in the medical scenario presented in Phase 1. The survey
administered after Phase 2 confirmed participant expertise in the medical scenario
presented in Phase 2.

With respect to expertise, we hypothesized that:

- **Resident physician hypothesis**: Expertise in the resident physician group
can be determined by the year of training in the residency program.

- **Nursing student hypothesis**: Expertise in the nursing student group can be
determined by participation in an undergraduate and graduate program.
This section contains a description of the survey questions, how they were administered to participants in the studies, and the results. Appendix E shows the survey questions as they appeared on the Web page to participants for both Phase 1 and Phase 2. Complete text for scenarios is shown in Appendix F and the application of the scenarios is described in Section 4.2.1.2 Application Task. The answers to the post-task survey questions for all participants divided into their respective subgroups are shown in Appendix G.

4.2.1.1.2 Description and Results of Post-Task Survey Question 1

Description

<table>
<thead>
<tr>
<th>ID</th>
<th>Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Have you ever used the Agency for Healthcare Research and Quality (AHRQ) to find guidelines?</td>
</tr>
<tr>
<td>Answer Options</td>
<td>Yes, No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Q1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>If Yes, did you use Search or the Hierarchy to find guidelines?</td>
</tr>
<tr>
<td>Answer Options</td>
<td>Search, Hierarchy</td>
</tr>
</tbody>
</table>

Question 1, identified as a combination of Q1 and Q1.1, was presented to participants to determine their experience with the guidelines search feature in the National Guideline Clearinghouse (NGC) to eliminate possible bias in the results due to prior knowledge of the MeSH hierarchy.
Results

Several participants replied that they previously used the Agency for Healthcare Research and Quality (AHRQ) to find guidelines. All participants that previously accessed AHRQ guidelines used the search feature. No participants replied that they previously used the hierarchy to search for guidelines and were thus unfamiliar with its structure.

4.2.1.2.2 Description and Results of Post-Task Survey Question 2

Description

<table>
<thead>
<tr>
<th>ID</th>
<th>Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>How familiar are you with the topic described in the scenario?</td>
</tr>
<tr>
<td>Answer Options</td>
<td>Very familiar</td>
</tr>
<tr>
<td></td>
<td>Familiar</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
</tr>
<tr>
<td></td>
<td>Unfamiliar</td>
</tr>
<tr>
<td></td>
<td>Very unfamiliar</td>
</tr>
</tbody>
</table>

Question 2, identified as Q2, confirmed the expertise of participants based on their self-assessment of knowledge and familiarity with the scenarios. Responses of Very Familiar or Familiar were considered a positive indicator of expertise in the scenario topic. Responses of Neutral, Unfamiliar, or Very Unfamiliar were considered a negative indicator of expertise in the scenario topic.

Results

Table 4-3 contains the results for Question 2 that verified participant expertise based on their self-assessment.
Table 4-3: Results for Question 2, participant self-assessment of familiarity with scenario topic

<table>
<thead>
<tr>
<th></th>
<th>Novice Resident Physicians</th>
<th>Expert Resident Physicians</th>
<th>Novice Nursing Students</th>
<th>Expert Nursing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase 1</td>
<td>Phase 2</td>
<td>Phase 1</td>
<td>Phase 2</td>
</tr>
<tr>
<td>Very Unfamiliar</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Neutral</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Familiar</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Very Familiar</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Very Familiar &amp; Familiar Total</td>
<td>7</td>
<td>5</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Participant N</td>
<td>13</td>
<td>13</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Percent “Expert” Per Phase</td>
<td>53.84%</td>
<td>38.46%</td>
<td>68.75%</td>
<td>62.50%</td>
</tr>
<tr>
<td>Percent “Expert” Average for Both Phases</td>
<td>46.15%</td>
<td>65.63%</td>
<td>40.00%</td>
<td>47.06%</td>
</tr>
</tbody>
</table>

The data in Table 4-3 demonstrates that expert resident physicians were the most knowledgeable about the scenario topics based on their own self-assessment of knowledge with a 65.63% average of Very Familiar or Familiar responses, considered a positive indicator of expertise in the scenario topic, to Question 2 for Phase 1 and Phase 2. Expert resident physicians were followed by expert nursing students with an average of 47.06%, novice resident physicians with an average of 46.15%, and novice nursing students with an average of 40.00%. The Very Familiar and Familiar total and percentages per group demonstrate the relative confidence about the scenario topics. Overall, the per-phase (Phase 1 and Phase 2) confidence and combined confidence were
higher in the expert groups over novices and resident physicians over nursing school students.

The proportions for total Very Familiar/Familiar responses for Question 2 in Phase 1 and Phase 2 for novices and experts were tested for significance \( p < .05 \) using the two-tailed two-proportion test for small sample sizes. Both the resident physician (Phase 1 \( p = 0.42 \), Phase 2 \( p = 0.21 \)) and nursing student (Phase 1 \( p = 0.52 \), Phase 2 \( p = 0.95 \)) groups showed no statistically significant difference in proportions. Although the results for the two groups were not statistically significant based on the number of participants (addressed in section 5.2.2), the general differences in proportions by percentages and by the order of the percentages for the groups support the expertise hypotheses that participants with more training (i.e. expressed by years in a residency program or participation in graduate program) can be considered more expert and have more knowledge about the topics within their domain of expertise.

4.2.1.1.2.3 Description and Results of Post-Task Survey Question 3

**Description**

<table>
<thead>
<tr>
<th>ID</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Were you able to find the keywords you wanted in the hierarchy?</td>
</tr>
<tr>
<td>Answer Options</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Q3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>If you answered No, which keywords were you looking for?</td>
</tr>
<tr>
<td>Answer Options</td>
<td>Unstructured text entry</td>
</tr>
</tbody>
</table>

Question 3, identified as a combination of Q3 and Q3.1, verified participant expertise by allowing the participant to demonstrate their knowledge based on keyword selection. The participants were asked in Q3 if they were able to find the keywords they wanted in
the hierarchy. If they answered “No”, they were asked in question Q3.1 to list the keywords they were looking for in the hierarchy. If the participants were not successful in finding the correct path, but could identify the correct keywords or determine keywords extrapolated from the scenario description and their knowledge of the topic, then their expertise could still be confirmed.

**Results**

Table 4-4 contains the results for Question 3 that verified participant expertise by allowing the participant to demonstrate their knowledge based on hierarchy navigation and keyword selection.

**Table 4-4: Results for Question 3, keyword selection**

<table>
<thead>
<tr>
<th>Familiarity with Topic of Participants Answering “No”</th>
<th>Novice Resident Physicians</th>
<th>Expert Resident Physicians</th>
<th>Novice Nursing Students</th>
<th>Expert Nursing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold=Correct Keywords (Kw)</td>
<td>Phase 1</td>
<td>Phase 2</td>
<td>Phase 1</td>
<td>Phase 2</td>
</tr>
<tr>
<td>1-Unfamiliar</td>
<td>13</td>
<td>13</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>1-Familiar</td>
<td>100%</td>
<td>92.30%</td>
<td>93.75%</td>
<td>87.50%</td>
</tr>
<tr>
<td>Percent Average for Both Phases</td>
<td>96.15%</td>
<td>90.63%</td>
<td>80.00%</td>
<td>91.18%</td>
</tr>
</tbody>
</table>
A participant reply of “Yes” and a reply of “No” with correct keywords listed were counted towards the total correct. In other words, \( Replied \, “Yes” + Replied \, “No” \), \textit{But Selected Correct Keywords} were counted as correct per phase (Phase 1 and Phase 2) and then averaged to get the percent average for both phases.

The data in Table 4-4 demonstrates that novice resident physicians were the most knowledgeable about the scenarios based on keyword selection with a total of 96.15% percent average reply to Question 3 for Phase 1 and Phase 2. Novice resident physicians were followed by expert nursing students with an average of 91.18%, expert resident physicians with an average of 90.63%, and novice nursing students with an average of 80.00%. The results for this question were somewhat mixed with novice resident physicians providing better keyword selection than the two expert groups.

A review of expert resident physician group answers showed that two physicians interpreted Scenario 3 describing Pulmonary Tuberculosis as Community Acquired Pneumonia or Pneumonia in general. This fact combined with a 100% response rate to Q3 by novice resident physicians as compared to other participant groups provided the best overall resulting performance. The novice nursing students performed as expected compared with other participant groups. The physician subgroups, with more expertise in diagnosis of disease, performed better than the nursing groups.

The proportions for participant reply of “Yes” and a reply of “No” with correct keywords listed for Question 3 in Phase 1 and Phase 2 for novices and experts were tested for significance \((p < .05)\) using the two-tailed two-proportion test for small sample sizes. Both the resident physician (Phase 1 \( p = 1.0 \), Phase 2 \( p = 0.68 \)) and nursing student
(Phase 1 $p = 0.13$, Phase 2 $p = 0.88$) groups showed no statistically significant difference in proportion (addressed in section 5.2.2). The Nursing student hypothesis was supported by the general differences in proportions, by percentages, and by the order of the percentages for the nursing student group. The Resident physician hypothesis was not supported based on the general differences in proportions and percent difference. The results for the resident physician group were skewed by the expertise of the individual novice and expert resident physicians in the two groups.

4.2.1.1.2.4 Description and Results of Post-Task Survey Question 4

**Description**

<table>
<thead>
<tr>
<th>ID</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</td>
</tr>
</tbody>
</table>

| Answer Options | Unstructured text entry |

Question 4, identified as Q4, verified participant knowledge by allowing participants to restate their understanding of the scenario using their own description of the information need created by the scenario stem and question. Question 4, which is an unstructured response question, confirmed responses to objective questions Q1 through Q3.

**Results**

Table 4-5 shows the interpretation of the scenario by participants for the four groups.
Table 4-5: Results for Question 4, participant interpretation of scenario topic

<table>
<thead>
<tr>
<th>Participant Groups</th>
<th>Task 1 Number Correct</th>
<th>Task 2 Number Correct</th>
<th>Total Number Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice Resident Physicians (N=13)</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Expert Resident Physicians (N=16)</td>
<td>9</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Novice Nursing Students (N=10)</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Expert Nursing Students (N=17)</td>
<td>11</td>
<td>6</td>
<td>17</td>
</tr>
</tbody>
</table>

The review of answers to the Question 4 in Appendix G shows that, on average, expert resident physician and expert nursing student subgroups were able to correctly identify more scenarios than the novice resident physician and novice nursing student subgroups. A correct identification means that participants answered Q4 by providing technical terms that reflect the topic described in the scenario stem and would lead to selection of returned guidelines that answer the question in the scenario. For example in Scenario 1, regarding hospital-acquired pneumonia (HAP) that can occur in patients who are on a breathing machine, correct identification is related to Ventilator-associated Pneumonia (VAP) or ventilators because VAP is a type of HAP caused by a ventilator. Responses that mention VAP were interpreted to be correct. Scenario 2 presented a patient in a coma caused by traumatic brain injury that required monitoring intracranial pressure. Responses that mention traumatic brain injury, TBI, or pressure monitoring were interpreted to be correct. Scenario 3 presented a patient with Pulmonary Tuberculosis. Responses that mention tuberculosis, TB, or management of TB were interpreted as correct. Question 4 confirmed the general self-assessment of knowledge for participant subgroups at the individual participant level. Overall, answers to Q4 verified the correct
division of the groups by the expertise variable by confirming individual participant responses within the groups. These results, again, support the expertise hypotheses that participants with more training can be considered more expert and have more knowledge about the topics within their domain of expertise.

Overall the post-task surveys support our expertise hypotheses: Resident physician hypothesis and Nursing student hypothesis. There is some evidence of an expertise difference based on the time of exposure to the domain topics as stated in research by Doerschug et al. (1999) and Ericsson et al. (2007). Specifically, expert groups performed better on the post-task survey questions in their own self-assessment and in the interpretation of the scenario topics. In general, the expertise hypotheses were also supported by their keyword listing where both expert subgroups performed better than the novice nursing student subgroup, and physicians performed better than nursing students. Based on the results we use Expertise as a variable to differentiate the User parameter, and can assume that our subgroup division of resident physicians by year of training in the residency program and division of nursing students by participation in undergraduate and graduate programs is acceptable.

4.2.1.2 Application Task

Both studies include the same application task. The application task is searching through medical guidelines based on a medical scenario. Participants were given a medical scenario and asked to select keywords to search for the appropriate medical guidelines to answer the question posed at the end of the scenario text. The scenarios were reviewed by domain experts to determine the final phrasing of the information need displayed to the users in the experiment. An example of a medical scenario describing
hospital-acquired pneumonia is shown in Figure 4-2 while the text for all three scenarios used in the studies is shown in Appendix F: Complete Text for All Scenarios.

A type of hospital-acquired pneumonia (HAP) can occur in people who are on a breathing machine through an endotracheal or tracheostomy tube for at least 48 hours. The pneumonia primarily occurs because the tube allows free passage of bacteria into the lower segments of the lung in a person who often has underlying lung or immune problems. Your patient shows the following signs: alternating fever and low body temperature, purulent sputum, and hypoxia.

What intervention is most likely to decrease the incidence of this type of pneumonia in the intensive care unit?

Figure 4-2: Example scenario

The scenario above describes ventilator-associated pneumonia (VAP), which is a hospital-acquired pneumonia defined as “pneumonia (infection of the lung) occurring in a person who has been assisted by mechanical ventilation (a breathing machine) within the past 48 hours” and is a “serious and life-threatening infection.” Individuals contracting ventilator-associated pneumonia are typically in a hospital ICU and are already critically ill. The death rate from ventilator-associated pneumonia is high. “Between 250,000 and 300,000 cases per year occur in the United States alone, which is an incidence rate of 5 to 10 cases per 1,000 hospital admissions. The mortality attributable to VAP has been reported to range between 0 and 50% (Koenig & Truwit, 2006).”

The information need is to first identify the correct type of pneumonia infection, and to then identify the common etiologic agents and intervention that is most likely to decrease the incidence of ventilator-associated pneumonia in the ICU. The physician or nurse should be able to identify that the information above is describing ventilator-

---

1 [http://jama.ama-assn.org/cgi/content/full/297/14/1616](http://jama.ama-assn.org/cgi/content/full/297/14/1616)
associated pneumonia, know that Staphylococcus aureus infection is one of the leading etiologic agents of ventilator associated pneumonia, and deduce the correct intervention of subglottic secretion drainage as the best intervention for the condition.

Looking up Diseases > Respiratory Tract Diseases > Respiratory Tract Infections > Pneumonia > Pneumonia, Ventilator-Associated in the hierarchy; selecting Pneumonia, Ventilator-Associated as the search terms; and submitting them as a query to the guideline database should return the relevant guidelines that relate to VAP. The physician or nurse can then select the guidelines associated with VAP and subglottic secretion drainage.

An expert resident in the third year of residency or above, or a nurse in the M.S.N. program, should be able to identify the information need and navigate to the correct section of the hierarchy based on their work in the hospital environment and probable exposure to ventilator-associated pneumonia policies and guidelines. A novice resident physician within the internship year or the second year of residency or a nurse in the last year of the B.S.N. program may not be familiar with the specifics of ventilator-associated pneumonia and related policies and guidelines. Experts will know technical terms and their interrelationships whereas novices will not know enough technical terms to formulate a complete query.

Each study consists of three steps. Each participant repeats the three steps during two phases as shown in Figure 4-1. The phases were supported by a changed and an unchanged pruned lightweight MeSH ontology. All participants read a medical scenario, traversed the ontology to select concept terms for a query, submitted the query, and selected medical guidelines to answer the question posed in the scenario. The study was
designed and implemented to only allow ontology concept terms to be added as keywords in the query. The participants identified keywords based on their knowledge of the domain, the information need presented in the medical scenario, and the presented ontology. The following steps depict the relevant content from the website used in the studies. Complete web page screenshots can be found in Appendix E.

**Step 1**

Participants are required to read a scenario. The scenario is one of three scenarios that are randomly assigned to participants during each phase. Scenario 1 describes ventilator-associated pneumonia, Scenario 2 describes a coma after a traumatic brain injury, and Scenario 3 describes Pulmonary Tuberculosis. The complete text for all scenarios including the stem and question can be seen in Appendix F. Figure 4-3 depicts the webpage excerpt for Step 1 showing the directions and scenario.

![Figure 4-3: Excerpt of the scenario step on the Task webpages](image)

**Step 2**

Participants search the ontology, presented as a navigable hierarchy, and select ontology concept terms as keywords. When a participant finds a desired keyword, double-clicking the keyword in the hierarchy allows the participant to select it to add to
the search string. Then, participants are instructed to press Search to retrieve the guidelines. The search is implemented so that only terms selected from the ontology are added to the search box. The participants have to identify keywords based on their knowledge of the domain and the presented ontology to determine keywords to fully elaborate the query that will completely answer the question posed at the end of the scenario. The participants are allowed to traverse the ontology multiple times and select the best keywords that they think will return the correct guidelines to answer the question. The residents are also allowed to select multiple keywords from different ontology paths connected with a Boolean OR to allow for the expansion of the returned results.

Figure 4-4 depicts the webpage excerpt for Step 2 showing the directions, hierarchy, and Search Keywords entry textbox.

**STEP 2:**
1. Search the hierarchy below for the keyword(s) to use to find guidelines that answer the question in the scenario. Try to get to the most specific keyword(s).
2. Click on the keyword(s) to add them to the Search Keywords textbox.
3. Click Search to retrieve the guidelines. The guidelines will appear in the Search Results section below.

<table>
<thead>
<tr>
<th>Search Keywords</th>
<th>Pneumonia, Bacterial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HIERARCHY**

Note: Use + to expand and - to collapse the hierarchy.

- Diseases
  - Bacterial Infections and Mycoses
    - Bacterial Infections
      - Bacteremia
      - Central Nervous System Bacterial Infections
        - Endocarditis, Bacterial
      - Eye Infections, Bacterial
      - Gram-Negative Bacterial Infections
      - Gram-Positive Bacterial Infections
      - Pneumonia, Bacterial
    - Sexually Transmitted Diseases, Bacterial
    - Skin Diseases, Bacterial
    - Spirochaetales Infections
    - Vaginosis, Bacterial
    - Central Nervous System Infections
  - Infection
  - Mycoses

**Figure 4-4: Excerpt of the hierarchy search step on the Task webpages**
Step 3

Participants examine the retrieved guidelines and select the ones they deem appropriate to answer the question posed at the end of the scenario. Figure 4-5 depicts the webpage excerpt for Step 3 showing the directions, returned guideline results, and Guidelines Selected entry textbox.

4.2.1.3 Domain

The selected medical domain consists of guidelines from the National Guideline Clearinghouse (NGC). The NGC is a public resource for evidence-based clinical practice guidelines\(^2\). It is maintained by the Agency for Healthcare Research and Quality (AHRQ)\(^3\), one of the operating divisions of the Department of Health and Human

\(^2\) http://guideline.gov/
\(^3\) http://www.ahrq.gov/
Services (HHS)⁴. The guidelines are available in PDF, MS Word, HTML, and XML format; of which the HTML format was utilized for the experiment.

A guideline contains the following information:

- Guideline Title
- Bibliographic Source(s)
- Guideline Status
- Scope
- Qualifying Statements
- Methodology
- Implementation of the Guideline
- Recommendations
- Institute of Medicine (IOM) National Healthcare Quality Report Categories
- Evidence Supporting the Recommendations
- Identifying Information and Availability
- Benefits/Harms of Implementing the Guideline Recommendations
- Disclaimer

A guideline also contains a classification section based on the U.S. National Library of Medicine (NLM) developed Unified Medical Language System (UMLS) concept vocabularies:

- Healthcare Common Procedure Coding System (HCPCS)
- International Classification of Diseases – Clinical Modification (ICD-9-CM)
- Medical Subject Headings (MeSH)
- Physician Data Query (PDQ)
- Standard Product Nomenclature (SPN)
- Systemized Nomenclature of Medicine (Clinical Terms) (SNOMED CT)
- UMLS Metathesaurus (MTH)
- Universal Medical Device Nomenclature System (UMDNS)

An example of a complete guideline titled “Strategies to prevent ventilator-associated pneumonia in acute care hospitals” is shown in Appendix C. Guidelines are the instances

⁴ http://www.hhs.gov/
of a category or concept similar to Wikipedia articles being instances of a category or concept (Yu et al., 2007).

4.2.1.4 **Specification of the Domain by the Lightweight MeSH Ontology**

The ontology utilized in the experiment is a lightweight ontology. A lightweight ontology contains concepts, a taxonomy structure, and possibly additional relationships linking the concepts outside of the normal taxonomical structure. The lightweight ontology is based on the Medical Subject Headings (MeSH) hierarchy in the National Guideline Clearinghouse (NGC). The lightweight ontology was created by pruning the Medical Subject Headings (MeSH) hierarchy to focus on Internal Medicine. An example of a relevant section is shown in Figure 4-6 (a more in-depth hierarchy section is shown in Appendix D). The Medical Subject Headings were developed by the U.S. National Library of Medicine (NLM). It is a “controlled vocabulary thesaurus consisting of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. At the most general levels of the hierarchical structure are very broad headings such as ‘Anatomy’ or ‘Mental Disorders.’ More specific headings are found at more narrow levels of the twelve-level hierarchy, such as ‘Pneumonia’ and ‘Pneumonia, Ventilator-Associated.’ There are 26,853 descriptors in 2013 MeSH.5”

MeSH is organized as a hierarchical tree structure which means that “each MeSH descriptor appears in at least one place in the tree, and may appear in as many additional places as may be appropriate. Those who index articles or catalog books are instructed to find and use the most specific MeSH descriptor that is available to represent each indexable concept. For example, articles concerning Streptococcus pneumoniae will be

---

found under the descriptor Streptococcus pneumoniae rather than the broader term Streptococcus, while an article referring to a new streptococcal bacterium which is not yet in the vocabulary will be listed directly under Streptococcus. This is important to the study because participants who identify the correct path to the most specific MeSH descriptor – or concept – will get better precision in the returned medical guidelines to answer the question posed the medical scenario.

<table>
<thead>
<tr>
<th>Disease/Condition</th>
<th>Conditions</th>
<th>Diseases</th>
<th>Respiratory Tract Diseases</th>
<th>Lung Diseases</th>
<th>Pneumonia</th>
<th>Pneumonia, Ventilator-Associated</th>
</tr>
</thead>
</table>

**Figure 4-6: Sample of the MeSH pneumonia related hierarchy**

**Extracting the MeSH Ontology from the NGC**

The MeSH ontology was extracted from the National Guideline Clearinghouse website by expanding the hierarchy structure and copying the nodes from the website pages. The ontology is part of the MeSH structure and starts at the topmost Diseases node and includes all nodes shown in the Original MeSH Ontology column in Table 4-6. The ontology hierarchy was pruned from 26 first level nodes underneath Diseases to 9 nodes. Pruning the hierarchy simplified it to reduce duplication of sub-topics and hierarchy pathways to keywords, made the hierarchy navigation more manageable for the

---

participant, and focused more on Internal Medicine topics. The number of navigable nodes was reduced from 4349 to 2205. The number of guidelines remained unchanged.

Table 4-6: MeSH Ontology categories before and after pruning

<table>
<thead>
<tr>
<th>Original MeSH Ontology</th>
<th>Experimental MeSH Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal Diseases</td>
<td>Bacterial Infections and Mycoses</td>
</tr>
<tr>
<td>Bacterial Infections and Mycoses</td>
<td>Cardiovascular Diseases</td>
</tr>
<tr>
<td>Cardiovascular Diseases</td>
<td>Digestive System Diseases</td>
</tr>
<tr>
<td>Congenital, Hereditary, and Neonatal Diseases and Abnormalities</td>
<td>Endocrine System Diseases</td>
</tr>
<tr>
<td>Digestive System Diseases</td>
<td>Hemic and Lymphatic Diseases</td>
</tr>
<tr>
<td>Disorders of Environmental Origin</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>Endocrine System Diseases</td>
<td>Nervous System Diseases</td>
</tr>
<tr>
<td>Eye Diseases</td>
<td>Respiratory Tract Diseases</td>
</tr>
<tr>
<td>Female Urogenital Diseases and Pregnancy Complications</td>
<td>Virus Diseases</td>
</tr>
<tr>
<td>Hemic and Lymphatic Diseases</td>
<td></td>
</tr>
<tr>
<td>Immune System Diseases</td>
<td></td>
</tr>
<tr>
<td>Male Urogenital Diseases</td>
<td></td>
</tr>
<tr>
<td>Musculoskeletal Diseases</td>
<td></td>
</tr>
<tr>
<td>Neoplasms</td>
<td></td>
</tr>
<tr>
<td>Nervous System Diseases</td>
<td></td>
</tr>
<tr>
<td>Nutritional and Metabolic Diseases</td>
<td></td>
</tr>
<tr>
<td>Occupational Diseases</td>
<td></td>
</tr>
<tr>
<td>Otorhinolaryngologic Diseases</td>
<td></td>
</tr>
<tr>
<td>Parasitic Diseases</td>
<td></td>
</tr>
<tr>
<td>Pathological Conditions, Signs and Symptoms</td>
<td></td>
</tr>
<tr>
<td>Respiratory Tract Diseases</td>
<td></td>
</tr>
<tr>
<td>Skin and Connective Tissue Diseases</td>
<td></td>
</tr>
<tr>
<td>Stomatognathic Diseases</td>
<td></td>
</tr>
<tr>
<td>Substance-Related Disorders</td>
<td></td>
</tr>
<tr>
<td>Virus Diseases</td>
<td></td>
</tr>
<tr>
<td>Wounds and Injuries</td>
<td></td>
</tr>
</tbody>
</table>

4.2.2 Structural Ontology Characteristics in the Studies

The studies we present in this chapter were conducted using structural characteristics. Specifically we selected the measures of depth, breadth, and fanoutness (Gangemi et al., 2005) (Gangemi et al., 2006). Table 4-7 shows the selected measures and their definition. These measures were chosen because of their simplicity and their impact on
the calculation of other measures, such as the density measure. For example, the density measure is defined as number of relations, subclasses, superclasses, and siblings for a specific concept (Alani & Brewster, 2006). Changing fanoutness will impact this measure by reducing the number of siblings for a concept. Of the 30 available measures, the selected set is easier to understand, calculate, and implement as ontology metadata than other measures, and will be more valuable when used in a recommendation framework in order to provide guidelines for selecting them for a specific context. Both researchers and practitioners in organizations will be able to understand and reproduce the results of our studies, and implement the measures during ontology engineering and evaluation. Additionally, these measures are more easily calculated, understood, and manipulated by non-expert developers in the ontology engineering domain.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Definition by Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td><strong>Defined</strong> as a graph property related to the cardinality of paths in a graph, where the arcs considered are only is-a arcs. <strong>Calculated</strong> as the average of the sum of all is-a paths starting at the top node and terminating in a leaf node.</td>
</tr>
<tr>
<td>Breadth</td>
<td><strong>Defined</strong> as a property related to the cardinality of levels (“generations”) in a graph, where the arcs considered here are only is-a arcs. <strong>Calculated</strong> as the sum of all is-a edges per level divided by the number of levels.</td>
</tr>
<tr>
<td>Fanoutness</td>
<td><strong>Defined</strong> as a relation to the “dispersion” of graph nodes, along is-a arcs. We distinguish the fanoutness measures related to leaf node sets, and fanoutness measures related to sibling node sets (i.e., “internal dispersion”). <strong>Calculated</strong> as the total sum of the average of is-a edges per node divided by the number of levels. Nodes with zero child nodes are not counted in the calculation.</td>
</tr>
</tbody>
</table>
Criteria were not selected for this research due to the lack of easily operationalized criteria, i.e. no criteria were directly defined by a formula, or defined well enough in the evaluation approach where a formula could be derived to express the criteria in a concise way symbolically.

4.2.3 Objective Quality Metrics in the Studies

The study includes two quality metrics that are appropriate for the selection of the search application task and user expertise variable. The quality metrics are **Path Selected** and **Guidelines Selected**.

Before performing the application task, study participants were asked to read a medical scenario. It is only after understanding the medical scenario, which motivated the need for searching for guidelines, that they traversed the ontology to select keywords to create a query. Participants were not limited to the traversal of only one path and selection of only one node to better emulate realistic searching using one to three terms per query. A specific path selected through the hierarchy that leads to query terms was the first result of using the ontology. This metric was adopted as the first indication of performance and called the **Path Selected** metric.

When the query was submitted, a set of guidelines was retrieved for a participant to select. The specific guidelines selected by participants to answer the question in the medical scenario were the second indication of performance adopted and were called the **Guidelines Selected** metric.
4.3 Methodology Steps in the Studies

This section describes the methodology steps across the two phases introduced in Section 3.2 as they were implemented in the two studies.

4.3.1 Phase 1

4.3.1.1 Select

In the Select step in the studies the healthcare context was defined by healthcare workers who searched for medical guidelines in the National Guideline Clearinghouse. These three context parameters define \( c \) in the healthcare context. The \( r \) was defined by three structural characteristics that are related to the search application task, the choice of the lightweight pruned MeSH ontology focusing on Internal Medicine, and the procedures of the search application task.

4.3.1.1.1 Define User Parameter Using the Expertise Variable

The studies focused on the user context parameter defined by the expertise variable with a value of \textit{novice} or \textit{expert}. The definition of novice and expert participants in the healthcare context can be found in Section 4.2.1.1 User.

4.3.1.1.2 Select Search Application Task

The studies focused on the search application task. Searching for information on the web, or in a health information system, is a typical task performed by users in the medical domain (Hughes et al., 2009). In order to properly assess the users and task we implement the expert-performance approach. The expert-performance approach uses “reproducibly superior performance on representative tasks within the domain (Ericsson et al., 2007),” in this case the search for medical guidelines based on a medical scenario.
4.3.1.1.3 Select Ontology Characteristics

The characteristics selected for the studies were breadth, depth, and fanoutness, which are structural characteristics (i.e., those based on the physical properties of the ontology) for implementation. These characteristics were selected for four reasons. First, breadth, depth, and fanoutness are well defined as described in Table 4-7. Second, the characteristics can be calculated from a formula. Third, the characteristics are easy to understand and thus likely to be applied and documented in an ontology engineering approach. Finally, the results based on structural measures can be immediately applied because they relate to all ontologies – lightweight and heavyweight.

4.3.1.1.4 Select the Location to Make the Change to the Characteristics

The location of the change to the selected ontology characteristics in the studies was based on theories related to categories and categorical structures. The change to ontology characteristics was performed around the basic level. Categories, including ontology categories, are organized in such a way as to have the basic level in the middle of the hierarchy with generalization moving upward in the hierarchy to superordinate categories and specialization moving downward in the hierarchy to subordinate categories (Uschold & King, 1995). Section 4.4 Changing Ontology Characteristics in the Applied Methodology details the change procedure and the reason behind the change.

4.3.1.1.5 Select Objective Quality Metrics

As stated previously, searching is a typical task performed by users in the medical domain. User performance on the search application task is measured by using two distinct metrics, the Path Selected and the Guidelines Selected. Before performing the application task, study participants were asked to read a medical scenario. It is only after
understanding the medical scenario that motivates the need for searching for guidelines, that they traversed the ontology to select keywords to create a query. The application task and steps are described in Section 4.2.1.2 Application Task. A specific path selected through the ontology hierarchy that leads to query terms is the first result of using the ontology. This was adopted in the studies as the first indication of performance called the Path Selected metric. When the query is submitted, a set of guidelines is retrieved for a participant to select. The specific guidelines selected by participants to answer the question posed in the medical scenario are the second indication of performance adopted in the studies and called the Guidelines Selected metric.

The Path Selected metric was based on the selection of correct and incorrect paths to ontology terms that could form the correct query to retrieve relevant guidelines. For each path selected, the path was labeled as:

- Correct: Path leads to a correct leaf term in the MeSH hierarchy to retrieve a guideline relevant to the medical scenario.

- Incorrect: Path does not lead to a correct leaf term in the MeSH hierarchy to retrieve a guideline relevant to the medical scenario.

The Guidelines Selected metric was based on the selection of correct and incorrect guideline to answer the question posed at the end of the scenario. For each guideline selected, the guideline was labeled as:

- Correct: Guideline can be used to answer the question posed at the end of the medical scenario.

- Incorrect: Guideline cannot be used to answer the question posed at the end of the medical scenario.
The correctness of the paths and guidelines relevant to the medical scenarios was determined by polling a group of physicians and nurses serving as an expert panel.

4.3.1.2 Measure Initial Performance

The initial performance ($p_{BC}$) of the search application task supported by the MeSH ontology before the change was measured for all groups in the two studies.

4.3.2 Phase 2

4.3.2.1 Change Values

The lightweight pruned MeSH ontology characteristic(s) were changed at the basic level by removing the superordinate level categories, thereby reducing the knowledge representation of the ontology by making it more difficult to determine the disease etiology. The change negatively impacted the performance of the search application task. This change was related to the user expertise variable, the search application task, and the choice of structural characteristics.

<table>
<thead>
<tr>
<th>L1. Bacterial Infections and Mycoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2. Bacterial Infections</td>
</tr>
<tr>
<td>L3. Gram-Negative Bacterial Infections</td>
</tr>
<tr>
<td>L4. Bordetella Infections</td>
</tr>
<tr>
<td>L5. Whooping Cough</td>
</tr>
<tr>
<td>L4. Enterobacteriaceae Infections</td>
</tr>
<tr>
<td>L5. Escherichia coli Infections</td>
</tr>
<tr>
<td>L5. Granuloma Inguinale</td>
</tr>
<tr>
<td>L5. Salmonella Infections</td>
</tr>
<tr>
<td>L6. Salmonella Food Poisoning</td>
</tr>
<tr>
<td>L4. Tick-Borne Diseases</td>
</tr>
<tr>
<td>L5. Tularemia</td>
</tr>
</tbody>
</table>

Figure 4-7: Excerpt of the unchanged MeSH ontology
Figure 4-7 is a sample section of the unchanged lightweight pruned MeSH ontology excerpt. Figure 4-8 is a sample of the changed lightweight pruned MeSH ontology excerpt to demonstrate how each of the selected measures was changed and the result of the change on the basic level and selected structural ontology characteristics of depth, breadth, and fanoutness. Section 4.4 Changing Ontology Characteristics in the Applied Methodology details the changes and the results of the changes.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Calculation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>$4 + 4 + 4 + 5 + 4 = 21/5 = 4.2$ nodes per path</td>
<td></td>
</tr>
<tr>
<td>Breadth</td>
<td>$1 + 1 + 3 + 5 + 1 = 11/5 = 2.2$ edges per level</td>
<td></td>
</tr>
<tr>
<td>Fanoutness</td>
<td>$(1/1 + 1/1 + 3/1 + (1 + 3 + 1)/3 + 1/1)/5 = 1.5$ edges per node</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4-8: Excerpt of the changed MeSH ontology.

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1.</td>
<td>Bacterial Infections and Mycoses</td>
</tr>
<tr>
<td>L2.</td>
<td>Bacterial Infections</td>
</tr>
<tr>
<td>L3.</td>
<td>Gram-Negative Bacterial Infections</td>
</tr>
<tr>
<td>L5.</td>
<td>Whooping Cough</td>
</tr>
<tr>
<td>L5.</td>
<td>Escherichia coli Infections</td>
</tr>
<tr>
<td>L5.</td>
<td>Granuloma Inguinale</td>
</tr>
<tr>
<td>L5.</td>
<td>Salmonella Infections</td>
</tr>
<tr>
<td>L6.</td>
<td>Salmonella Food Poisoning</td>
</tr>
<tr>
<td>L5.</td>
<td>Tularemia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Calculation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>$3 + 3 + 3 + 4 + 3 = 16/5 = 3.2$ nodes per path</td>
<td></td>
</tr>
<tr>
<td>Breadth</td>
<td>$1 + 1 + 5 + 1 = 16/4 = 4.0$ edges per level</td>
<td></td>
</tr>
<tr>
<td>Fanoutness</td>
<td>$(1/1 + 1/1 + 5/1 + 1/1)/4 = 2.0$ edges per node</td>
<td></td>
</tr>
</tbody>
</table>
In the example the change results in decreased average depth and increased average breadth and fanoutness. The largest impact in our study was on breadth, as it is in this example.

4.3.2.2 Measure Resultant Performance

The resultant performance \( p_{AC} \) of the search application task supported by the MeSH ontology after the change was measured for all groups in the two studies.

4.3.2.3 Analyze

The difference in results for performance \( p \) for the search application task for the selected characteristics of depth, breadth, and fanoutness in the defined context was analyzed. The context was defined as resident physicians and nursing students searching for medical guidelines in the National Guidelines Clearinghouse supported by the MeSH ontology. Performance \( p \) showed a significant difference as measured by the Path Selected and the Guidelines Selected metrics. The discussion can be found in section 4.7 Results.

4.3.2.4 Save and Apply

The results described in Sections 4.7.1.1 and 4.7.2.1 show statistical significance in the difference for novice users for the two studies. In this case we can attempt to create and document a guideline for the specific combination ontology characteristics of depth, breadth, and fanoutness and context of resident physicians and nursing students searching for medical guidelines in the National Guidelines Clearinghouse, along with the selected change to the characteristics.
4.4 Changing Ontology Characteristics in the Applied Methodology

The changes to the breadth, depth, and fanoutness measures, which are structural characteristics selected for the studies, were made at the basic level. The unchanged ontology contained ten (10) hierarchical levels with Level 5 identified as the basic level. The following procedure was carried out on the pruned and simplified lightweight MeSH ontology (see Section 4.2.1.4 for more detail) to attain the desired change to the characteristics.

Step 1: Count nodes and branches per node

All leaf nodes and branches per node were counted.

Step 2: Calculate measures

The selected measures of breadth, depth, and fanoutness were calculated using the values from Step 1 and the formulas defined in Table 4-7. For the lightweight ontology based on the pruned Medical Subject Headings (MeSH) hierarchy focusing on Internal Medicine in the studies the figures are Depth = 4.63 nodes per path, Breadth = 433.5 edges per level, Fanoutness = 4.31 edges per node.

Step 3: Determine basic level

The basic level was determined by the amount of information gained when moving from a more general level of abstraction to a more specific level. The basic level is determined by identifying basic level categorization, primacy, and functional embodiment. Basic level categorization implies that categories are not merely organized in a hierarchy from the most general to the most specific, but organized so that categories that are cognitively basic are in the middle of a general-to-specific hierarchy; with generalization proceeding upward from the basic level and specialization proceeding
Basic level primacy implies that basic level categories are functionally and epistemologically primary in gestalt perception, image formation, motor movement, knowledge organization, ease of cognitive processing (learning, recognition, memory), and ease of linguistic expression (Lakoff, 1987). Functional embodiment implies that certain concepts are not merely understood intellectually but used automatically, unconsciously, and without noticeable effort as part of normal functioning (Lakoff, 1987).

**Example**

**Level 4:** *Meningitis, Bacterial* – does not have a specific cure because it can be caused by different types of bacteria. There are several pathogens (types of germs) that can cause bacterial meningitis. Some of the leading causes of bacterial meningitis in the United States include *Haemophilus influenzae* (most often caused by type b, Hib), *Streptococcus pneumoniae*, group B *Streptococcus*, *Listeria monocytogenes*, and *Neisseria meningitides* (Bacterial Meningitis, 2012).

**Level 5:** *Meningitis, Haemophilus* – has a specific cure and thus a better functional embodiment because it is caused by a specific organism. *Haemophilus Meningitis* is caused by the bacteria *Haemophilus influenzae* type b. *Haemophilus influenzae* type b vaccine prevents this type of meningitis (an infection of the covering of the brain and spinal cord), and other serious infections caused by a type of bacteria called *Haemophilus influenzae* type b. It is recommended for all children younger than 5 years old in the US, and it is usually given to infants starting at 2 months old. The Hib vaccine can be combined with other vaccines (Hib Vaccination, 2012).

The above example shows the increasing amount of information gained when moving down the hierarchy to the basic level. The basic level (Level 5 in this case) is the level in the middle of the 10 level hierarchy. It also has basic level primacy with respect to the medical domain because it is specifically acted upon by a physician. It has functional embodiment in that it represents one object or concept – a specific disease.
Step 4: Make the change

Level 4 of the hierarchy, immediately above the basic level was removed. This reduced the hierarchy. The reference to a bacterial etiology was removed. Meningitis can be caused by viruses, bacteria, fungi, or protozoa (parasites). Therefore, the participant must know that this is bacterial and that treatment for other causes e.g. antivirals or corticosteroids will not be effective. This is due to recognition of the name. In the medical scenario the name of the disease or required procedure is not stated directly; only the signs and symptoms are stated leaving the participant to diagnose the disease and determine the path through hierarchy. This change removes a link to knowing that an object belongs to a category which would allow the participant to plausibly infer not only the properties of that category, but also those of categories higher or lower in the ontology hierarchy (Markman & Wisniewski, 1997).

For the lightweight ontology in the studies the final figures are Depth = 4.41 nodes per path, Breadth = 245.8 edges per level, Fanoutness = 4.13 edges per node. The largest impact was on Breadth as shown in Table 4-8.

<table>
<thead>
<tr>
<th>Selected Structural Characteristic (Measure)</th>
<th>Phase 1 Before Change</th>
<th>Phase 2 After Change</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth in nodes per path</td>
<td>4.63</td>
<td>4.41</td>
<td>4.75% decrease</td>
</tr>
<tr>
<td>Breadth in edges per level</td>
<td>433.5</td>
<td>245.8</td>
<td>43.29% decrease</td>
</tr>
<tr>
<td>Fanoutness in edges per node</td>
<td>4.31</td>
<td>4.13</td>
<td>4.17% decrease</td>
</tr>
</tbody>
</table>
4.5 Basic Level Categories, Expertise, and Impact on Search

Basic level categories are middle level categories in a hierarchy or taxonomy and are functionally primary in gestalt perception, knowledge organization, image formation, and linguistic expression (Lakoff, 1987). There is a wealth of research about the basic level in categorization (Hajibayova, 2013) and about expertise and its relationship to the basic level (Tanaka & Taylor, 1991; Johnson & Mervis, 1997). The research demonstrates that domain experts can identify more differences or distinctions per level in a hierarchy within their domain of expertise than novices. Knowledge of the distinctions about domain-specific terms across hierarchy levels allows an expert to correctly traverse and identify appropriate terms to use to formulate a query for information retrieval.

Superordinate categories are upper level categories in a hierarchy that have fewer defining attributes than basic level categories, and are typically mass nouns (Lakoff, 1987) – meaning they cannot be counted individually (e.g. furniture, wine). They are highly distinctive, but not very informative in that they clearly distinguish objects by major differences but do not provide as high a number of differences as lower level categories (Tanaka & Taylor, 1991; Johnson & Mervis, 1997). In this way superordinate categories act as gateways to the basic level. The basic level and subordinate level terms act as more specific differentiators. As described previously in Section 2.12.3 Search and Expertise, experts know more technical terms in the domain and by extension in the pruned lightweight MeSH ontology hierarchy in our studies. The experts can therefore navigate and select more relevant technical terms from the lightweight ontology.

Therefore, when we remove superordinate categories from the lightweight MeSH ontology hierarchy, it is reasonable to expect that this removal is detrimental to the ability
of that ontology to support the novice by limiting the clear and distinctive gateway. In our research it is the disease etiology that is made less clear for the novice. At the same time, we promote more technical terms that are less recognizable by the novice. In the studies we specifically made a change related to expertise by removing the level linking superordinate categories to the basic level. We hypothesized that this change would negatively impact novice participants because novice participants depend on the link between the superordinate categories, the basic level, and subordinate categories to supply domain knowledge. The impact may be negative for experts depending on the individual expert performance and knowledge of the topics presented in the scenarios.

4.6 Hypotheses in the Studies

Ontologies are a knowledge representation of a domain. That knowledge representation is used to support various application tasks. This research tests the relationship between the design of the ontology (described by its characteristics such as breadth, depth, and fanoutness) and its context. The exact role of the ontology is to complement the users understanding of the domain. For this reason the research focuses on the user as a primary parameter of context. To ensure a thorough experimental design and produce sufficiently explanatory results the user parameter is delineated by two expertise levels of novice and expert across participant groups of nursing students and resident physicians. Literature explains that both novices and experts use ontologies in search (Liu & Wacholder, 2008; White et al., 2009), and that search is a common task in the medical domain (Hughes et al., 2009). The search application task is therefore utilized in our experimental design. Because the role of an ontology is to support a user performing an application task in a domain, it is reasonable to hypothesize that the
removal of a part of the ontology would be detrimental to that role (Coletti & Bleich, 2001; Eisinger et al., 2013).

Table 4-9: Hypotheses testing the impact of a change to structural ontology characteristics on novices and experts performing a search application task in the medical domain

<table>
<thead>
<tr>
<th>Hypothesis No.</th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1</td>
<td>The removal of the superordinate level above the basic level in the ontology (causing a change in the selected ontology characteristics of depth, breadth, and fanoutness) will negatively impact the performance of the search application task when users are novice resident physicians.</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>The removal of the superordinate level above the basic level in the ontology (causing a change in the selected ontology characteristics of depth, breadth, and fanoutness) will negatively impact the performance of the search application task when users are expert resident physicians.</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>The removal of the superordinate level above the basic level in the ontology (causing a change in the selected ontology characteristics of depth, breadth, and fanoutness) will negatively impact the performance of the search application task when users are novice nursing students.</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>The removal of the superordinate level above the basic level in the ontology (causing a change in the selected ontology characteristics of depth, breadth, and fanoutness) will negatively impact the performance of the search application task when users are expert nursing students.</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>The removal of the superordinate level above the basic level in the ontology (causing a change in the selected ontology characteristics of depth, breadth, and fanoutness) will have a greater impact on the performance of the search application task for novice users than expert users.</td>
</tr>
</tbody>
</table>

Table 4-9 shows the hypotheses tested in the experimental design described in Section 4.1. We hypothesize that there will be a negative impact on performance when
the selected ontology characteristics are changed at the basic level. The negative impact is related to the effect of the change on participant expertise and on the structure of the lightweight MeSH ontology. The hypotheses are presented as one-tailed directional alternative hypotheses. They are divided into the two participant groups of nursing students and resident physicians. For each participant group, the hypotheses are divided by the expertise variable of novice and expert; and a general hypothesis is made about the difference in expertise.

The supposition that there will be a negative, or detrimental, impact on performance is based on the relationship of the change to expertise. The discussion about search and expertise in Sections 2.12.3 and 4.5 describes the differences in the exploitation of knowledge for search by experts and novices supported by a MeSH ontology. Novices are not adept at utilizing MeSH terms due to a lack of recognition of the domain-specific lexicon. They are not able to utilize domain knowledge to navigate through the terms in the MeSH ontology hierarchy, or recognize the differences between subordinate MeSH terms. Novices rely on the superordinate categories to act as gateways to the basic level and use the MeSH ontology hierarchy and the distinction of superordinate category terms for knowledge representation. We expected the impact of the change to the ontology to be detrimental for the novice groups due to the promotion of more technical domain terms to the upper levels of the hierarchy and the novices’ inability to recognize these terms and their interrelationships. Additionally, the removal of the superordinate categories that provide a distinctive gateway to the disease etiology knowledge encoded at the basic level will also negatively affect novice performance. Because the ontology is
a knowledge base and complements the users understanding of the domain, any
detrimental change to the ontology will have a more negative impact on the novice user.  

Generally experts are better able to utilize domain knowledge to navigate through the
terms in the MeSH ontology hierarchy, recognize the differences between MeSH terms,
and then utilize the terms in a query. Experts rely less on the superordinate categories to
act as gateways to the subordinate categories. Overall, the validation of expertise in
Section 4.2.1.1.2 shows that, as a group, experts were more knowledgeable about the
scenario topics based on their own self-assessment, keyword selection, and interpretation
of the scenarios. However, this section also shows that experts with less knowledge
within the group performed worse than their peers. We expect the impact of the change
to the ontology to be detrimental, although not as severe, for the expert groups because of
the same limit to the disease etiology and impact on the less knowledgeable experts.

The experiment utilized a lightweight MeSH ontology. The lightweight MeSH
ontology is an established representation of expert medical knowledge based on its
acceptance in the medical community and widespread use in information systems at the
National Center for Biotechnology Information, National Institutes of Health, PubMed,
National Guideline Clearinghouse, and other systems and sites related to medical
information (Coletti & Bleich, 2001; Eisinger et al., 2013). It provides demonstrated
support for the search application task in the healthcare domain (Coletti & Bleich, 2001;
Eisinger et al., 2013). In the experimental studies the lightweight MeSH ontology was
changed by removing the superordinate categories above the basic level with the result of
changing the selected ontology characteristics as described in Section 4.4. This change
was related to expertise, the variable selected to differentiate the user context parameter
as described in Section 4.2. Removing the superordinate categories made it more difficult to determine the disease etiology with the result that the ontology was less able to support the search application task for both experts and novices. We hypothesize that this general change to make the ontology less able to support the search application task by removing a significant part of the ontology should have a negative impact on all participant groups.

4.7 Results

This section describes the results of the two studies conducted for the resident physician and nursing student participant groups. Table 4-10 shows the results summary for the Path Selected and Guidelines Selected metrics for the novice and expert resident physician and nursing student groups. The significant differences in the selected objective quality metrics are highlighted in bold. Sections 4.7.1 and 4.7.2 present the detailed results for each hypothesis tested for the two metrics. The results are based on performance of the groups in Phase 1 with the search application task supported by the initial (i.e. unchanged) lightweight MeSH ontology and in Phase 2 with the search application task supported by the changed lightweight MeSH ontology.

A one-tailed binomial test (p < .05) was performed on the results of the resident physician and nursing student groups for the Path Selected and the Guidelines Selected metrics to determine statistical significance. The binomial test was used on the nominal Path Selected and the Guidelines Selected metrics for the dichotomous correct and incorrect categorical values. The test compared the difference in the correct versus incorrect results ratios of resident physicians and nursing students denoting a change in performance for the search application task.
Table 4-10: Results for novice and expert resident physician and nursing student groups with significant changes in performance in bold

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Novice Resident Physicians</th>
<th>Expert Resident Physicians</th>
<th>Novice Nursing Students</th>
<th>Expert Nursing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase 1 Correct %</td>
<td>Phase 2 Correct %</td>
<td>Phase 1 Correct %</td>
<td>Phase 2 Correct %</td>
</tr>
<tr>
<td>Path Selected</td>
<td>62</td>
<td>23</td>
<td>56</td>
<td>69</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>62</td>
<td>23</td>
<td>56</td>
<td>50</td>
</tr>
</tbody>
</table>

4.7.1 Results for Resident Physician Participants

This section presents results for the novice resident physician subgroup and expert resident physician subgroup that constituted the resident physician group.

4.7.1.1 Results and Statistical Significance for Hypothesis 1

The results for Hypothesis 1 are presented in Table 4-11. For the **Path Selected** metric, novice residents selected the correct path to the keywords for 62% (8/13) of their attempts in Phase 1 and for 23% (3/13) of their attempts in Phase 2.

For the **Guidelines Selected** metric, novice residents were able to select a correct guideline for 62% (8/13) of their selection attempts in Phase 1, and selected a correct guideline for 23% (3/13) of their selection attempts in Phase 2.

Table 4-11: Results for novice resident physicians

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Phase 1 Correct %</th>
<th>Phase 2 Correct %</th>
<th>% Change in Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path Selected</td>
<td>62</td>
<td>23</td>
<td>decrease of 63</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>62</td>
<td>23</td>
<td>decrease of 63</td>
</tr>
</tbody>
</table>
The novice resident physicians had a 63% decrease in performance for both the Path Selected and the Guidelines Selected metrics from 62% correct in Phase 1 to 23% correct in Phase 2. The results show a significant decrease in the correct Path Selected (p=0.03, p < .05) and the Guidelines Selected (p=0.03, p < .05) metrics between the unchanged and changed lightweight MeSH ontology. The results confirm Hypothesis 1 for both metrics.

4.7.1.2 Results and Statistical Significance for Hypothesis 2

The results for Hypothesis 2 are presented in Table 4-12. For the Path Selected metric, expert residents selected the correct path to the keywords for 56% (9/16) of their attempts in Phase 1 and for 69% (11/16) of their attempts in Phase 2.

The change to the ontology eliminated the link between the superordinate and subordinate disease concepts thereby decreasing the specific disease etiology. Nevertheless, expert residents were able to select a path through the ontology to the correct keywords. Inspection of the post-task survey revealed that expert residents were knowledgeable about the scenario topics based on their “Very Familiar” and “Familiar” responses to Question 2 (Q2) and the correct restatement of the scenario topic in their own words for Question 4 (Q4).

For the Guidelines Selected metric, expert residents were able to select a correct guideline for 56% (9/16) of their selection attempts in Phase 1, and selected a correct guideline for 50% (8/16) of their selection attempts in Phase 2.
Expert resident physicians had an increase of 23% in performance for the *Path Selected* metric from 56% correct in Phase 1 to 69% correct in Phase 2 and an 11% decrease in performance for the *Guidelines Selected* metric from 56% correct in Phase 1 to 50% correct in Phase 2. The results show no significant increase in the correct *Path Selected* (p=0.24, p > .05) and no significant decrease in the correct *Guidelines Selected* (p=0.36, p > .05) metrics between the unchanged and changed lightweight MeSH ontology. The results do not confirm Hypothesis 2 for either metric. Although there was a decrease in performance for the *Guidelines Selected* metric, the decrease was not significant.

4.7.1.3 Results and Statistical Significance for Hypothesis 5

The combined results for novice resident physicians and expert resident physicians for Hypothesis 5 are presented in Table 4-13. For the *Path Selected* metric, novice residents selected the correct path to the keywords for 62% (8/13) of their attempts in Phase 1 and for 23% (3/13) of their attempts in Phase 2 while expert residents selected the correct path to the keywords for 56% (9/16) of their attempts in Phase 1 and for 69% (11/16) of their attempts in Phase 2. For the *Guidelines Selected* metric, novice residents were able to select a correct guideline for 62% (8/13) of their selection attempts in Phase 1, and for 23% (3/13) of their selection attempts in Phase 2 while expert residents were...
able to select a correct guideline for 56% (9/16) of their selection attempts in Phase 1, and selected a correct guideline for 50% (8/16) of their selection attempts in Phase 2.

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Study performed with:</th>
<th>Novice Resident Physicians</th>
<th>Expert Resident Physicians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phase 1 Correct %</td>
<td>Phase 2 Correct %</td>
<td>Phase 1 Correct %</td>
</tr>
<tr>
<td>Path Selected</td>
<td>62</td>
<td>23</td>
<td>56</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>62</td>
<td>23</td>
<td>56</td>
</tr>
</tbody>
</table>

The results show a significant decrease in the correct Path Selected (p=0.03, p < .05) and the Guidelines Selected (p=0.03, p < .05) metrics for novice residents. The results show no significant increase in the correct Path Selected (p=0.24, p > .05) and no significant decrease in the correct Guidelines Selected (p=0.36, p > .05) metrics for expert residents. These results confirm Hypothesis 5.

4.7.2 Results for Nursing Student Participants

This section presents results for the novice nursing student subgroup and expert nursing student subgroup that constituted the nursing student group.

4.7.2.1 Results and Statistical Significance for Hypothesis 3

The results for Hypothesis 3 are presented in Table 4-14. For the Path Selected metric novice nurses selected the correct path to the keywords for 70% (7/10) of their attempts in Phase 1 and for 30% (3/10) of their attempts in Phase 2.
For the *Guidelines Selected* metric, novice nurses were able to select a correct guideline for 70% (7/10) of their selection attempts in Phase 1. They were able to select a correct guideline for 30% (3/10) of their attempts in Phase 2.

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Phase 1 correct %</th>
<th>Phase 2 correct %</th>
<th>% Change in Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path Selected</td>
<td>70</td>
<td>30</td>
<td>decrease of 57</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>70</td>
<td>30</td>
<td>decrease of 57</td>
</tr>
</tbody>
</table>

The novice nursing students had a 57% decrease in performance for both the *Path Selected* and the *Guidelines Selected* metrics from 70% correct in Phase 1 to 30% correct in Phase 2. The results show a significant decrease in the correct *Path Selected* ($p=0.04$, $p < .05$) and the *Guidelines Selected* ($p=0.04$, $p < .05$) metrics between the unchanged and changed lightweight MeSH ontology. The results confirm Hypothesis 3 for both metrics.

### 4.7.2.2 Results and Statistical Significance for Hypothesis 4

The results for Hypothesis 4 are presented in Table 4-15. For the *Path Selected* metric, expert nurses selected the correct path to the keywords for 53% (9/17) of their attempts in Phase 1 and for 47% (8/17) of their attempts in Phase 2.

For the *Guidelines Selected* metric, expert nurses were able to select a correct guideline for 35% (6/17) of their selection attempts in Phase 1 and for 24% (4/17) of their attempts in Phase 2.
Table 4-15: Results for expert nursing students

<table>
<thead>
<tr>
<th>Quality Metric</th>
<th>Phase 1 correct %</th>
<th>Phase 2 correct %</th>
<th>% Change in Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path Selected</td>
<td>53</td>
<td>47</td>
<td>decrease of 11</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>35</td>
<td>24</td>
<td>decrease of 31</td>
</tr>
</tbody>
</table>

The expert nursing students had an 11% decrease for the *Path Selected* metric from 53% in Phase 1 to 47% in Phase 2. The students had a 31% decrease for the *Guidelines Selected* metric from 35% correct in Phase 1 to 24% correct in Phase 2. The results show no significant decrease in the correct *Path Selected* (p=0.38, p > .05) and the *Guidelines Selected* (p=0.23, p > .05) metrics between the unchanged and changed lightweight MeSH ontology. The results do not confirm Hypothesis 4 for either metric. Both the *Path Selected* metric and the *Guidelines Selected* metric decreased, but the decrease was not significant.

4.7.2.3 Results and Statistical Significance for Hypothesis 5

The combined results for novice nursing students and expert nursing students are presented in Table 4-16. For the *Path Selected* metric, novice nursing students selected the correct path to the keywords for 70% (7/10) of their attempts in Phase 1 and for 30% (3/10) of their attempts in Phase 2 while expert nursing students selected the correct path to the keywords for 53% (9/17) of their attempts in Phase 1 and for 47% (8/17) of their attempts in Phase 2. For the *Guidelines Selected* metric, novice nursing students were able to select a correct guideline for 70% (7/10) of their selection attempts in Phase 1 and for 30% (3/10) of their attempts in Phase 2 while expert nursing students were able to
select a correct guideline for 35% (6/17) of their selection attempts in Phase 1 and for 24% (4/17) of their attempts in Phase 2.

Table 4-16: Results for novice vs. expert nursing student task performance

<table>
<thead>
<tr>
<th>Study performed with:</th>
<th>Novice Nursing Students</th>
<th>Expert Nursing Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality Metric</td>
<td>Phase 1 Correct %</td>
<td>Phase 2 Correct %</td>
</tr>
<tr>
<td>Path Selected</td>
<td>70 30</td>
<td>53 47</td>
</tr>
<tr>
<td>Guidelines Selected</td>
<td>70 30</td>
<td>35 24</td>
</tr>
</tbody>
</table>

The results show a significant decrease in the correct Path Selected (p=0.03, p < .05) and the Guidelines Selected (p=0.03, p < .05) metrics for novice nursing students. The results show no significant decrease in the correct Path Selected (p=0.38, p > .05) and the Guidelines Selected (p=0.23, p > .05) metrics for expert nursing students. These results confirm Hypothesis 5.

4.8 Discussion

This section presents a discussion of the results based on the expertise variable that defined the user context parameter. This variable was selected in the Select step in Phase 1 of the methodology. The results in Sections 4.7.1 and 4.7.2 demonstrated that the selected change to the lightweight MeSH ontology had a significant impact only on the novice participants, but did affect all subgroups for the Guidelines Selected metric causing worse performance in the search application task.
The discussion also takes into account the results of the post-task surveys administered to study participants immediately after the completion of the search application task during Phase 1 and Phase 2, which are described in Section 4.2.1.1.2 Confirmation of Expertise.

4.8.1 Novice Participants

The results suggest that novices obtained better support from the unchanged lightweight MeSH ontology. The results confirm that an ontology detrimentally changed at the basic level provides poor support to novices that rely on the knowledge representation of the ontology to perform the search application task. For the novice participants, Hypothesis 1 for novice resident physicians and Hypothesis 3 for novice nursing students were confirmed by the results for both metrics. The confirmed hypotheses indicate that, for novices, a relationship exists between changes to the structural characteristics of depth, breadth, and fanoutness of the lightweight MeSH ontology and user expertise associated through the basic level. The results were tested and found to be statistically significant. Statistical significance implies that the selected structural characteristics, the selected ontology change, and the context parameters can be used in concert to indicate the performance of the ontology in the tested context and a guideline can be created and used for future ontology engineering as depicted in Figure 5-1.

To elaborate on the experimental results based on the post-task surveys, the novice resident physician subgroup was shown to have a higher self-assessment of knowledge about the topics presented in the medical scenarios averaging 46.15% for Phase 1 and Phase 2 combined. Novice nursing students had a lower self-assessment of knowledge
about the topics presented in the medical scenarios averaging 40.00% for Phase 1 and Phase 2 combined. Novice resident physicians also had a higher result for interpreting the scenario topic for the two tasks at 11 correct scenario interpretations versus 10 correct scenario interpretations for novice nursing school students.

The difference in actual performance for both the Path Selected and the Guidelines Selected metrics between novice resident physicians and novice nursing students can be attributed to a slightly higher presentation of Scenario 1 – related to the topic of selecting the correct intervention to decrease the incidence of Ventilator-associated Pneumonia in the intensive care unit. Novice resident physicians received the topic for 7 of 13 random scenario presentations and novice nursing school students received the topic for 8 of 10 random scenario presentations. This topic is somewhat more familiar to nursing students as it is closely aligned with the nursing scope of practice as demonstrated in Appendix G: Post Task Survey Answers.

In Phase 1, for every case where novice nursing students had a correct Path Selected metric result they had a correct Guidelines Selected metric result, signifying that the selection of the correct path; and thus the correct search keywords typically led to the correct identification of guidelines.

Novice resident physicians had an incorrect Path Selected metric but a correct Guidelines Selected metric result. The resident physician with this result had “Unfamiliar” selected for Q2: How familiar are you with the topic described in the scenario?, but was able to correctly identify the Scenario 1 topic by answering “looking for prevention and vap” [Ventilator-associated Pneumonia] for question Q4: Briefly explain what you were looking for to determine if the guidelines you selected were
relevant to the scenario. In cases where there was a correct Path Selected metric result and an incorrect Guidelines Selected metric result, the novice resident physician selected “Unfamiliar” for Q2 and was not able to correctly determine the Scenario 1 topic based on his answer to Q4 of “Evidence based recommendations.” This is a typical result where poor knowledge of a topic and the inability to find the correct path and keywords leads to the inability to find correct guidelines.

In Phase 2, novice nursing students had two (2) incorrect Path Selected metric and correct Guidelines Selected metric results. This is an atypical result because the keywords determined by the incorrect Path Selected metric make it difficult to find the correct guideline.

The first nursing student with this result was close to the correct conceptual path with the eventual keywords of “Cerebral Hemorrhage, Traumatic”. The correct hierarchy path leads to the “Coma” keyword, which results in the most relevant guidelines returned for the topic. The nursing student with this result selected “Neutral” for Q2, but was able to correctly interpret the Scenario 2 topic as evidenced by the answer “Anything related to assessment, monitoring of brain hemorrhage s/p traumatic injury.” The student indicated that he was not able to find the correct keywords in the hierarchy based on his inability to find the correct path and answer of “No” to Q3: Were you able to find the keywords you wanted in the hierarchy? However, the student demonstrated the knowledge of the topic by looking for keywords of “Monitor, ICP, intracranial pressure,” in his answer to question Q3.1: If you answered No, which keywords were you looking for?

The second nursing student was closer to the correct path and keyword selection with the selection of the path leading to “Tuberculosis”. The correct path would have led to
the keyword “Tuberculosis, Pulmonary”, which is a subordinate category of Tuberculosis. The nursing student with this result selected “Neutral” for Q2, but again was able to correctly interpret the Scenario 3 topic based on the answer “Skin testing, treatment, and control of possible TB” for Q4.

Resident physicians had one incorrect Path Selected metric and a correct Guidelines Selected metric result. The resident physician with this result had selected the hierarchy path leading to “Trauma, Nervous System” as the keyword. The correct path would lead to the “Coma” keyword. The resident had “Familiar” selected for Q2 and correctly identified the Scenario 2 topic by answering Q4 “…I was searching for the guidelines required to monitor the patient with acute head injury. If any hematoma is causing raised intracranial tension…” Again, the knowledge of the topic demonstrated by looking for guidelines with intracranial pressure monitoring helped the novice resident to identify the correct guidelines from the returned results based on a path selected that was not optimal.

Overall, novices obtained better support from the unchanged lightweight MeSH ontology confirmed by the significant results. In most cases the change to the basic level of the ontology, and thus its knowledge representation, was detrimental to novices that rely on that knowledge to perform the search application task. However, specific individual knowledge of the scenario allowed a few novices to compensate for the poor ontological representation and obtain the correct final answer to the scenario questions.

4.8.2 Expert Participants

For the experts, the detrimental change at the basic level of the ontology decreased the performance for one subgroup of experts for the Path Selected metric and both subgroups for the Guidelines Selected metric. However, the results did not confirm that
an ontology detrimentally changed at the basic level negatively impacts experts performing the search application task. Hypothesis 2 for expert resident physicians and Hypothesis 4 for expert nursing students were initially supported by results for the \textit{Guidelines Selected} metric. For the \textit{Path Selected} metric only Hypothesis 4 for expert nursing students was initially supported by the results. The results were tested and found not to be statistically significant. Therefore, while a relationship appears to exist between changes to the structural characteristics and ontology performance it cannot be verified based on the results as both Hypothesis 2 and Hypothesis 4 were not confirmed. The relationship cannot be used to indicate the performance of the ontology in the tested context in the future as shown in Figure 5-1.

The post-task surveys support the results for expert participants. The expert resident physician subgroup has the highest self-assessment of knowledge about the topics presented out of all the participant groups averaging 65.63\% for Phase 1 and Phase 2 combined. The expert nursing student subgroup had the second highest self-assessment of knowledge with a 47.06\% average rating for Phase 1 and Phase 2 combined. Both the expert resident physician subgroup and the expert nursing student subgroup had higher results than novice subgroups for interpreting the scenario topic for the two phases at 18 and 17 correct scenario interpretations respectively.

In Phase 1, for every correct \textit{Guidelines Selected} metric result, expert resident physicians had a correct \textit{Path Selected} metric result. Expert nursing students had two (2) instances of incorrect \textit{Path Selected} metric and correct \textit{Guidelines Selected} metric results.
The first expert nursing student had followed the hierarchy path leading to the keyword “Bacterial Infections and Mycoses.” The correct hierarchy path would lead to the “Pneumonia, Ventilator-Associated” keyword. This nursing student selected “Unfamiliar” for Q2, but was able to correctly identify the Scenario 1 topic by answering “… the guideline specifically addressing hospital acquired pneumonia specifically related to ventilators” for question Q4.

The second nursing student with this result followed the path with the eventual keywords of “Ventilator-Induced Lung Injury”, which is a superordinate category on the path leading to the correct keyword of “Pneumonia, Ventilator-Associated.” This nursing student selected “Familiar” for Q2 and correctly identified the Scenario 1 topic by answering Q4 with “interventions regarding hospital-acquired ventilator pneumonia.”

In Phase 2, both the expert resident physician group and the expert nursing student group selected the correct Guidelines Selected metric result for a correct Path Selected metric result. In the case of correct Path Selected metric results and incorrect Guidelines Selected metric results, both groups had missed the topic of the scenario. An example is an expert resident physician misidentifying the topic as Pneumonia instead of Tuberculosis for Scenario 3 as evident by his answer of “I was looking for guidelines for pneumonia in varying clinical settings” for Q4. For the expert nursing student group, these results reflected a lack of knowledge about the topics by the “Unfamiliar” to “Neutral” responses to Q2 for all nurses that had this result, although some were able to get close to the scenario topics by their answers to Q4 such as “hematoma cranial” for Scenario 2 and “I was looking for differential diagnosis for pneumonia vs. TB” for Scenario 3.
Overall, knowledge of the scenario allowed most experts to compensate for the poor ontological representation and obtain the correct final answer to the scenario questions.

4.8.3 Results for the Groups Overall

Participants in our study were either experts or novices in order to account for our examination of the impact of lightweight ontologies in human expertise. Figure 4-9 depicts the combined results of the four participant subgroups with expert and novice user performance for the search application task supported by the unchanged and changed lightweight MeSH ontologies using the Path Selected and Guidelines Selected metrics.

The results from the studies show that there is a significant impact on the performance of an ontology-supported task by novices when that ontology is changed to the detriment of its ability to support that task with respect to expertise. Removing the ontology layer containing the superordinate concepts to the basic layer breaks the link to the disease etiology. The change to the ontology had the greatest impact on novice nursing students, who in general were the least knowledgeable subgroup about the specific medical conditions in the scenarios presented, followed by novice residents, expert nursing students, and then expert resident physicians. Overall, the change to the ontology had a greater impact on the performance of the search application task for novices than experts as demonstrated by the confirmation of Hypothesis 5 for both the resident physician and nursing student participant groups.
These results may be attributable to the type of training conducted within the two participant groups. Physician training has a greater focus on diagnostic skills for the identification of disease and its appropriate treatment as compared to training in nursing. Because the lightweight MeSH ontology is based on a taxonomy of disease, a change that made it more difficult to link the symptoms in the scenario to the disease impacted nurses to a greater extent than physicians. Expert physicians were able to compensate for the lack of knowledge during the navigation through the ontology, and had the smallest decrease in the final task of identifying guidelines appropriate to the medical scenario. There are two ways to look at expertise: as knowledge of the general domain and as knowledge of a particular topic. In the knowledge of the domain, the subgroups’ rated their knowledge as expected when averaged for the subgroup. In the knowledge of a
particular topic, the results show that some individuals have specific knowledge of the topic presented in the medical scenario helping them to boost their performance for the search application task for that specific scenario.

Chapter 5 will describe how the methodology introduced in Chapter 3 and implemented in studies in a healthcare context in Chapter 4 answers the research questions and meets our overall goals.
CHAPTER 5: Contributions, Limitations, and Future Work

This chapter describes the contributions of the research, the limitations, and future work. Section 5.1 elaborates the five contributions based on the research questions, the proposed methodology, and the knowledge representation of lightweight ontologies. Section 5.2 describes limitations of the research. Section 5.3 describes future research that will build upon and expand the scope of the current research and address limitations.

The diagram in Figure 5-1 depicts the reasoning flow by which the methodology answers the research questions. To reiterate, the research questions were:

- **R1:** Is there a relationship (either positive or negative) between characteristics and performance for a given context?
- **R2:** Can ontology characteristics be used as indicators of ontology quality?
- **R3:** Can we develop guidelines for ontology engineering and evaluation taking context into account?
The methodology developed for this research employs a standard hypothesis in each implementation:

\[ H: \text{A change in the selected ontology characteristics } r_j \text{ has an impact on the performance } p \text{ of the application task for defined context } c_i. \]

If there is a change in performance – either positive or negative – of the application task based on the desired change to the values of the selected ontology characteristics, then the hypothesis is true and question R1 is answered: there is a relationship. If the results show statistical significance, then questions R2 and R3 are answered: ontology characteristics can be used as indicators of ontology quality for the given context, and guidelines for ontology engineering and evaluation can be built for the context. With statistical significance question R2 is answered because the difference in performance is not attributable to chance. When question R2 is answered with statistical significance, the results can be used to develop a guideline answering R3. If the results do not show
statistical significance, the difference in performance can be attributable to chance. In this case questions R2 and R3 are not fully answered and the selected characteristics should not be used as indicators of ontology quality, nor used for developing guidelines.

If there is no change in performance of the application task based on the change in values of the selected characteristics, then the alternative hypothesis is false and question R1 is answered: there is no relationship between the selected characteristics for the given context. The selected characteristics cannot be used as indicators of ontology quality and guidelines cannot be developed.

5.1 Contributions

This research has five contributions. The first three contributions relate to the research questions and our goal of finding a potential relationship between ontology characteristics and context parameters, specifically structural characteristics and the user context parameter defined by the expertise variable. The forth contribution relates to the proposed methodology. The fifth contribution relates to the ability of a lightweight ontology to represent knowledge.

5.1.1 R1: Is there a relationship between characteristics and performance for a given context?

The research demonstrated that a relationship does exist between structural ontology characteristics and the user context parameter defined by the expertise variable. This relationship was found to be directional and negative for the novice groups and the expert nursing student subgroup. A detrimental change made to the structural characteristics of the ontology that impacts expertise and lessens the ability of the ontology to represent


knowledge decreased the performance of the search application task in the medical domain for users that rely on that knowledge to be successful.

Ontology engineering and evaluation approaches are based on ontology characteristics. Table 2-1 and Table 2-2 show the ontology characteristics utilized by authors of ontology evaluation approaches. Structural characteristics tested by our methodology can be utilized in similar future ontology development efforts to determine a possible evaluation approach to use when the user context parameter is defined by expertise. When a relationship with a statistically significant difference in performance is found, the approach utilizing those characteristics can be selected for ontology evaluation. For example, in our studies changes to structural characteristics were found to have a statistically significant impact on performance of users differentiated by expertise. Therefore, the evaluation framework developed by Gangemi, et al. (2006) can be applied because it uses these structural characteristics.

5.1.2 R2: Can ontology characteristics be used as indicators of ontology quality?

When a relationship with a statistically significant difference in performance is found the approach utilizing those characteristics can be selected for ontology engineering and evaluation. A significant result means that the difference in performance is attributable to the change. Therefore, this combination of structural characteristics and the user context parameter can be used as an indicator of ontology quality. This is especially true if the ontology will be used to support an application in the same context. The same combination of characteristics and context can provide a priori knowledge about the quality of the same or similar task.
The results demonstrated that a change in structural characteristics that negatively impacts the knowledge representation of the ontology negatively impacts its ability to support an application task. This further supports the idea that characteristics can be used to indicate ontology quality; with the only requirement being its use in a similar context.

In our research the results show that novice users, who are non-expert users with less knowledge about a domain, are negatively impacted in their ability to perform a task supported by an ontology whose structural characteristics were changed to affect the ability of the ontology to represent knowledge. The change was made to the basic level of the ontology hierarchy. We can therefore infer that an ontology with a poorly constructed on insufficient basic level as reflected by the calculation of structural characteristics will perform poorly in supporting search for novice users.

5.1.3 R3: Can we develop guidelines for ontology engineering and evaluation taking context into account?

As previously explained through Figure 5-1, the results of the implementation of the methodology must have a difference in performance on the application task for the selected quality metric, and the difference in performance has to be statistically significant.

A guideline can be created for any combination of selected ontology characteristics (that describe the ontology), context parameters (that describe the context), and change (that describes the result to the characteristics and impact on the selected context parameter). The guideline can only be created where there is a statistically significant difference in performance measured by the quality metrics. An example of the above
components for an ontology engineering guideline is depicted in Figure 5-2. All components are required for a guideline to be useful for engineering and evaluation.

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Change</th>
<th>removal of the superordinate level above the basic level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Selected Ontology Characteristics</td>
<td>depth, breadth, and fanoutness</td>
</tr>
<tr>
<td>Result</td>
<td>break in the disease etiology and flattening of the hierarchy</td>
<td></td>
</tr>
<tr>
<td>Impact</td>
<td>detrimental</td>
<td></td>
</tr>
<tr>
<td>Context Parameters</td>
<td>User</td>
<td>novices</td>
</tr>
<tr>
<td></td>
<td>Application Task</td>
<td>search</td>
</tr>
<tr>
<td></td>
<td>Domain</td>
<td>medical guidelines</td>
</tr>
</tbody>
</table>

Figure 5-2: Guideline components

The components can be combined into the following guideline in paragraph form:

**Removal of the superordinate level above the basic level** in the ontology that impacts **depth, breadth, and fanoutness** resulting in a **break in the disease etiology** and flattening of the **hierarchy** will be **detrimental** to **novices** performing a **search** in the **medical guidelines**.

5.1.4 Methodology to Test the Relationship

The methodology utilized in this research was successfully implemented. The methodology allows for repetitive testing of ontology characteristics across various contexts. Most importantly, it also allows for the inclusion of ontology data at a more granular level to that of other methods. This lower level granularity allows linking of specific ontology characteristics to specific context parameters and parameter variables.

The results could be appended to currently available ontology resources such as AKTiveRank (Alani & Brewster, 2006), ontology search engines (e.g. Swoogle),
websites (e.g. semanticweb.org), and repositories (e.g. OwlSeek.com) to improve selection of ontologies for a specific context.

5.1.5 Lightweight Ontologies as Knowledge Representation

The selected change was directed at the context user parameter represented by the expertise variable. The change in the ontology was related to knowledge representation. The change was selected to impact the ability of the lightweight MeSH ontology to represent knowledge. Experts possessing domain knowledge were not significantly impacted by the change while novices who do not have domain knowledge were significantly impacted. This was confirmed across two participant groups: resident physicians and nursing students.

From the results presented in this research, we demonstrate that lightweight ontologies do serve as a replacement for human expertise in a context of use. This is important because many of the current taxonomy and is-a hierarchies on the Web, examples of lightweight ontologies, are supporting novices in performing tasks that might require expertise. The studies show that changes to characteristics of ontologies can impact their support of an application task for a context for which they were designed. Specifically, we demonstrated that less complete knowledge structures will interfere with the ability of novice users, who lack expertise about the domain, to perform a task in that domain supported by the knowledge structure.
5.2 Limitations

This section examines several limitations to the research originating from the elements of the methodology operationalized as a user performing an application task supported by a lightweight ontology in some domain.

5.2.1 Human Participation in the Methodology

The methodology developed for this research requires human participation. Human participation limits the applicability of the methodology as it may be difficult to find a homogeneous group of participants for every context. The correct number of participants to test statistical significance may also be difficult to determine and find. The difference in performance of the participants on the application task influences the statistical calculation and the required number of participants.

5.2.2 Distinction Between Novice and Expert Participants

The participants in the studies included resident physicians and nursing students. These groups embody the user context parameter. Within the two groups, participants were differentiated by the expertise variable with the values of novice and expert resulting in participant classifications of novice resident physician, expert resident physician, novice nursing student, and expert nursing student. Supporting research demonstrates that both resident physicians (Doerschug et al., 1999) and nursing students (Ericsson et al., 2007) can be differentiated by expertise based on their duration of training, especially deliberate practice under supervision as normally occurs in residency and university programs. Besides the research, the difference in nursing student expertise was also verified by faculty from the nursing school.
In practice however, the exact division between expert and novice is not easily discernible. Although the overall distinction in the groups between novice and expert was generally shown to be valid by the post-task surveys as described in Section 4.2.1.2 they were not statistically significant. Individual expertise in the resident physician group skewed the results for post-task survey Question 3. The experimental results confirmed that expertise in general improves performance but made it difficult to cleanly differentiate participant classifications of novice resident physician, expert resident physician, novice nursing student, and expert nursing student. The results of the post-task surveys showed a definite trend towards the correct separation of the subgroups by the expertise variable; however, additional participants would be required to get statistical significance based on the results of the post-task surveys. Due to the difficulty of obtaining human participants as described in Section 5.2.1 and because of the statistical significance obtained for the experimental results as described in Section 4.7 we chose to limit the number of participants in the studies as described in Section 4.2.1.1.

With respect to the two studies, the subgroup results differed based on the expertise of individual participants about the scenario topic as well. Nevertheless, the overall decrease in performance for the novice groups was greater than the overall decrease in performance for the selected metrics for the expert groups and was statistically significant. The distinction was made clearer by participant answers to the post-task survey questions as explained in Section 4.8.

5.2.3 Application Task

The studies include only one application task. The choice to utilize the search task in the studies was based on a common on-the-job task performed by healthcare
professionals. Although not as common, the converse task of classification and its complement task of annotation are other tasks commonly performed in the healthcare context, and in many other contexts. If the research would have included two tasks, such as search and annotation, the results would be more readily applicable to a variety of contexts. However, a secondary task implemented in the same studies or as additional studies would have the effect of decreasing study participation in case of the former or extending the research in case of the latter.

Additionally, selecting a task such as annotation would have required the involvement of expert reviewers in the studies to a much greater extent. The search application task results are based on the selection of predetermined guidelines from a defined set in the domain. Only one review was required by experts to determine the correct set of guidelines which could be verified for all participants in the studies. The annotation task would have required the review of all participant performance by experts for each document to determine correct annotation.

5.2.4 Lightweight Ontology

The studies use a lightweight ontology to support the search application task. The research did not include a heavyweight ontology. The lightweight ontology was sufficient to demonstrate the relationship between the selected ontology characteristics and the selected context parameter. Lightweight ontologies are currently more widely applied across the Internet. Lightweight ontologies are also a backbone for all heavyweight ontologies; consequently results based on a lightweight ontology can be applied to a heavyweight ontology. Using a lightweight ontology enables the transfer of
the results to other similar knowledge representation structures that are not ontologies but have the same characteristics or similar attributes.

5.2.5 Characteristics

The studies conducted in this research tested structural characteristics represented by the measures of depth, breadth, and fanoutness. There are a number of other characteristics, especially non-structural characteristics, which can be utilized in the studies from 22 available criteria and 30 available measures. Using similar selection principles of the characteristic being well defined, easy to understand and apply, and applicable to the majority of ontologies and many hierarchical structures we can select a non-structural criterion of Clarity. As defined by Gruber (1993), Clarity reflects the ontology’s ability to effectively communicate the intended meaning of defined terms. The variation of the definition of the concept, or the concept term, can be seen as important to the ability of the ontology to support an application task. The user performing an application task supported by an ontology can have a different culture, background, age, literacy level, or as in this research expertise. The differences will have an impact on the interpretation of either the definition or the term for a concept, and will affect the interpretation and selection of query terms in the case of a task such as search.

The selection of the tested structural characteristics of depth, breadth, and fanoutness was deemed more important to the ontology because of their central role in the formation of the ontology is-a backbone over other characteristics. Furthermore, the selection of inadequate concept terms can be mitigated more easily than an inferior arrangement of the terms into the physical hierarchical structure. However, additional study into this area is warranted.
5.2.6 Impact of Changes on Characteristics

The studies conducted in this research include only one type of change. The selected change in the study was made to have an impact on the expertise variable defining the user parameter while at the same time having an impact on the selected structural characteristics and the search application task. A systematic change can be made to the hierarchy structure to have a similar impact, for example through pruning ontology branches or through creating a new ontology. However, these changes are not expected to have any specific impact on the expertise variable. It would then be difficult to understand the cause for the difference in performance for the two selected metrics. The methodology recommends that the change should be based on the selected context parameter, task, and ontology characteristics so that an actionable guideline can be created if the results are statistically significant.

The research could include an additional structural characteristic represented by a measure such as tangledness. Tangledness defines the interconnection of concepts through other relationships besides the is-a hierarchy relationship. Adding an additional structural characteristic increases the possibility of confounding the study results because it would be an independent change to the change selected in the study.

Some changes to the ontology characteristics cannot be done independently. In the case of this research, the selection of the breadth, depth, and fanoutness measures limited the possibility of making an independent change to a single measure without changing the ontology. The selected change resulted in a simultaneous update to all three measures. As long as the guideline is made for the three measures in aggregate, then the impact of non-independent results is minimized. This limitation will be true for most
structural characteristics as they are based on the physical properties of the ontology which are interconnected. Future research can mitigate this issue by selecting a criterion or any of the non-structural measures that are not interconnected.

5.2.7 Application in the Medical Domain

The studies in this research were conducted in a medical domain. Performing additional studies in other domains would provide for better generalization of the research from the additional results and would demonstrate the diverse applicability of the methodology. In order to conduct the studies in another domain, but keep the research methodology consistent, we would have to equalize the two other context parameters of user and application task. The user parameter was differentiated by the expertise variable with the values of novice and expert. The values were defined in the studies by the duration of training that was supported by research specific to the medical domain. It would be difficult to find a participant population in another domain with the same type of training program and incorporate the results into a unified analysis. Additionally, the search application task presents a problem because of the dependence on the medical scenario, which would be difficult to replicate in an analogous way in another domain for the purpose of including the results in this research. We therefore selected to limit the studies to the same domain and select two different groups with a similar background and expertise but different training to differentiate the context but keep the focus on the same context parameter.
5.3 Future Work

The research presented in this dissertation can be broadened to provide a more comprehensive implementation of the methodology and lessen the effect of the limitations of the studies.

5.3.1 Application Task

We plan to expand the application task to include annotation. Physicians commonly annotate electronic medical records with codes, for example the International Classification of Diseases - Clinical Modification codes (ICD-9-CM and ICD-10-CM)\(^7\), to assign diagnostic and procedure codes for physician’s office and outpatient visits and inpatient stays. The codes are also utilized for billing purposes and epidemiology and health management.

The MeSH or the ICD codes hierarchy can be used to categorize medical scenarios. The hierarchy would be changed following the same criteria as the previously selected change to the characteristics. This will provide a direct or near direct comparison between two application tasks within the same domain.

5.3.2 Characteristics

We plan to test non-structural characteristics. In order to maintain the same study implementation the candidate non-structural characteristic will be selected based on its relationship to expertise. Correct recognition and interpretation of the concept name is an expertise related action. The studies would vary the concept name with various synonymous terms with a variance in their recognition between novice and expert users to determine the impact on the search application task in the same domain.

\(^7\) http://www.cdc.gov/nchs/icd.htm
5.3.3 Change of Context

Finally, we plan to execute a similar study outside of the healthcare field. Conducting a study utilizing the user, defined through the expertise variable, performing a search application task based on a scenario in an alternative domain will allow for better generalization of the research. The study will allow for the comparison between more varied contexts, and provide a better foundation for the guideline based on the results of the two studies with resident and nursing student participants.
List of References


### Appendix A: Candidate Criteria and Measures

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Measure Type</th>
<th>Measure or Criteria</th>
<th>Calculation / Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gangemi et al.</td>
<td>Structural</td>
<td>Depth</td>
<td>Average arc path length for ‘is a’ arcs</td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>Breadth</td>
<td>Average number of ‘is a’ arcs per node</td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>Tangledness</td>
<td>Average number of relationships other than inheritance relationships per node</td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>Fanoutness</td>
<td>Total number of arcs from all nodes as compared to the total number of nodes with more than one ‘is a’ arc</td>
</tr>
<tr>
<td></td>
<td>Structural</td>
<td>Density</td>
<td>The number of arcs from a selected leaf node set as compared to the total number of arcs from all nodes</td>
</tr>
<tr>
<td>Tatir et al.</td>
<td>Schema</td>
<td>Relationship richness</td>
<td>The ratio of the number of (non-inheritance) relationships (P), divided by the total number of relationships defined in the schema</td>
</tr>
<tr>
<td></td>
<td>Schema</td>
<td>Attribute richness</td>
<td>The average number of attributes (slots) per class;</td>
</tr>
<tr>
<td>Researcher</td>
<td>Measure Type</td>
<td>Measure or Criteria</td>
<td>Calculation / Implementation</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>computed as the number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>attributes for all classes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>divided by the number of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>classes</td>
</tr>
<tr>
<td>Schema</td>
<td>Inheritance richness</td>
<td>The average number of subclasses per class</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>Class richness</td>
<td>The percentage of the number of non-empty classes (classes with instances) divided by the total number of classes defined in the ontology schema</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>Class Importance</td>
<td>The percentage of the number of instances that belong to the inheritance subtree rooted at a class in the ontology compared to the total number of class instances in the ontology</td>
<td></td>
</tr>
<tr>
<td>Instance</td>
<td>Class Connectivity</td>
<td>The total number of relationships instances of the class have with instances of other classes</td>
<td></td>
</tr>
<tr>
<td>Gruninger and Fox</td>
<td>Criterion: Not Applicable</td>
<td>Competency</td>
<td>Correctly answering competency questions related to the ontology and assessing the number successfully answered</td>
</tr>
<tr>
<td>Researcher</td>
<td>Measure Type</td>
<td>Measure or Criteria</td>
<td>Calculation / Implementation</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Guarino and Welty</td>
<td>Not Applicable</td>
<td>Correctness (Essence, Rigidity, Identity, Unity)</td>
<td>Use the OntoClean method to verify essence (property that must hold true for an example of the entity), rigidity (property that must hold true for all the instances of an entity), identity (property of recognizable distinct individual entities), and unity (property of a recognizable group of parts that form an individual entity) with a Yes/No for each property</td>
</tr>
</tbody>
</table>
## Appendix B: Ontology Resources

### Ontology Libraries

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Text Description from Home Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAML Ontology Library</td>
<td><a href="http://www.daml.org/ontologies/">http://www.daml.org/ontologies/</a></td>
<td>Web page listing ontologies that can be sorted by various criteria, such as by keyword, by class, by property, by submission date, etc.</td>
</tr>
<tr>
<td>Protégé Ontology Library</td>
<td><a href="http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library">http://protegewiki.stanford.edu/index.php/Protege_Ontology_Library</a></td>
<td>Wiki page listing submitted ontologies in various available formats. Provides the ontology name and description, along with links to ontology related plug-ins.</td>
</tr>
<tr>
<td>SchemaWeb</td>
<td><a href="http://www.w3.org/wiki/SchemaWeb">http://www.w3.org/wiki/SchemaWeb</a></td>
<td>SchemaWeb is a directory of RDF schemas expressed in the RDFS, OWL and DAML+OIL schema languages. It provides a comprehensive directory of RDF schemas to be browsed and searched by human agents and also an extensive set of web services to be used by software agents that wish to obtain real-time schema information whilst processing RDF data.</td>
</tr>
<tr>
<td>OBO Foundry / Bioportal</td>
<td><a href="http://obofoundry.org">http://obofoundry.org</a></td>
<td>A collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain.</td>
</tr>
<tr>
<td>Name</td>
<td>Link</td>
<td>Text Description from Home Pages</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Semantic Web Wiki</td>
<td>[<a href="http://semanticweb.org/wiki/Main">http://semanticweb.org/wiki/Main</a> Page](<a href="http://semanticweb.org/wiki/Main">http://semanticweb.org/wiki/Main</a> Page)</td>
<td>A wiki page that lists and describes ontologies and related tools, people, and events.</td>
</tr>
<tr>
<td>OBO (Open Biomedical Ontologies) Foundry</td>
<td><a href="http://obo.sourceforge.net">http://obo.sourceforge.net</a> or <a href="http://www.obofoundry.org/">http://www.obofoundry.org/</a></td>
<td>A collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain.</td>
</tr>
</tbody>
</table>
## Semantic Search Engines

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swoogle</td>
<td><a href="http://swoogle.umbc.edu/">http://swoogle.umbc.edu/</a></td>
<td>Swoogle is a search engine for the Semantic Web on the Web. Swoogle index only Semantic Web documents, currently including those written in RDF/XML, N-Triples, N3(RDF) and some documents that embed RDF/XML fragments.</td>
</tr>
<tr>
<td>ONTOSEARCH2</td>
<td><a href="http://ontosearch.org/">http://ontosearch.org/</a></td>
<td>ONTOSEARCH2 is a project at the University of Aberdeen Computer Science Department to help facilitate reuse of ontologies on the Semantic Web by allowing structures searches of a large database of ontologies and other semantic web documents.</td>
</tr>
<tr>
<td>Ontaria</td>
<td><a href="http://www.w3.org/2004/ontaria/">http://www.w3.org/2004/ontaria/</a></td>
<td>A searchable and browsable directory of semantic web data. Our focus is RDF vocabularies with OWL ontologies, but all the RDF data we index is visible. The site is primarily intended for people creating RDF content who want to better understand which vocabularies are available and how they are being used.</td>
</tr>
</tbody>
</table>
Appendix C: Sample Guideline from the National Guideline Clearinghouse

Guideline Summary NGC-6807

Guideline Title

Strategies to prevent ventilator-associated pneumonia in acute care hospitals.

Bibliographic Source(s)


Guideline Status

This is the current release of the guideline.

Scope

Disease/Condition(s)

Ventilator-associated pneumonia (VAP)

Guideline Category

Prevention
Risk Assessment

Clinical Specialty

Critical Care
Infectious Diseases
Internal Medicine
Nursing
Preventive Medicine
Pulmonary Medicine
Surgery
**Intended Users**

Advanced Practice Nurses
Allied Health Personnel
Hospitals
Nurses
Physician Assistants
Physicians
Respiratory Care Practitioners
Utilization Management

**Guideline Objective(s)**

To highlight practical recommendations in a concise format designed to assist acute care hospitals in implementing and prioritizing their ventilator-associated pneumonia (VAP) prevention efforts

**Target Population**

Patients in acute care hospitals with an endotracheal tube for mechanical ventilation

**Interventions and Practices Considered**

1. Basic practices for prevention and monitoring of ventilator-associated pneumonia (VAP) including:
   - Healthcare personnel education
   - Surveillance of VAP
   - Implementing practices for disinfection and sterilization of respiratory equipment, maintaining patients in a semirecumbent position, antiseptic oral care
   - Assignment of accountability

2. Special approaches for prevention of VAP in hospitals with unacceptably high rates of VAP including:
   - Using an endotracheal tube with in-line and subglottic suctioning for all eligible patients
   - Ensuring that all intensive care unit (ICU) beds used for patients undergoing ventilation have a built-in tool to provide continuous monitoring of the angle of incline

The following approaches should not be considered a routine part of VAP prevention:

- Routine administration of intravenous immunoglobulin, white-cell-stimulating factors (filgrastim or sargramostim), enteral glutamine, or chest physiotherapy
- Routine use of rotational therapy with kinetic or continuous lateral rotational therapy beds
- Routine administration of prophylactic aerosolized or systemic antimicrobials

**Major Outcomes Considered**

- Incidence density of VAP, reported as the number of episodes of VAP per 1,000 ventilator-days
- Length of hospitalization (days)
- Length of mechanical ventilation (days)
- Morbidity
- Mortality
- Healthcare resource utilization
Methodology

Methods Used to Collect/Select the Evidence

Searches of Electronic Databases

Description of Methods Used to Collect/Select the Evidence

For this compendium, the Society for Healthcare Epidemiology of America/Infectious Diseases Society of America (SHEA/IDSA) reviewed previously published guidelines and recommendations relevant to each section and performed computerized literature searches using PubMed. Searches of the English-language literature focused on human studies published after existing guidelines through 2007, using the subject headings listed in Table 2 of the Compendium document (see "Availability of Companion Documents" field).

Number of Source Documents

Not stated

Methods Used to Assess the Quality and Strength of the Evidence

Weighting According to a Rating Scheme (Scheme Given)

Rating Scheme for the Strength of the Evidence

Quality of Evidence*

I. Evidence from ≥1 properly randomized, controlled trial
II. Evidence from ≥1 well-designed clinical trial without randomization, from cohort or case-controlled analytic studies (preferably from ≥1 center), from multiple time-series studies, or from dramatic results of uncontrolled experiments
III. Evidence from opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees

*Adapted from the Canadian Task Force on the Periodic Health Examination.

Methods Used to Analyze the Evidence

Review of Published Meta-Analyses
Systematic Review

Description of the Methods Used to Analyze the Evidence

In evaluating the evidence regarding the prevention and monitoring of healthcare-associated infections (HAIs), the HAI Allied Task Force followed a process used in the development of other Infectious Diseases Society of America (IDSA) guidelines, including a systematic weighting of the quality of the evidence and the grade of recommendation (see the "Rating Scheme for the Strength of the Evidence" and "Rating Scheme for the Strength of the Recommendations" fields).

Methods Used to Formulate the Recommendations
Expert Consensus

**Description of Methods Used to Formulate the Recommendations**

Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA) Standards and Practice Guidelines Committee convened experts in the prevention and monitoring of healthcare-associated infections (HAIs).

The HAI Allied Task Force met on 17 occasions via teleconference to complete the compendium. The purpose of the teleconferences was to discuss the questions to be addressed, make writing assignments, and discuss recommendations. All members of the HAI Allied Task Force participated in the preparation and review of the draft documents. The compendium was then submitted to a subgroup of the HAI Allied Task Force with implementation expertise that, through a series of additional teleconferences and communications, performed extensive editing and reformatting to create implementation-focused text.

**Rating Scheme for the Strength of the Recommendations**

**Strength of Recommendation***

A. Good evidence to support a recommendation for use  
B. Moderate evidence to support a recommendation for use  
C. Poor evidence to support a recommendation

*Adapted from the Canadian Task Force on the Periodic Health Examination.

**Cost Analysis**

Guideline developers reviewed published cost analyses.

**Method of Guideline Validation**

External Peer Review  
Internal Peer Review

**Description of Method of Guideline Validation**

**Review and Approval Process**

A critical stage in the development process is peer review. Peer reviewers are relied on for expert, critical, and unbiased scientific appraisals of the documents. The Society for Healthcare Epidemiology of America/Infectious Diseases Society of America (SHEA/IDSA) employed a process used for all SHEA/IDSA guidelines that includes a multilevel review and approval. Comments were obtained from several outside reviewers who complied with the SHEA/IDSA policy on conflict of interest disclosure. In addition, 8 stakeholder organizations provided comments on the document. Finally, the guideline was reviewed and approved by the IDSA Standards and Practice Guidelines Committee and the Board of Directors of the SHEA and the IDSA prior to dissemination.

**Recommendations**

**Major Recommendations**
Recommendations for Implementing Prevention and Monitoring Strategies

Recommendations for preventing and monitoring ventilator-associated pneumonia (VAP) are summarized below. They are designed to assist acute care hospitals in prioritizing and implementing their VAP prevention efforts.

Each recommendation includes a ranking for the strength and the quality of evidence supporting it. Definitions of the levels of evidence (I-III) and grades of recommendation (A-E) are provided at the end of the "Major Recommendations" field.

Basic Practices for Prevention and Monitoring of VAP: Recommended for All Acute Care Hospitals

Education

1. Educate healthcare personnel who care for patients undergoing ventilation about VAP, including information about the following (A-II):
   - Local epidemiology
   - Risk factors
   - Patient outcomes
2. Educate clinicians who care for patients undergoing ventilation about noninvasive ventilatory strategies (B-III).

Surveillance of VAP

1. Perform direct observation of compliance with VAP-specific process measures (B-III).
   - VAP-specific process measures include hand hygiene, bed position, daily sedation interruption and assessment of readiness to wean, and regular oral care.
   - Use structured observation tools at regularly scheduled intervals.
2. Conduct active surveillance for VAP and associated process measures in units that care for patients undergoing ventilation who are known or suspected to be at high risk for VAP on the basis of risk assessment (A-II).
   - Collect data that will support the identification of patients with VAP and calculation of VAP rates (i.e., the number of VAP cases and number of ventilator-days for all patients who are undergoing ventilation and in the population being monitored).

Practice

1. Implement policies and practices for disinfection, sterilization, and maintenance of respiratory equipment that are aligned with evidence-based standards (e.g., guidelines from the Centers for Disease Control and Prevention and professional organizations) (A-II) (Tablan et al., 2004).
   - See the Appendix in the original guideline document for a list of recommended practices.
2. Ensure that all patients (except those with medical contraindications) are maintained in a semirecumbent position (B-II).
3. Perform regular antiseptic oral care in accordance with product guidelines (A-I).
4. Provide easy access to noninvasive ventilation equipment and institute protocols to promote the use of noninvasive ventilation (B-III).

Accountability

1. The hospital's chief executive officer and senior management are responsible for ensuring that the healthcare system supports an infection prevention and control program to effectively prevent VAP.
2. Senior management is accountable for ensuring that an adequate number of trained personnel are assigned to the infection prevention and control program.
3. Senior management is accountable for ensuring that healthcare personnel, including licensed
and nonlicensed personnel, are competent to perform their job responsibilities.

4. Direct healthcare providers (such as physicians, nurses, aides, and therapists) and ancillary personnel (such as housekeeping and equipment-processing personnel) are responsible for ensuring that appropriate infection prevention and control practices are used at all times (including hand hygiene, standard and isolation precautions, cleaning and disinfection of equipment and the environment, aseptic techniques when suctioning secretions and handling respiratory therapy equipment, patient positioning, sedation and weaning protocols, and oral care).

5. Hospital and unit leaders are responsible for holding their personnel accountable for their actions.

6. The person who manages the infection prevention and control program is responsible for ensuring that an active program to identify VAP is implemented, that data on VAP are analyzed and regularly provided to those who can use the information to improve the quality of care (e.g., unit staff, clinicians, and hospital administrators), and that evidence-based practices are incorporated into the program.

7. Personnel responsible for healthcare personnel and patient education are accountable for ensuring that appropriate training and educational programs to prevent VAP are developed and provided to personnel, patients, and families.

8. Personnel from the infection prevention and control program, the laboratory, and information technology departments are responsible for ensuring that systems are in place to support the surveillance program.

Special Approaches for the Prevention of VAP

Perform a VAP risk assessment. These special approaches are recommended for use in locations and/or populations within the hospital that have unacceptably high VAP rates despite implementation of the basic VAP prevention procedures listed above.

1. Use an endotracheal tube with in-line and subglottic suctioning for all eligible patients (B-II).
2. Ensure that all intensive care unit (ICU) beds used for patients undergoing ventilation have a built-in tool to provide continuous monitoring of the angle of incline (B-III).

Approaches That Should Not Be Considered a Routine Part of VAP Prevention

1. Do not routinely administer intravenous immunoglobulin (Tablan et al., 2004), white-cell–stimulating factors (filgrastim or sargramostim) (Tablan et al., 2004), enteral glutamine (Tablan et al., 2004), or chest physiotherapy (Tablan et al., 2004; Ntoumenopoulos et al., 2002). (A-III)
2. Do not routinely use rotational therapy with kinetic or continuous lateral rotational therapy beds (B-II) (Tablan et al., 2004; Goldhill et al., 2007).
3. Do not routinely administer prophylactic aerosolized or systemic antimicrobials (B-III) (Richards et al., 2000; Tablan et al., 2004; Hoth et al., 2003).

Unresolved Issues

1. Avoidance of H2 antagonist or proton pump inhibitors for patients who are not at high risk for developing gastrointestinal bleeding (Collard, Saint, & Matthay, 2003; Cook, 1995; Cook et al., 1996; Cook et al., 1998; Khan, Doctor, & Rubenfeld, 2006; Kantorova et al., 2004; Yildizdas, Yapicioglu, & Yilmaz, 2002; Levy et al., 1997)
2. Selective digestive tract decontamination for all patients undergoing ventilation (Liberati et al., 2004; van Nieuwenhoven et al., 2001; Bonten, 2006; de Jonge et al., 2003; Krueger et al., 2002; Silvestri et al., 2007
3. Use of antiseptic-impregnated endotracheal tubes (Pacheco-Fowler et al., 2004; Berra et al.,
4. Intensive glycemic control (Collier et al., 2005; van den Berghe et al., 2001; Toschlog et al., 2007; Brunkhorst et al., 2008)

Definitions:
Quality of Evidence*

I. Evidence from ≥1 properly randomized, controlled trial

II. Evidence from ≥1 well-designed clinical trial without randomization, from cohort or case-controlled analytic studies (preferably from ≥1 center), from multiple time-series studies, or from dramatic results of uncontrolled experiments

III. Evidence from opinions of respected authorities based on clinical experience, descriptive studies, or reports of expert committees

Strength of Recommendation*

A. Good evidence to support a recommendation for use

B. Moderate evidence to support a recommendation for use

C. Poor evidence to support a recommendation

*Adapted from the Canadian Task Force on the Periodic Health Examination.

Clinical Algorithm(s)

None provided

Evidence Supporting the Recommendations

References Supporting the Recommendations


Bonten MJ. Selective digestive tract decontamination--will it prevent infection with multidrug-resistant gram-negative pathogens but still be applicable in institutions where methicillin-resistant Staphylococcus aureus and vancomycin-resist enterococci are endemic?. Clin Infect Dis 2006 Sep 1;43 Suppl 2:S70-4. [35 references] PubMed


Richards MJ, Edwards JR, Culver DH, Gaynes RP. Nosocomial infections in combined medical-


**Type of Evidence Supporting the Recommendations**

The type of supporting evidence is identified and graded for each recommendation (see "Major Recommendations").

The recommendations in this guideline are largely based on previously published healthcare-associated infection (HAI) prevention guidelines available from a number of organizations, including the Healthcare Infection Control Practices Advisory Committee and the Centers for Disease Control and Prevention, Society for Healthcare Epidemiology of America (SHEA), the Infectious Diseases Society of America (IDSA), and the Association for Professionals in Infection Control and Epidemiology, and relevant literature published after these guidelines.

**Benefits/Harms of Implementing the Guideline Recommendations**

**Potential Benefits**

Appropriate use of ventilation leading to prevention of ventilator-associated pneumonia (VAP) in acute care hospitals
Potential Harms

Not stated

Qualifying Statements

Qualifying Statements

Recommendations that might ordinarily be included in a guideline with a C-level strength of recommendation were excluded from the recommendations and are discussed in the "unresolved issues" sections (see original guideline document); this was done to help hospitals to focus their implementation efforts on the most strongly recommended prevention practices. Hospitals can prioritize their efforts by initially focusing on implementation of the prevention approaches listed as basic practices recommended for all acute care hospitals. If healthcare-associated infection (HAI) surveillance or other risk assessments suggest that there is ongoing transmission despite implementation of basic practices, hospitals should then consider adopting some or all of the prevention approaches listed under the "special approaches" section of this document. These can be implemented within specific locations or patient populations or can be implemented hospital wide, depending on outcome data, risk assessment, and/or local requirements. Most of the special approaches listed in this document are supported by studies based on the control of HAI outbreaks and require additional personnel and financial resources for implementation.

The following issues remain unresolved:

- Avoidance of H2 antagonist or proton pump inhibitors for patients who are not at high risk for developing gastrointestinal bleeding
- Selective digestive tract decontamination for all patients undergoing ventilation
- Use of antiseptic-impregnated endotracheal tubes
- Intensive glycemic control

Implementation of the Guideline

Description of Implementation Strategy

An implementation strategy was not provided.

Implementation Tools

Audit Criteria/Indicators
Foreign Language Translations
Patient Resources

For information about availability, see the Availability of Companion Documents and Patient Resources fields below.

Institute of Medicine (IOM) National Healthcare Quality Report

Categories

IOM Care Need
Staying Healthy

**IOM Domain**

Effectiveness
Patient-centeredness
Safety

**Identifying Information and Availability**

**Bibliographic Source(s)**


**Adaptation**

Not applicable: The guideline was not adapted from another source.

**Date Released**

2008 Oct

**Guideline Developer(s)**

Infectious Diseases Society of America - Medical Specialty Society
Society for Healthcare Epidemiology of America - Professional Association

**Source(s) of Funding**

Society for Healthcare Epidemiology of America (SHEA)/Infectious Diseases Society of America (IDSA)

**Guideline Committee**

Healthcare-Associated Infections Task Force

**Composition of Group That Authored the Guideline**

*Primary Authors:* Susan E. Coffin, MD, MPH; Michael Klompas, MD; David Classen, MD, MS; Kathleen M. Arias, MS, CIC; Kelly Podgorny, RN, MS, CPHQ; Deverick J. Anderson, MD, MPH; Helen Burstin, MD; David P. Calfee, MD, MS; Erik R. Dubberke, MD; Victoria Fraser, MD; Dale N. Gerding, MD; Frances A. Griffin, RRT, MPA; Peter Gross, MD; Keith S. Kaye, MD; Evelyn Lo, MD; Jonas Marschall, MD; Leonard A. Mermel, DO, ScM; Lindsay Nicolle, MD; David A. Pegues, MD; Trish M. Perl, MD; Sanjay Saint, MD; Cassandra D. Salgado, MD, MS; Robert A. Weinstein, MD; Robert Wise, MD; Deborah S. Yokoe, MD, MPH

*Task Force Members:* David Classen, MD, MS; Infectious Diseases Society of America Co-Chair
Financial Disclosures/Conflicts of Interest

All members of the Healthcare-Associated Infections (HAI) Allied Task Force and the external peer reviewers complied with the Infectious Diseases Society of America (IDSA) policy on conflicts of interest, which requires disclosure of any financial or other interest within the past 2 years that might be construed as constituting an actual, potential, or apparent conflict. Members of the HAI Allied Task Force and the external reviewers were provided with the IDSA conflicts of interest disclosure statement and were asked to identify ties to companies developing products that might be affected by promulgation of the compendium. Information was requested regarding employment, consultancies, stock ownership, honoraria, research funding, expert testimony, and membership on company advisory committees. The task force made decisions on a case-by-case basis as to whether an individual's role...
should be limited as a result of a conflict.

D.S.Y. has received a research grant from Sage Products. L.A.M. has received research grants from and served as a consultant to 3M, Angiotech, and Cadence and is a consultant to Ash Access Technology. D.J.A. has received a research grant from Pfizer and has served on advisory councils for Schering-Plough and Pfizer. K.M.A. is the immediate past president of the Association for Professionals in Infection Control and Epidemiology and serves on its board of directors. H.B.'s participation does not represent official endorsement of the compendium by the National Quality Forum. D.P.C. is a member of the speakers' bureau for Enturia. S.E.C. has received a research grant from Merck. E.R.D. is a member of the speakers' bureaus for Elan, Enzon, Schering-Plough, Viropharma, Pfizer, and Astellas and serves on the advisory boards of Schering-Plough, Genzyme, and Salix. V.F. is the past president of the Society for Healthcare Epidemiology of America, has been a consultant to Steris, Verimetrix, and Merck, and is a member of the speakers' bureaus for Cubist and Merck. P.G. has received a research grant from Becton, Dickinson and Company (BD); has been on the speakers' bureau for Ortho-McNeil; and served on the Zostavax advisory board of Merck. K.S.K has received research grants from Pfizer, Merck, and Cubist; is a member of the speakers' bureaus for Pfizer, Merck, Cubist, Schering-Plough, and Wyeth; and serves on the advisory board for Schering-Plough. J.M. has received a research grant from the Swiss National Science Foundation. T.M.P. is a past president of the Society for Healthcare Epidemiology of America; is on the advisory board or the speakers' bureau for Theradoc, 3M, Replydine, and Ortho-McNeil; and has received honoraria from VHA and the Institute for Healthcare Improvement. S.S. has received an honorarium from VHA. C.D.S. is a member of the speakers' bureau for Pfizer. R.A.W. has received research grants from Sage Products and the Centers for Disease Control and Prevention and has been a consultant on Tolevamer for Genzyme and a consultant to the Centers for Disease Control and Prevention. D.C. is co-chair of the National Quality Forum Patient Safety Taxonomy Committee and an employee of CSC, a healthcare technology consulting company, and has ownership in Theradoc, a medical software company. All other authors report no relevant conflicts of interest.

**Guideline Endorser(s)**

American Organization of Nurse Executives - Professional Association
Association for Respiratory Care - Professional Association
Infusion Nurses Society - Professional Association
Pediatric Infectious Diseases Society - Professional Association
Society for Hospital Medicine - Professional Association
Society of Critical Care Medicine - Professional Association
Surgical Infection Society - Professional Association

**Guideline Status**

This is the current release of the guideline.

**Guideline Availability**


Print copies: Available from the Reprints Coordinator, University of Chicago Press, 1427 E. 60th St., Chicago, IL 60637 ([reprints@press.uchicago.edu](mailto:reprints@press.uchicago.edu)) or contact the journal office ([iche@press.uchicago.edu](mailto:iche@press.uchicago.edu)).
Availability of Companion Documents

The following are available:


Print copies: Available from the Reprints Coordinator, University of Chicago Press, 1427 E. 60th St., Chicago, IL 60637 (reprints@press.uchicago.edu) or contact the journal office (iche@press.uchicago.edu).

Performance measures and a urinary catheter reminder form (in appendix) are available in the original guideline document.

Patient Resources

The following is available:


Please note: This patient information is intended to provide health professionals with information to share with their patients to help them better understand their health and their diagnosed disorders. By providing access to this patient information, it is not the intention of NGC to provide specific medical advice for particular patients. Rather we urge patients and their representatives to review this material and then to consult with a licensed health professional for evaluation of treatment options suitable for them as well as for diagnosis and answers to their personal medical questions. This patient information has been derived and prepared from a guideline for health care professionals included on NGC by the authors or publishers of that original guideline. The patient information is not reviewed by NGC to establish whether or not it accurately reflects the original guideline's content.

NGC Status

This NGC summary was completed by ECRI Institute on January 21, 2009. The information was verified by the guideline developer on March 30, 2009.

Copyright Statement

This NGC summary is based on the original guideline, which is subject to the guideline developer's copyright restrictions.

Disclaimer

NGC Disclaimer

The National Guideline Clearinghouse™ (NGC) does not develop, produce, approve, or endorse the guidelines represented on this site.

All guidelines summarized by NGC and hosted on our site are produced under the auspices of medical specialty societies, relevant professional associations, public or private organizations, other government agencies, health care organizations or plans, and similar entities.
Guidelines represented on the NGC Web site are submitted by guideline developers, and are screened solely to determine that they meet the NGC Inclusion Criteria which may be found at http://www.guideline.gov/about/inclusion-criteria.aspx.

NGC, AHRQ, and its contractor ECRI Institute make no warranties concerning the content or clinical efficacy or effectiveness of the clinical practice guidelines and related materials represented on this site. Moreover, the views and opinions of developers or authors of guidelines represented on this site do not necessarily state or reflect those of NGC, AHRQ, or its contractor ECRI Institute, and inclusion or hosting of guidelines in NGC may not be used for advertising or commercial endorsement purposes.

Readers with questions regarding guideline content are directed to contact the guideline developer.
Appendix D: Sample MeSH Ontology

Disease/Condition

Diseases (2336)

Animal Diseases (7)
Bacterial Infections and Mycoses (348)
Cardiovascular Diseases (463)
Congenital, Hereditary, and Neonatal Diseases and Abnormalities (187)
Digestive System Diseases (320)
Disorders of Environmental Origin (345)
Endocrine System Diseases (218)
Eye Diseases (88)
Female Urogenital Diseases and Pregnancy Complications (583)
Hemic and Lymphatic Diseases (140)
Immune System Diseases (328)
Male Urogenital Diseases (411)
Musculoskeletal Diseases (203)
Neoplasms (489)
Nervous System Diseases (567)
Nutritional and Metabolic Diseases (296)
Otorhinolaryngologic Diseases (67)
Parasitic Diseases (38)
Pathological Conditions, Signs and Symptoms (1068)

Respiratory Tract Diseases (340)

Bronchial Diseases (53)
Laryngeal Diseases (5)
**Lung Diseases (220)**

Acute Chest Syndrome (1)  
Cystic Fibrosis (12)  
Hemoptysis (3)  
Hepatopulmonary Syndrome (1)  
Hypertension, Pulmonary (16)  
Lung Diseases, Fungal (12)  
Lung Diseases, Interstitial (5)  
Lung Diseases, Obstructive (66)  
Lung Injury (8)  
Lung Neoplasms (55)  

**Pneumonia (33)**

Pneumonia, Aspiration (2)  
Pneumonia, Bacterial (5)  
  Pneumonia, Pneumococcal (1)  
Pneumonia, Pneumocystis (7)  

**Pneumonia, Ventilator-Associated (5)**

Pneumonia, Viral (2)  
Pulmonary Atelectasis (1)  
Pulmonary Edema (3)  
Pulmonary Embolism (25)  
Pulmonary Emphysema (2)  
Pulmonary Fibrosis (2)  
Respiratory Distress Syndrome, Adult (3)  
Respiratory Distress Syndrome, Newborn (1)  
Solitary Pulmonary Nodule (1)  
Tuberculosis, Pulmonary (4)  

Nose Diseases (19)
Pleural Diseases (11)

Respiration Disorders (97)

Respiratory Hypersensitivity (45)

Respiratory Tract Fistula (1)

Respiratory Tract Infections (111)

Respiratory Tract Neoplasms (60)

Thoracic Diseases (3)

Tracheal Diseases (1)

Skin and Connective Tissue Diseases (206)

Stomatognathic Diseases (76)

Virus Diseases (291)
Appendix E: Complete Screenshots of Website

Home Web Page

GUIDELINE SEARCH

WELCOME TO GUIDELINE SEARCH.
This research project will run from August 1 through December 31. The research involves investigating the relationship between the structure of an ontology (its depth, breadth, and fanout) and the difference in performance for an online search for information based on a medical scenario.

DIRECTIONS
This process consists of 5 steps:
1. Completing a short Intro Survey.
2. Completing the first search task.
3. Completing the first post-task survey.
4. Completing the second search task.
5. Completing the second post-task survey.

The search tasks require you to:
1. Read a short medical scenario and determine its topic.
2. Search the provided hierarchy for terms related to the topic, add them to a search box, and then submit the search.
3. Examine the returned guidelines and select the best guidelines related to the topic of the scenario.

The entire process should take approximately 20 minutes.

START THE STUDY

START THE STUDY or click Intro Survey above to Log In and start.

Thank you for participating in this research.
You can go to the about page to get more information, or email the research team if you have any questions, concerns, or want additional information.
Intro Survey Web Page

![Guideline Search](image)

**Intro Survey**

Please complete all fields in the survey below.

- **Participant user**

  - Your program *(e.g. BSN ACE, MSN)*
    - OR resident medical specialty *(e.g. Emergency Medicine, Internal Medicine)*
  - If applicable, your track *(e.g. Adult Acute Care, Pediatric Primary Care)*
    - OR medical sub-specialty *(e.g. Hematology, Nephrology, Rheumatology)*

- **Your year**
  - [ ] 1st
  - [ ] 2nd
  - [ ] 3rd
  - [ ] 4th

- **Submit**
Task 1 Web Page

MeSH ontology hierarchy expanded to demonstrate navigation.

Hello user1, please follow STEPS 1 through 3 on the page to complete Task 1.

**STEP 1:**
1. Read the scenario.

**SCENARIO**
A type of hospital-acquired pneumonia (HAP) can occur in ICU patients who have pulmonary disease or are immuno suppressed and who are receiving mechanical ventilation through an endotracheal or tracheostomy tube for at least 48 hours. The pneumonia primarily occurs because the tube allows free passage of bacteria into the lower segments of the lung. Your patient shows the following signs: alternating fever and low body temperature, pendulant abdomen, and hypoxia. What intervention is most likely to decrease the incidence of this patient developing HAP?

**STEP 2:**
1. Search the hierarchy below for the keyword(s) to use to find guidelines that answer the question in the scenario. Try to get to the most specific keyword(s).
2. Click on the keyword(s) to add them to the Search Keywords textbox.
3. Click Search to retrieve the guidelines. The guidelines will appear in the Search Results section below.

**HIERARCHY**

*Note:* Use + to expand and - to collapse the hierarchy.
- Diseases
  - Bacterial Infections and Mycoses
  - Cardiovascular Diseases
    - Cardiovascular Abnormalities
    - Cardiovascular Infections
    - Heart Diseases
      - Pregnancy Complications, Cardiovascular
  - Vascular Diseases
  - Digestive System Diseases
  - Endocrine System Diseases
  - Hematologic and Lymphatic Diseases
  - Neoplasms
  - Nervous System Diseases
  - Respiratory Tract Diseases
  - Virus Diseases

**STEP 3:**
1. Scroll through the guidelines in Search Results and click on the guideline(s) that may answer the question in the scenario.
2. Examine the Recommendations section of the guideline by clicking on the Recommendations link on the Jump To tab of the guideline.
3. Use the browser back button to return to the results list.
4. Copy the NGC code number (e.g., <i>001234</i>) from NGC into the Guidelines Selected textbox. Separate multiple code numbers with commas.
5. Click Save to submit the guidelines selected.

*Note:*
Guidelines must have a NGC code after the guideline name e.g. Healthy lifestyles, 2008 Mar(revised 2011 May), NGC:008572. Ignore other results.

**Search Results**
Task 1 Web Page

Showing search results with ventilator-associated pneumonia results from the National Guideline Clearinghouse in the SEARCH RESULTS window.

**SEARCH RESULTS**
**Post Task 1 Survey Web Page**

Hello user1, please complete all fields in the survey below.

Have you ever used the Agency for Healthcare Research and Quality (AHRQ) to find guidelines?
- Yes
- No

If Yes, did you use Search or the Hierarchy to find guidelines?
- Search
- Hierarchy
- Unfamiliar

How familiar are you with the topic described in the scenario?
- Very familiar
- Somewhat familiar
- Unfamiliar

Were you able to find the keywords you wanted in the hierarchy?
- Yes
- No

If you answered No, which keywords were you looking for?

Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.

Submit
Task 2 Web Page

GUIDELINE SEARCH

Welcome user1! [Log Out]

Home   Intro Survey   Task 1   Post Task 1 Survey   Task 2   Post Task 2 Survey   About

STEP 1:
1. Read the scenario.

SCENARIO
A 38-year-old man has sustained a severe, closed, traumatic brain injury. On admission to the ED, the Glasgow coma scale score is 8 and his blood pressure is 85/50 mm Hg. The patient is intubated and receives 2 L of 0.9% saline, which improves his blood pressure. A CT scan of the head shows bilateral frontal hematomas. What should be monitored as part of this patient’s appropriate management?

STEP 2:
1. Search the hierarchy below for the keyword(s) to use to find guidelines that answer the question in the scenario. Try to get to the most specific keywords.
2. Click on the keyword(s) to add them to the Search Keywords textbox.
3. Click Search to retrieve the guidelines. The guidelines will appear in the Search Results section below.

Search Keywords

Search
Clear Search Keywords

HIERARCHY
Note: Use + to expand and - to collapse the hierarchy.

% Diseases
% Infectious Diseases
% Bacterial Infections and Mycoses
% Cardiovascular Diseases
% Cardiovascular Abnormalities
% Cardiovascular Infections
% Heart Diseases
- Pregnancy Complications, Cardiovascular
% Vascular Diseases
% Digestive System Diseases
% Endocrine System Diseases
% Hemic and Lymphatic Diseases
% Neoplasms
% Nervous System Diseases
% Respiratory Tract Diseases
% Virus Diseases

STEP 3:
1. Scroll through the guidelines in Search Results and click on the guideline(s) that may answer the question in the scenario.
2. Examine the Recommendations section of the guideline by clicking on the Recommendations link on the Jump To tab of the guideline.
3. Use the browser back button to return to the results list.
4. Copy the NGC code number (e.g., 001234 from NGC:001234) into the Guidelines Selected textbox. Separate multiple code numbers with commas.
5. Click Save to submit the guidelines selected.

Note:
Guidelines must have a NGC code after the guideline name: e.g., Healthy lifestyles, 2008 Mar (revised 2011 May), NGC:008572. Ignore other results.

Guidelines Selected

Search

Search Results
Post Task 2 Survey Web Page

Hello user1, please complete all fields in the survey below.

How familiar are you with the topic described in the scenario?

- [ ] Very unfamiliar

Were you able to find the keywords you wanted in the hierarchy?

- [ ] Yes

If you answered No, which keywords were you looking for?

Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.

Submit

About Web Page

Not displayed. Participants do not access the Web page as part of the study procedure.
## Appendix F: Complete Text for All Scenarios

<table>
<thead>
<tr>
<th>No.</th>
<th>Scenario Text</th>
<th>Disease and Treatment / Management Options</th>
</tr>
</thead>
</table>
| 1   | A type of hospital-acquired pneumonia (HAP) can occur in people who are on a breathing machine through an endotracheal or tracheostomy tube for at least 48 hours. The pneumonia primarily occurs because the tube allows free passage of bacteria into the lower segments of the lung in a person who often has underlying lung or immune problems. Your patient shows the following signs: alternating fever and low body temperature, purulent sputum, and hypoxia. What intervention is most likely to decrease the incidence of this type of pneumonia in the intensive care unit? | Ventilator-associated Pneumonia  
Requires subglottic secretion drainage                                                                               |
| 2   | A 38-year-old man has sustained a severe, closed, traumatic brain injury. On admission to the ED, the Glasgow coma scale score is 8 and his blood pressure is 85/50 mm Hg. The patient is intubated and receives 2 L of 0.9% saline, which improves his blood pressure. A CT scan of the head shows bilateral frontal hematomas. What should be monitored as part of appropriate management? | Coma  
Requires intracranial pressure monitoring                                                                                  |
| 3   | A 45-year-old white woman comes to the emergency room because of fever, chills, night sweats, and a dry cough for three weeks. The patient reports no other medical problems and does not recall any prior skin testing. She has been living in a homeless shelter for several months. The patient is thin and in moderate respiratory distress. Temperature is 39.5 C (102.6 F). Pulse rate is 110 per minute, and rhythm is regular. Respiration are 22 per minute. Blood pressure is 100/70 mm Hg. The oral mucous membranes appear somewhat dry. Chest examination reveals fine crackles in the right upper lobe. A grade 2/6 systolic ejection murmur is heard best at the left sternal border. Hemoglobin is 9.8 g/dL, and serum electrolyte levels are normal. Chest radiograph reveals a patchy density in the right upper lobe. What are the next most appropriate steps for treatment? | Pulmonary Tuberculosis  
Can require admittance to a negative-pressure isolation room; obtaining induced sputum samples for staining, culture, and drug susceptibility testing. |
Appendix G: Post Task Survey Answers

Novice Resident Physicians

Post Task 1 Survey

<table>
<thead>
<tr>
<th>Physician ID</th>
<th>Questions</th>
<th>If No, which keywords were you looking for?</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>physician 1</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>physician 2</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td>in this case, i was looking for practice guidelines, monitoring, traumatic brain injury</td>
</tr>
<tr>
<td>physician 3</td>
<td>Familiar</td>
<td>Yes</td>
<td>Prevention of ventilator associated pneumonia</td>
</tr>
<tr>
<td>physician 4</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>Did it answer question</td>
</tr>
<tr>
<td>physician 5</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>If the guidelines had evidence of what to monitor for bilateral hematoma</td>
</tr>
<tr>
<td>physician 6</td>
<td>Familiar</td>
<td>Yes</td>
<td>I was looking for general recommendations evidence based. The title would tell me that.</td>
</tr>
<tr>
<td>physician 7</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>Evidence based recommendations</td>
</tr>
<tr>
<td>physician 8</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>physician 9</td>
<td>Familiar</td>
<td>Yes</td>
<td>Just made sure whether the recommended guidelines are addressing the issue of concern, in this case prevention of pneumonia in intubated patients.</td>
</tr>
<tr>
<td>physician 10</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>physician 11</td>
<td>Unfamiliar</td>
<td>No</td>
<td>Looking for prevention and vap</td>
</tr>
<tr>
<td>physician 12</td>
<td>Familiar</td>
<td>Yes</td>
<td>ventilator acquired pneumonia; pneumonia prevention</td>
</tr>
<tr>
<td>physician 13</td>
<td>Familiar</td>
<td>Yes</td>
<td>general monitoring parameters after acute traumatic head injury, particularly with hematoma (unclear from prompt what type of hematoma specifically)</td>
</tr>
</tbody>
</table>
## Post Task 2 Survey

<table>
<thead>
<tr>
<th>Physician ID</th>
<th>Questions</th>
<th>Were you able to find the keywords you wanted in the hierarchy?</th>
<th>If No, which keywords were you looking for?</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>physician 1</td>
<td>Very familiar</td>
<td>Yes</td>
<td></td>
<td>I was looking for a clinical explanation of the management of traumatic closed head injuries, namely, searching for literature to support ICP monitoring, as I assumed that was the correct answer.</td>
</tr>
<tr>
<td>physician 2</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td>treatment of infective endocarditis</td>
</tr>
<tr>
<td>physician 3</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>paitent monitoring recommendations in traumatic brain injury</td>
</tr>
<tr>
<td>physician 4</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td>how to treat pneumonia</td>
</tr>
<tr>
<td>physician 5</td>
<td>Very familiar</td>
<td>Yes</td>
<td></td>
<td>I was looking for CAP guidelines in adults.</td>
</tr>
<tr>
<td>physician 6</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td>I was looking for a guideline that mention severe head trauma management</td>
</tr>
<tr>
<td>physician 7</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>Evidence based treatment</td>
</tr>
<tr>
<td>physician 8</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 9</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>Basically, I was searching for the guidelines required to monitor the patient with acute head injury. If any hematoma is causing raised intracranial tension, it is an absolute neurosurgical emergency.</td>
</tr>
<tr>
<td>physician 10</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 11</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td>imaging studies and monitoring parameters</td>
</tr>
<tr>
<td>physician 12</td>
<td>Neutral</td>
<td>No</td>
<td>pneumonia in homeless patient</td>
<td>homeless; pneumonia; management; possibility of tuberculosis</td>
</tr>
<tr>
<td>physician 13</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Expert Resident Physicians

### Post Task 1 Survey

<table>
<thead>
<tr>
<th>Physician ID</th>
<th>Questions</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>physician 1</td>
<td>Familiar</td>
<td>Prevention of VAP</td>
</tr>
<tr>
<td>physician 2</td>
<td>Very familiar</td>
<td>Searching for literature supporting tactics to prevent/decrease occurrence of Ventilator Associated Pneumonia in Acute Care Patients.</td>
</tr>
<tr>
<td>physician 3</td>
<td>Neutral</td>
<td>infections in the lungs</td>
</tr>
<tr>
<td>physician 4</td>
<td>Familiar</td>
<td>Bacterial respiratory infections as well as lowered immune response and foreign, indwelling devices.</td>
</tr>
<tr>
<td>physician 5</td>
<td>Familiar</td>
<td>I was looking for guidelines regarding ventilator-associated pneumonia and prevention of hospital or ventilator associated pneumonia</td>
</tr>
<tr>
<td>physician 6</td>
<td>Familiar</td>
<td>Pertinent details regarding the disease in question, namely &quot;VAP&quot;, &quot;prevention&quot;, and &quot;intervention.&quot;</td>
</tr>
<tr>
<td>physician 7</td>
<td>Very familiar</td>
<td>I was looking for articles on prevention on infections in the healthcare setting</td>
</tr>
<tr>
<td>physician 8</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>physician 9</td>
<td>Familiar</td>
<td></td>
</tr>
<tr>
<td>physician 10</td>
<td>Very familiar</td>
<td>Title of Guideline</td>
</tr>
<tr>
<td>physician 11</td>
<td>Familiar</td>
<td>I am familiar with the topic of ventilator associated pneumonia. I am familiar with basic strategies for prevention was looking for these strategies in the guidelines.</td>
</tr>
<tr>
<td>physician 12</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>Physician ID</td>
<td>Questions</td>
<td>How familiar are you with the topic described in the scenario?</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>physician 13</td>
<td>I was looking for mention of <strong>hematoma under the neurological trauma</strong> section. I was a bit confused whether the hematoma was epidural or subdural, as this was not listed in the stem.</td>
<td>Neutral</td>
</tr>
<tr>
<td>physician 14</td>
<td>I was looking for an overview of the disease process, diagnostic tools and management criteria.</td>
<td>Familiar</td>
</tr>
<tr>
<td>physician 15</td>
<td>key words that reflect a summary of the scenario and your treatment plan</td>
<td>Familiar</td>
</tr>
<tr>
<td>physician 16</td>
<td>Hematoma, hypotension, cerebrovascular injury</td>
<td>Unfamiliar</td>
</tr>
</tbody>
</table>

**Post Task 2 Survey**

<table>
<thead>
<tr>
<th>Physician ID</th>
<th>Questions</th>
<th>How familiar are you with the topic described in the scenario?</th>
<th>Were you able to find the keywords you wanted in the hierarchy?</th>
<th>If No, which keywords were you looking for?</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>physician 1</td>
<td>Initial eval and management of possible TB</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 2</td>
<td>Looking for treatment and <strong>management of active tuberculosis</strong>, involving antibiotic choice and containment of infection.</td>
<td>Very familiar</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 3</td>
<td><strong>tuberculosis management</strong></td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 4</td>
<td>The scenario is suspicious for <strong>MTB</strong>, however other GNR/anaerobes as well as endocarditis would also be on the list.</td>
<td>Familiar</td>
<td>No</td>
<td>Tuberculosis is, hematologic associated night sweats</td>
<td></td>
</tr>
<tr>
<td>physician 5</td>
<td>Guidelines for <strong>intracranial pressure monitoring</strong> in closed trauma to brain</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physician ID</td>
<td>Questions</td>
<td>Were you able to find the keywords you wanted in the hierarchy?</td>
<td>If No, which keywords were you looking for?</td>
<td>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>physician 6</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>Keywords &quot;TB&quot; and &quot;diagnosis&quot;. Also relevant were recommendations regarding various tests available and what is the most current information regarding efficacy vs cost/benefit.</td>
<td></td>
</tr>
<tr>
<td>physician 7</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td>I was looking for articles on what to monitor for after a traumatic cerebral hemorrhage.</td>
<td></td>
</tr>
<tr>
<td>physician 8</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 9</td>
<td>Neutral</td>
<td>No</td>
<td></td>
<td>I was looking in general for Community Acquired Pneumonia. The closer thing was Pneumococcal Pneumonia.</td>
<td></td>
</tr>
<tr>
<td>physician 10</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td>title of guideline and relevance of recommendations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 11</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td>I was looking for goals of management related to the clinical scenario. I wanted specific answers on what to look for while the patient is on the medical floor.</td>
<td></td>
</tr>
<tr>
<td>physician 12</td>
<td>Familiar</td>
<td>Yes</td>
<td>The stage at which the scenario is presented; i.e., TB appears likely, but the diagnosis needs to be confirmed to guide management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>physician 13</td>
<td>Very familiar</td>
<td>No</td>
<td>pneumonia</td>
<td>I was looking for guidelines for pneumonia in varying clinical settings.</td>
<td></td>
</tr>
<tr>
<td>physician 14</td>
<td>Very familiar</td>
<td>Yes</td>
<td></td>
<td>I was searching for guidelines for Pneumonia and TB. Isolation recommendations.</td>
<td></td>
</tr>
<tr>
<td>physician 15</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>again, used key words in the scenario.</td>
<td></td>
</tr>
<tr>
<td>physician 16</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>pneumonia, homeless/shelter, SIRS.</td>
<td></td>
</tr>
</tbody>
</table>
# Novice Nursing School Students

## Post Task 1 Survey

<table>
<thead>
<tr>
<th>Nursing Student ID</th>
<th>Questions</th>
<th>How familiar are you with the topic described in the scenario?</th>
<th>Were you able to find the keywords you wanted in the hierarchy?</th>
<th>If No, which keywords were you looking for?</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nursing student 1</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td></td>
<td>I was looking for key words including: <strong>ventilator/ventilation</strong>, hospital acquired, pneumonia. I then looked at the recommended interventions to see if it matched what the scenario was looking for.</td>
<td></td>
</tr>
<tr>
<td>nursing student 2</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nursing student 3</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td><strong>Ventilator-acquired pneumonia,</strong> <strong>ventilator-associated pneumonia,</strong> <strong>prevention</strong></td>
<td></td>
</tr>
<tr>
<td>nursing student 4</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
<td><strong>HAP and ventilators.</strong></td>
<td></td>
</tr>
<tr>
<td>nursing student 5</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>I was looking for keywords in the title of the article, such as &quot;prevention.&quot; This keyword would be in relation to the subject searched</td>
<td></td>
</tr>
<tr>
<td>nursing student 6</td>
<td>Neutral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nursing student 7</td>
<td>Unfamiliar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nursing student 8</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>I suspected <strong>VAP</strong> so I searched it and found results on preventing its occurrence to answer the question.</td>
<td></td>
</tr>
<tr>
<td>nursing student 9</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
<td>I was looking for prevention management in ventilator associated pneumonia because I know from that the <strong>most common cause of HAP is due ventilators.</strong> I need to search for guidelines that went into specific details of care for ventilators to prevent pneumonia.</td>
<td></td>
</tr>
<tr>
<td>nursing student 10</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
<td>look for topics that matched the statement scenario</td>
<td></td>
</tr>
</tbody>
</table>
### Post Task 2 Survey

<table>
<thead>
<tr>
<th>Nursing Student ID</th>
<th>Questions</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nursing student 1</td>
<td>Very unfamiliar</td>
<td>I decided to investigate the cough and related symptoms first to see if it would lead me to guidelines on how to proceed.</td>
</tr>
<tr>
<td>nursing student 2</td>
<td>Very unfamiliar</td>
<td></td>
</tr>
<tr>
<td>nursing student 3</td>
<td>Neutral</td>
<td>Monitor, ICP, intracranial pressure, Anything related to assessment, monitoring of monitoring of brain hemorrhage s/p traumatic injury.</td>
</tr>
<tr>
<td>nursing student 4</td>
<td>Neutral</td>
<td>Skin testing, treatment, and control of possible TB</td>
</tr>
<tr>
<td>nursing student 5</td>
<td>Familiar</td>
<td>The question was looking for ways to manage a traumatic brain injury so I wanted to make sure the title included the word &quot;management.&quot; Based on the recommendations from the article, it looks to give appropriate information for the management of a TBI.</td>
</tr>
<tr>
<td>nursing student 6</td>
<td>Neutral</td>
<td></td>
</tr>
<tr>
<td>nursing student 7</td>
<td>Familiar</td>
<td>Since it was bilateral Frontal lobe hematoma, there are high chances of optic nerve damage</td>
</tr>
<tr>
<td>nursing student 8</td>
<td>Familiar</td>
<td>I was looking for treatment/intervention guidelines for the treatment of either community-acquired pneumonia or tuberculosis (since the history mentioned skin testing).</td>
</tr>
<tr>
<td>nursing student 9</td>
<td>Familiar</td>
<td>I knew the scenario related to some type of respiratory infection. Since it mentioned dry cough, I immediately thought of the flu, CAP, or TB, especially since she was living in homeless shelter. This increases the risk of disease transmission. I searched for the diseases they could be and wanted signs and symptoms to show up. Then, if it correlated to the presenting symptoms of the patient, I also wanted specifics on what labs to collect, to confirm a diagnosis. Only then will treatment be</td>
</tr>
<tr>
<td>Nursing Student ID</td>
<td>Questions</td>
<td>Were you able to find the keywords you wanted in the hierarchy?</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>nursing student 10</td>
<td>Unfamiliar</td>
<td></td>
</tr>
</tbody>
</table>
## Expert Nursing School Students

### Post Task 1 Survey

<table>
<thead>
<tr>
<th>Nursing Student ID</th>
<th>Questions</th>
<th>If No, which keywords were you looking for?</th>
<th>Briefly explain what you were looking for to determine if the guidelines you selected were relevant to the scenario.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nursing student 1</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>Key words related to the clinical scenario, clear concise information</td>
</tr>
<tr>
<td>nursing student 2</td>
<td>Familiar</td>
<td>Yes</td>
<td>I was looking for topics that included prevention of ventilator acquired pneumonia (VAP).</td>
</tr>
<tr>
<td>nursing student 3</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>Was the guideline specifically addressing hospital acquired pneumonia specifically related to ventilators?</td>
</tr>
<tr>
<td>nursing student 4</td>
<td>Familiar</td>
<td>Yes</td>
<td>Looking for research to show how to prevent HAP infections to serve as a guide for staff. Through the use of research to implement a new procedures for the care of patients.</td>
</tr>
<tr>
<td>nursing student 5</td>
<td>Neutral</td>
<td>Yes</td>
<td>If it included what needed to be monitored on this patient</td>
</tr>
<tr>
<td>nursing student 6</td>
<td>Familiar</td>
<td>Yes</td>
<td>I was only able to find ventilator associated pneumonia</td>
</tr>
<tr>
<td>nursing student 7</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td>Looking for information on blunt trauma and hematoma. What to look for to make sure this patient is stable.</td>
</tr>
<tr>
<td>nursing student 8</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>Looking for the keyword &quot;hematoma&quot; and monitoring in the recommendation sections of the provided guidelines.</td>
</tr>
<tr>
<td>nursing student 9</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td>pneumonias and ventilated patients</td>
</tr>
<tr>
<td>nursing student 10</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td>Prevention of hospital acquire pneumonia and infections.</td>
</tr>
<tr>
<td>nursing student 11</td>
<td>Familiar</td>
<td>Yes</td>
<td>Ventilator dependent, pneumonia prevention, immunocompromised</td>
</tr>
<tr>
<td>nursing student 12</td>
<td>Familiar</td>
<td>Yes</td>
<td>VAP Protocol</td>
</tr>
<tr>
<td>nursing student 13</td>
<td>Familiar</td>
<td>Yes</td>
<td>interventions regarding hospital-acquired ventilator pneumonia</td>
</tr>
<tr>
<td>nursing student 14</td>
<td>Familiar</td>
<td>Yes</td>
<td>Guidelines to prevent the occurrence of VAP in the inpatient setting</td>
</tr>
<tr>
<td>nursing student 15</td>
<td>Familiar</td>
<td>Yes</td>
<td>If the guidelines listed specific recommendations on how to prevent VAP, which is what the questions asked</td>
</tr>
<tr>
<td>Nursing Student ID</td>
<td>Questions</td>
<td>How familiar are you with the topic described in the scenario?</td>
<td>Were you able to find the keywords you wanted in the hierarchy?</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>nursing student 16</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 17</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 1</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 2</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 3</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 4</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 5</td>
<td>Neutral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nursing student 6</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 7</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Nursing Student ID</td>
<td>Questions</td>
<td>How familiar are you with the topic described in the scenario?</td>
<td>Were you able to find the keywords you wanted in the hierarchy?</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>nursing student 8</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 9</td>
<td>Very unfamiliar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 10</td>
<td>Neutral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nursing student 11</td>
<td>Neutral</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 12</td>
<td>Neutral</td>
<td>No</td>
<td>neurological assessment - baseline indicators (not mra, ct)</td>
</tr>
<tr>
<td>nursing student 13</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 14</td>
<td>Familiar</td>
<td>No</td>
<td>Treatment and diagnosis of respiratory infection</td>
</tr>
<tr>
<td>nursing student 15</td>
<td>Unfamiliar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 16</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>nursing student 17</td>
<td>Familiar</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
Vita
Ilya Michael Waldstein

Education

2013 Drexel University
Ph.D. Information Science
Thesis: The Relationship Between User Expertise and Structural Ontology Characteristics
Advisor: Dr. Rosina O. Weber, Drexel University

2003 Drexel University
M.S. Information Systems

1994 Rutgers University
B.A. Economics

Publications

Refereed Conference Papers


Posters
