

**Exploring the Semantic Meaning of Constructs that Lead  
to Human Decisions**

by

**Chih How Bong**

M.S., University of Colorado, 2010

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Computer Science

2011

This thesis entitled:  
Exploring the Semantic Meaning of Constructs that Lead to Human Decisions  
written by Chih How Bong  
has been approved for the Department of Computer Science

---

James Martin

---

Kai Larsen

Date \_\_\_\_\_

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Bong, Chih How (Ph.D., Computer Science)

Exploring the Semantic Meaning of Constructs that Lead to Human Decisions

Thesis directed by James Martin

This study examines automated approaches to discovering behavioral knowledge that are encoded as constructs in social and behavioral science disciplines. To date, constructs relationships are ordinarily revealed through laborious psychometric methods, but this study has shown that it is possible to extract these relationships through automated computational approaches. By building on text similarity measures from prior literature, we are able to predict construct relationships through construct name, definition and items. The predicted relationships were woven into an interlock system to demonstrate construct interplays, even though they have not been studied. The construct interlock could be seen as a theory map to understand human decision-making. Two use cases were presented to demonstrate the efficacy of the proposed measures: measuring the root constructs in UTAUT and visualizing network of construct perceived usefulness. The encouraging results showed that the proposed measures could dramatically expedite theory development, at the same time also expedite progression of human science.

## Acknowledgements

First of all, I would like to honor God the mighty Creator for giving me the opportunity, patience, energy and wisdom to complete the study. Without You, nothing in this world is possible.

I would like to sincerely thank my beloved wife, Ding Kiew (Aivee) for her sacrifices, understanding and patience in these not-too-great 5 years. I believe you deserve a better life than this! Thank you for supporting and believing in me.

Also to my adorable daughter, Faith, for throwing tantrums at me when you were tired, for kidding and cheering me up when I was down. Your existence really lights up my life.

I would also like to thank Kai Larsen for all the creativities, inspirations, encouragements, humors, ideas and every fruitful discussions we had. You are also the one who makes this thesis possible and accomplished. You have been so patient with my buggy works, your are flexible whenever you were needed and you kindly support when I was in difficulties.

I would like to thank James Martin for your time and support in making this study possible. I am truly appreciate your willingness to take me as your advisee, knowing that there are many outstanding students under your supervision.

Next, I also like to thank Tamara Sumner for your support and understanding when I have to switch to this field of study. I really sorry to see our previous work-together project to be called off.

Thanks to all my committee members, especially Clayton Lewis and Martha Palmer for all your valuable time and constructive feedback.

I would like to show my appreciation to my friends Ovo, Keith Maul, Jing Jing Li, Jeffrey

Sweeney, and Lee Becker who willing to read and comment the draft of my proposal or thesis even though you guys are constrained by time.

Not to forget my family and colleague in Malaysia, especially Narayanan Kulathuramaiyer, for all kind of support.

Lastly, I would like to thank the sponsors of my study, UNIMAS, NSF, CCTSI, SSHRC, making my study here in CU Boulder, USA a dream comes true.

## Contents

### Chapter

<b>1</b>	Introduction	1
1.1	Overview . . . . .	1
1.2	Motivation . . . . .	1
1.3	Problems . . . . .	2
1.3.1	Construct proliferation . . . . .	2
1.3.2	Linguistic ambiguity . . . . .	2
1.3.3	Disconnected constructs . . . . .	3
1.4	Research Questions . . . . .	3
1.5	Research Goal and Objectives . . . . .	4
1.6	Research Scope . . . . .	4
1.7	Chapter Overview . . . . .	4
<b>2</b>	Constructs	6
2.1	Overview . . . . .	6
2.2	What are constructs . . . . .	6
2.3	Why constructs are important . . . . .	7
2.4	How constructs are created . . . . .	8
2.4.1	Create a clear definition . . . . .	8
2.4.2	Specify a construct's dimension and their relationships . . . . .	9

2.4.3	Explore how constructs can be applied to other context . . . . .	9
2.4.4	Expand the conceptualization of constructs . . . . .	9
2.5	Psychometrics . . . . .	10
2.5.1	Validity and Reliability in Construct . . . . .	10
2.6	Problems . . . . .	11
2.6.1	Construct Proliferation . . . . .	12
2.6.2	Linguistic Ambiguity . . . . .	12
2.6.3	Construct Disconnect . . . . .	13
2.7	Evaluation of Constructs by Experts . . . . .	14
2.8	Nomological Network . . . . .	14
2.9	From Nomological Network to ConstructNet . . . . .	15
2.10	Construct Collection . . . . .	16
2.11	Conclusion . . . . .	20
<b>3</b>	<b>Background</b>	<b>22</b>
3.1	Overview . . . . .	22
3.2	Human Behavior Project (HBP) . . . . .	22
3.3	Human Decision Making through HBP . . . . .	22
3.4	Theory . . . . .	23
3.5	Solutions . . . . .	25
3.5.1	Search Engines . . . . .	25
3.5.2	Reference Resources . . . . .	26
3.5.3	Citation Analysis . . . . .	27
3.5.4	Meta Analysis . . . . .	28
3.6	Conclusion . . . . .	29
<b>4</b>	<b>Computational Approaches to Detect Similar Constructs</b>	<b>31</b>
4.1	Introduction . . . . .	31

4.2	Text Similarity Measurement . . . . .	31
4.2.1	Document Vector and word-occurrence methods . . . . .	32
4.2.2	Corpus-based methods . . . . .	33
4.2.3	Hybrid Method . . . . .	34
4.2.4	Descriptive method . . . . .	35
4.3	Review of Selected Sentence Similarity Approaches . . . . .	35
4.3.1	Latent Semantic Indexing (LSA) . . . . .	36
4.3.2	Li et al. Similarity Measures . . . . .	38
4.3.3	Mihalcea et al. Similarity Measure . . . . .	43
4.3.4	Measuring Sentence Similarity . . . . .	44
4.4	Limitations of the Reviewed Sentence Similarity Measures . . . . .	44
4.4.1	LSA Limitations . . . . .	45
4.4.2	Limitations in Li et al. and Mihalcea et al. Similarity Measures . . . . .	45
4.5	Proposed Similarity Measures for Construct Relationships . . . . .	46
4.5.1	Construct Similarity Measure derived from Li et al. (ConSimLi) . . . . .	46
4.5.2	Construct Similarity Measure derived from Mihalcea et al. (ConSimMi) . . . . .	46
4.6	Predicting Construct Similarity with the Proposed Approaches . . . . .	47
4.7	Evaluation . . . . .	48
4.8	Conclusion . . . . .	48
<b>5</b>	<b>Building a Gold Standard</b>	<b>50</b>
5.1	Overview . . . . .	50
5.2	Objective . . . . .	50
5.3	Related Works . . . . .	51
5.4	Our Challenges . . . . .	51
5.5	Our Approach . . . . .	51
	Step 1: Defining Construct Relationships . . . . .	52



Step 2: Annotator Training . . . . .	52
Step 3: Colocating Identical Constructs . . . . .	53
Step 4: Rough Categorization . . . . .	54
Step 5: Refined Categorization . . . . .	54
Step 6: Reconciling Changes . . . . .	57
Step 7: Evaluating Categories . . . . .	58
5.6 Gold Standard Validity . . . . .	60
5.7 Conclusion . . . . .	61
<b>6 Evaluation</b>	<b>62</b>
6.1 Overview . . . . .	62
6.2 LSA Dimensionality . . . . .	62
6.3 Comparison of Construct Similarity Measures . . . . .	65
6.4 Similarity Scores Distribution . . . . .	70
6.4.1 Cumulative Frequency Distributions of Correspondents Constructs . . . . .	71
6.4.2 Frequency Distributions of New Correspondents Constructs . . . . .	73
6.5 Combined Measurement . . . . .	74
6.6 Conclusion . . . . .	80
<b>7 Use Case 1: Predicting Construct Similarity with a Unified Theory</b>	<b>81</b>
7.1 Overview . . . . .	81
7.2 Background . . . . .	81
7.2.1 Unified Theory of Acceptance and Use of Technology (UTAUT) . . . . .	81
7.2.2 Conceptualization . . . . .	83
7.3 Task Description . . . . .	84
7.4 Results and Discussion . . . . .	85
7.5 Conclusion . . . . .	90

<b>8</b>	<b>Use Case 2: ConstructNet</b>	<b>92</b>
8.1	Overview . . . . .	92
8.2	Task Description . . . . .	92
8.3	Results and Discussion . . . . .	93
8.3.1	Perceived Ease of Use and Perceived Usefulness . . . . .	95
8.3.2	Perceived Usefulness and Cognitive Absorption . . . . .	98
8.4	Conclusion . . . . .	102
<b>9</b>	<b>Conclusion, Contributions and Future Works</b>	<b>104</b>
9.1	Overview . . . . .	104
9.2	Conclusion . . . . .	104
9.3	Contributions . . . . .	106
9.4	Future Work . . . . .	107
9.4.1	Improving the Proposed Similarity Measures . . . . .	107
9.4.2	Attesting our Proposed Measures on a Much Large Scale Unified Model . . .	107
9.4.3	Qualitative Analysis of Construct Relationships in the ConstructNet . . . . .	107
	<b>Bibliography</b>	<b>109</b>
	<b>Appendix</b>	
<b>A</b>	<b>Training Document</b>	<b>115</b>
A.1	Criteria for Classifying Constructs . . . . .	115
<b>B</b>	<b>Construct Categories in the Gold Standard</b>	<b>118</b>
<b>C</b>	<b>UTAUT Constructs</b>	<b>120</b>

## Tables

### Table

2.1	The table shows that the construct, Perceived Usefulness, as reported in Davis[17] has three textual properties: name, definition and items. . . . .	7
4.1	Deriving semantic vector for $S_1$ . The similarity scores were computed using Equation 4.4 . . . . .	39
4.2	Deriving a semantic vector for $S_2$ . The similarity scores were computed using Equation 4.4 . . . . .	40
4.3	Pilot study: How Li et al. and Mihalcea et al. similarity measures operating on Wordnet are not able to capture domain-specific words. . . . .	49
5.1	The degree of agreement between the experts and the gold standard. . . . .	61
6.1	Comparison of different dimensionality model in retrieving relevant measurement items. . . . .	64
6.2	Area Under Curve for 12 models. . . . .	67
6.3	The number of total matches and the accuracy in the nine models. . . . .	77
7.1	Two most similar items and their score between Perceived Behavioral Control(PBC) and Perceived Ease of Use (PEU). . . . .	86
7.2	Two most similar items and their score between Perceived Behavioral Control(PBC) and Ease of Use (EU). . . . .	87

7.3	Two most similar items and their score between Perceived Behavioral Control(PBC) and Compatibility (CMPT). . . . .	87
7.4	Two most similar items and their score between Perceived Behavioral Control(PBC) and Facilitating Control (FC). . . . .	87
8.1	Perceived Usefulness constructs from different articles. . . . .	96
8.2	Similarity scores on two set of items. . . . .	97
8.3	Perceived Usefulness properties. . . . .	100
C.1	Performance Expectancy: Root Constructs, Definitions, and Items . . . . .	121
C.2	Effort Expectancy: Root Constructs, Definitions, and Items . . . . .	122
C.3	Social Influence: Root Constructs, Definitions, and Items . . . . .	122
C.4	Facilitating Conditions: Root Constructs, Definitions, and Items . . . . .	123

## Figures

### Figure

- |     |   |    |
|-----|---|----|
| 2.1 | The Nomological Network developed by Cronbach and Meehl [14]. The figure shows the relationships of theoretical constructs to observables, observable properties or quantities to each other, and different theoretical constructs to each other. . . . .                                       | 15 |
| 2.2 | Excerpt of article content published in MISQ. Black boxes represent boundary of paragraphs to be extracted. . . . .   | 18 |
| 2.3 | In some cases, the constructs and their corresponding properties are contained nicely in a structured table. . . . .  | 18 |
| 2.4 | Auditing screen. . . . .  | 20 |
| 3.1 | Technology Acceptance Model [18]. . . . .   | 24 |
| 3.2 | The returned results with Google Scholar using Ease of Use. It returns 222,000 hits, which is beyond human capability of examination. The returned results also include documents that have any of the keywords. . . . .  | 26 |
| 3.3 | The wiki-based resource from York University is a site that provides researchers with summarized information on theories widely used in information systems (IS) research. Visitors can find detailed information of IS theories, which are created and edited by visiting researchers. . . . . | 27 |

3.4	Grid-Enabled Measures(GEM) is an online repository of behavioral and social science measures. On this website, researchers have the ability to search, download, submit and provide feedback on measures. As of 7/20/11, the site contained a total of 147 constructs. . . . .	28
3.5	Theory Maps is another online repository for variables and measurement items. It is a collaborative project where the participant can input theories from publication and seel them as graph. The main purpose of the project is to allow researchers to manage bibliographic sources. . . . .	29
4.1	A fragment of WordNet 2.1 . . . . .	34
5.1	Main categories, correspondences and constructs. . . . .	55
5.2	Reconciliation in progress. . . . .	56
5.3	Example of hierarchical taxonomy in the gold standard. IS Development: Risk Factor and IS Development: Etc are the smaller categories which are broken down from IS Development. . . . .	58
5.4	Screen shot of the web-based application used to collect expert annotations. . . . .	60
6.1	ROC with LSA in terms of cosine cutoffs between 0 and 1. . . . .	67
6.2	ROC with ConSimMi in terms of cosine cutoffs between 0 and 1. . . . .	68
6.3	ROC with ConSimMi in terms of cosine cutoffs between 0 and 1. . . . .	69
6.4	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff . . . . .	72
6.5	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff . . . . .	73
6.6	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff. . . . .	74

6.7	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff. . . . .	75
6.8	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff. . . . .	76
6.9	Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff. . . . .	77
6.10	Frequency and cumulative frequency distribution of correspondent pairs across different cutoff overlayed with precision (success rate) at each cutoff. . . . .	78
6.11	The performance of possible combined models generated by ConSimLi in finding new correspondent construct pairs. . . . .	79
7.1	Unified Theory of Acceptance and Use of Technology Research Model. . . . .	82
7.2	Visual representation of the pattern of similarity of 14 root constructs appearing in Venkatesh et al.[84] that are computed with ConSimLi(I). Different colors are used to indicate the same group of root construct as reported in the paper. . . . .	86
7.3	Visual representation of pattern similarity generated with ConSimLi(I+D) . . . . .	89
7.4	Visual representation of pattern similarity generated with ConSimLi(I+N) . . . . .	90
7.5	Visual representation of pattern similarity generated with ConSimLi(I+D+N) . . . .	91
8.1	ConstructNet. . . . .	93
8.2	Connectivity for constructs pertaining to Perceived Usefulness and Perceived Ease of Use. . . . .	95
8.3	Connectivity for constructs pertaining to Cognitive Absorption. . . . .	98
8.4	Consolidate connectivity for constructs pertaining to Cognitive Absorption. . . . .	102

## **Chapter 1**

### **Introduction**

#### **1.1 Overview**

We present the main motivation that prompted this research. Next, we describe the common problems most scientists face when integrating existing knowledge. Then we discussed the goal and objectives of the research. Lastly, we present the brief outline of the thesis.

#### **1.2 Motivation**

For the last several decades, social and behavioral science has grown enormously. The increasing volume of theory has produced ample knowledge that is highly validated and solid, partly due to the strict operations and procedures of psychometrics. We do not have an exact number of theories, but we know there are thousands to tens of thousands and many thousands of extension articles. In the course of the last 70 years of research and development, constructs in social and behavioral science have been developed to cover the entire spectrum of human experience. This highly validated and broadly covered pool of knowledge is a gold mine of information that exists in the human brain. If this pool of knowledge could be mined and engineered with proper Machine Learning (ML) and Natural Language Processing (NLP) approaches, its potential could become invaluable to social, behavioral and computer science. Never before has anyone thought to tap into this gold mine to address prominent human decisions related to important social problems and health issues. In addition, it can be seen as a valuable source to leverage phenomena or concepts that only exist in the brain such as beliefs, intentions, perceived truths, motivation states,



expectancies, needs, emotions and social role perceptions in machine learning research.

### 1.3 Problems

Today, many scientists consider combining developed knowledge the greatest challenge of science. The following discussion highlights the obstacles currently hindering theory development research in multiple disciplines. We have chosen to focus on three prominent problems that relate to facilitation in social and behavioral science: construct proliferation, linguistic ambiguity, and disconnected constructs.

#### 1.3.1 Construct proliferation

In many social and behavioral disciplines, research focused on theory development has gained in prominence in the past decades, but the utilization of knowledge embedded in the development efforts has not kept pace [44]. Evidence shows that researchers are not making effective use of existing research studies [42]. The plethora and fragmentation of constructs in social and behavioral science has suggested that researchers prefer to propose new constructs over using existing constructs when developing new papers [42]. Normative science is additive, new research allows the theorist to refine, change, and adapt the existing theories. To date, effective approaches for discovery of newly developed or past theories from an integrated knowledge base do not exist.

#### 1.3.2 Linguistic ambiguity

In theory development, researchers use words to facilitate conceptual dialogues. Often, experts can disagree on the vocabulary and the meaning of the words used to represent concepts. Linguistic ambiguity is created if the wrong words are chosen or concepts are not put in context. Past research [24] reveals that people are less than 20% likely to express the same idea using the same words, which lead to construct correspondence, a scenario where different names have identical meanings. On the other hand, researchers, unaware of related works, tend to create independent constructs when using identical names with different definitions within different research areas.

### 1.3.3 Disconnected constructs

To the best of our knowledge, large-scale integration of constructs across multiple disciplines has never been attempted. Nunnally [61] made clear that constructs do not exist as isolated instances but are inter-related to one another. For instance, both the verbal and the mathematical dimensions of the construct intelligence must be assessed before judging an individual's intelligence. A nomological network represents the way in which one construct relates to other constructs and how those constructs interplay, allowing predictions to be made about unobserved constructs and their relations even before we create them.

## 1.4 Research Questions

Knowledge integration begins with synthesizing multiple knowledge bases into a common construct. Operationally, two constructs are deemed similar if the domain experts determine that one or more of the construct measurement items can also be used to measure the other construct. It is also part of the process in construct validity. In practice, two constructs presumed similar are validated by experts “operationally” through (1) formulating the assumptions of the relationships between constructs, (2) determining the corresponding relationships with the empirical measurement results (3) comparing the empirical relationships between the measures results with the corresponding assumed relationships between constructs and (4) interpreting the results of the comparison in order to determine the validity of measures. Measuring construct similarities this way is complex, laborious and time consuming.

Measuring constructs operationally becomes almost impossible when dealing with thousands of constructs collected from the past literature in social and behavioral science. This leads to the question of whether any of the current advancements in Natural Language Processing (NLP), specifically computational semantics, are capable of predicting construct relationships.

## 1.5 Research Goal and Objectives

The goal of the study is to explore and devise automated computational approaches, as an alternative approach to the conventional psychometric processes to predict construct relationships based on their textual properties. The study is further solidified and sharpened with the following three objectives:

- (1) Investigate, propose and evaluate computational text similarity measures for construct similarity computation based on names, definitions and items.
- (2) Appoint experts to manually label construct relationships, which will serve as a gold standard to benchmark the measures proposed in the objective mentioned above.
- (3) Demonstrate the applicability of the similarity measures inducing human decisions related to technology adoption through use cases.

## 1.6 Research Scope

The scope of our study of construct relationships is restricted to the Information System (IS) discipline only. We restrict our efforts to finding binary class relationships when measuring similarity: correspondent and independent, by using preset thresholds on the similarity scores.

## 1.7 Chapter Overview

Chapter 2 provides an in depth discussion of the main entity of the study, the constructs. We begin the discussion on the importance of the constructs and how they can be operationalized. Next, we give an overview of the problems that are hindering theory development. The chapter ends with our approach to collecting constructs for the study.

Chapter 3 presents various background material that are related to construct development. We discuss briefly the role of constructs in behavioral science and its connection to psychometrics. We also present existing attempts to address the problems.

Chapter 4 elaborates on our attempt to reveal construct relationships through computational approaches. We review a number of text similarity measures and carry out a pilot study where they are able to predict construct relationships.

Chapter 5 describes the creation of a gold standard, which is used to evaluate alternative approaches to prediction of construct relationships.

Chapter 6 presents various evaluation methods for finding the best prediction measure for construct relationships.

Chapter 7 describes the application of the similarity measure in predicting corresponding constructs in the Venkatesh et al. [84] unified model.

Chapter 8 presents a use case study on the network of constructs, which is known as ConstructNet.

Chapter 9 concludes the thesis by discussing the overall contribution of the research in the context of related work in the area. In addition, it addresses the limitations of our study and suggests future work.

## Chapter 2

### Constructs

#### 2.1 Overview

In this chapter, we begin our discussion by defining constructs and explaining how they are created throughout different disciplines. We then tackle the problems that are hindering theory development where constructs are used

#### 2.2 What are constructs

Constructs are the elements of behavioral theories. Cronbach and Murphy [15] (p. 464) defines a construct as “an intellectual device by means of which one construes events. It is a means of organizing experience into categories.” Constructs are also known as latent variables. The term latent variable implies two features of constructs (a) they are unobservable, e.g. anxiety and aspiration and (b) they are variable rather than constant, e.g. the level of anxiety changes over time. Although the constructs are latent and cannot be observed directly, their magnitude can be quantified through behavior. The phenomenon of behavioral constructs is usually reflected with a set of measurement items (or scales), which are used to quantify construct reflection through behavior, e.g. determining usefulness through productivity. For example, the construct *Perceived Usefulness* shown in Table 2.1 which is first appeared in Davis [17] has 6 measurement items.

Measurement items (or measurement instrument or sometimes known as scales), are a collection of statements or questions intended to reveal the levels of theoretical concepts or constructs. They are used to measure a phenomena we believe to exist but which cannot be observed and

Name	Perceived Usefulness
Definition	The degree to which a person believes that using a particular system would enhance his or her job performance.
Items	<ol style="list-style-type: none"> <li>1. Using the system in my job would enable me to accomplish tasks more quickly.</li> <li>2. Using the system would improve my job performance.</li> <li>3. Using the system in my job would increase my productivity.</li> <li>4. Using the system would enhance my effectiveness on the job.</li> <li>5. Using the system would make it easier to do my job.</li> <li>6. I would find the system useful in my job.</li> </ol>

Table 2.1: The table shows that the construct, Perceived Usefulness, as reported in Davis[17] has three textual properties: name, definition and items.

assessed directly. For example, if a person was given an opportunity to rate their productivity by how strongly they agree with each of the items, their underlying Perceived Usefulness should influence their responses[17]. Each item should be an indicator of how strong the Perceived Usefulness is. The score obtained on the item is caused by the strength or quality of the construct for that person at the particular time and space. Thus, these items have the cause of relationships to the construct. and they are intended as a measure to estimate the actual magnitude of the construct.

Constructs are non-specific concept representation in words and are easily fabricated, manipulated and interpreted [92]. When dealing with constructs, the two major concerns are (a) do the measurement components of constructs measure what they purport to measure? and (b) what are the structural components that interrelate constructs with one another? Construct validation is a process “to employ one or more measures whose results generalize to a broader class of measures that legitimately employ the same name” [61]. Psychometrics is a body of study governing the design of valid constructs.

### 2.3 Why constructs are important

Constructs are the cornerstone of behavioral theory. Constructs represent valuable units of concept and knowledge of theory in the fields of study. Concept and knowledge are important as they help us to understand and explain what actually is going on. They can be seen as the

building blocks of hypotheses in the behavioral and social sciences. For example, in psychology, they represent ideas such as *anxiety*, *aspiration* and *usefulness* that cannot be deemed as behavior. Constructs should be consistently relevant to a variety of domains within a field of study.

Constructs also serve as common language through which theoretical ideas and research findings are conveyed among researchers. They directly contribute to the progression of existing knowledge and theory.

## 2.4 How constructs are created

The transformation of concepts into constructs is known as construct conceptualization. It is a process of constructing scientists' imaginations into something that are abstract rather than concrete [61]. The lack of concreteness of construct conceptualization implies that there is no one prescribed empirical approach to define and create constructs. However, Barki [7] in his paper has generalized that they are four non-exclusive approaches to construct conceptualization, which are

- (1) Create a clear definition;
- (2) Specify a construct's dimensions and their relationships;
- (3) Explain how a construct applies to alternative contexts; and
- (4) Expan the conceptualization of a construct.

### 2.4.1 Create a clear definition

A clear and explicit definition has to be established for a construct once a concept is conceptualized. For example, to conceptualize *Perceived Usefulness*, a suitable definition must be created which reflects the research context. *Perceived Usefulness* has existed for a long time in the IS discipline within an extensive body of research. However conceptualizing the construct with a clear definition clarifies its relationship with the existing literature. In addition, it is suggested that different construct names be used to distinguish between behavioral, cognitive and attitudinal

constructs. This step is very useful in explaining time-dependent relationships that frequently exist between the constructs of different meta-categories.

#### **2.4.2 Specify a construct's dimension and their relationships**

Another way to conceptualize constructs is to identify the dimensions of a construct and the relationships that exist between them. This can be done by studying the related studies or other fields theories and discovering the specification of multidimensional constructs. The researcher has to analyze and decide whether or not all construct dimensions must be simultaneously present for the construct to exist. It should be noted that the conceptualized constructs may overlap when exploring other contexts. Such overlapping is expected when considering the conceptualized constructs are relevant with other concepts.

#### **2.4.3 Explore how constructs can be applied to other context**

The third approach in conceptualizing a construct is by analyzing how a given construct can apply in different technological, organizational, or individual contexts. For example, how constructs relationships in Technology Acceptance Model (TAM) [16] are vary between hedonic and instrumental contexts.

#### **2.4.4 Expand the conceptualization of constructs**

Construct development is part of the theory development where new theories are constantly revised and new knowledge is discovered through research by introducing new constructs. Constructs in most research models are narrowly defined to suit the particular context being studied. Such constructs can be conceptualized better by expanding the conceptualization to include a wider context so that it would better explain and support understanding of multifaceted and complex realities.



## 2.5 Psychometrics

Constructs are latent behavioral variables. Since they are unobservable, they are derived through a systematic procedure comparing the behavior of two or more people with carefully designed instruments. The systematic procedure is known as psychological testing. The key question here is how to ensure that constructs produced through psychological testing and reported in publications represent what they are supposed to conceptualize. Psychometrics comes into play here.

Psychometrics is the study of operations and procedures used to measure individual differences in behavior and to link those measurements to psychological theory or phenomena. The statement reveals three psychometric attributes: (1) tests involving behavior samples, that are (2) collected in a systematic way, and (3) with the goal of comparing the behavior of two and more people. In other words, it is about generating instruments and procedures use to quantify or measure a particular theory or phenomena [25]. The science of psychometrics is particularly concerned with the design of valid measurement instruments and reliable tests, which in consequence allows administration and analysis of the results with sustainable empirical evidence.

### 2.5.1 Validity and Reliability in Construct

Two key concepts in psychometrics are validity and reliability of a psychological test. Validity is the degree to which a test measures what is it meant to measure. Validity is an extremely important attribute in psychometrics as it concerns whether instruments used are capable of predicting specific events or an event's relationship to measures of other constructs. There is no single nominal index of test validity, rather there are a number of different approaches established to measure validity of a test. The more prominent type is construct validity. Construct validity is an effort to correlate between a theoretical concept and measurement items. For example, can the ability to solve mathematical problems be appropriately used to measure Intelligence? A test is only valid if it measures what it claims to measure.

Reliability of a test is about the proportion of variance related to the true score of the

construct to be measured, based on the observed scores obtained through measurement items. Reliability reflects the correlation between observed scores, true scores and error. True scores are unobservable “actual value” of constructs, whereas observed scores are raw data. The reliability is high when the differences among participants’ observed scores are consistent with the differences among their true scores, factoring in the error<sup>1</sup>. The reliability does not vary because of the differences in time, space and different measurement items being used. For example, the psychological trait intelligence is quite stable across time and its true value does not change much in weeks or months and it can be measured either through problem-solving and verbal ability. That says a reliable test should highly correlate and has reasonable low test-retest variability if it administered to the same group on two occasions.

The purpose of discussing psychometrics here is to make clear that construct conceptualization is strictly bound to standards and guidelines to safeguarding their quality. All constructs found in articles published in reputable journals have gone through thorough reviewing process and stringent selection to ensure that the results reported are of high validity and reliability. Despite this, a discipline’s validity and reliability evolves over time, and not all review processes are of identical quality. The different levels of validity and reliability affect the work in this dissertation and are identified when appropriate.

## 2.6 Problems

Though constructs contribute to the progression of science, theory development research these days has become multi-disciplinary, thanks to the number of new theories developed every year in all the relevant fields. However, questions arise about how to appropriately manage and integrate the knowledge. Currently, theory development is facing three problems: construct proliferation, linguistic ambiguity and construct disconnect.

---

<sup>1</sup> Classical measurement theory assumes that Observed score (O) is equate to True score (T) and Error (E). If E is reduced, O will converge on the true score of the concept.

### 2.6.1 Construct Proliferation

Theory development has grown over the past decades across multiple disciplines. For example, in the behavioral science discipline, Lee, Lee and Gosain [45] show over two hundred theories being developed and used in information system research alone, and Straker [76] lists over three hundred theories or models that have some bearing on persuasion. Although growth is positive, as more conceptualization efforts pour in from every direction to solidify scientific discoveries and develop more specific theories, the valuable knowledge embedded in the efforts does not keep pace nor converge to a conclusive body of evidence. For example, behavioral science research in IS, which is considered relatively well-defined, is being characterized as theoretically scattered [39, 62], fragmented [6], and chaotic [51]. One major factor that contributes to this predicament is researchers who are not well-informed of or considering related works. They are disconnected from earlier research even in cases where they conduct similar or even identical research. If these problems are left unattended, disciplinary knowledge is likely to continue to grow apart, resulting in the problem of related and even identical constructs that are seldom cited or reused. For example, IS research is derived from multiple disciplines such as psychology, economics, computer science, as well as management. As an applied discipline, IS research should be grounded in one or more of the underlying disciplines. Testing of hypotheses should be derived from the existing theories rather than from intuition. Furthermore, methods of testing should be compatible and derived from the existing disciplines. Normative science is additive, with new research allowing theorists to refine, change, and adapt the existing theories. To date, effective approaches for discovery of newly developed or past theories from an integrated knowledge base do not exist.

### 2.6.2 Linguistic Ambiguity

Zmud [92] (p. 29) described construct as "abstract entities and not specific, observable variables", and stated that "they are easily fabricated." Theory development makes use of words to facilitate conceptual dialogues. The problem of this is to decide what words to select, which refer

to events or things that are unobserved or functions that are unclear and unspecific. Researchers can disagree on the vocabulary and the meaning. If researchers fail to understand the specific conceptualization context and adopt the wrong words, linguistic ambiguity results. As Zmud pointed out, linguistic ambiguity "is not unique to the IS field," it happens in every discipline. Linguistic ambiguity causes failure among researchers to reach consensus on the underlying semantics of concepts and ideas. One of the main factors contributing to construct confusion is "simply lack of understanding or awareness of the structure and breadth of these fields and the measurement tools used to assess them" [28]. Also, Furnas, Landauer, Gomez and Dumais [24] found that different people are less than 20% likely to express the same idea using the same words. Larsen [42] discovered that in the IS implementation area alone, 83 unique concepts were measured using 948 different instruments and most of the research papers employing these instruments did not build on similar and existing ones. Thus, the use of different words creates ambiguity which leads to correspondent (similar) and independent (dissimilar) constructs.

### **2.6.3 Construct Disconnect**

Nunnally [61] suggested that a construct "... literally ... does not exist as an isolated, observable dimension of behavior." Relationships between constructs are of paramount interest to theory practitioners and researchers. Cronbach [14] envisioned a nomological network as an "interlocking system of laws which constitute a theory." A nomological network represents the way in which a construct relates to other constructs and demonstrates the interplay of constructs. Accordingly, defining a theoretical construct is a matter of elaborating the nomological network in which it occurs, "with the net, adding a construct or a relation to theory is justified if it generates nomologicals that are confirmed by observation" [14]. Although nomological networks sound like a promising panacea to all the discussed predicaments discussed in theory development, methodology to develop such a net has never been developed. Part of the reason is because the effort to weave constructs of theory into a graph-like presentation is formidable and requires huge financial support, laborious effort and great scholarship across different disciplines.

## 2.7 Evaluation of Constructs by Experts

There are always infinite ways to describe constructs. The assessment of constructs quality should not be focused on the “truthfulness” of the construct’s conceptualization, but should rather emphasize the constructs efficacy in predicting or explaining interesting or important phenomena. This is where the constructs contribute to the progression of science.

## 2.8 Nomological Network

A nomological network is a network of constructs and is originally proposed to ensure constructs validity. A nomological network is defined as the interlocking system of laws which constitutes a theory [14] (p. 290). It is a network representation of concepts (construct) of interest in a study, observable manifestations, and the relationships among them. This network would include the theoretical framework of what we are trying to measure, an empirical framework of how we are going to measure it, and interrelationships between these two frameworks. Constructs constitute a crucial part of these laws, and Cronbach and Meehl outlined the importance of learning more about a theoretical construct through elaborating the nomological network in which it occurs.

For example, while preparing items to measure a developing construct, **C**, it is useful to identify and study how **C** is semantically similar to and theoretically distinct from other constructs as this insight contributes to a clear definition for the construct to reflect its research context. Once having established the interrelationship specification to **C**, we can distinguish between construct measurement items that exhibit Convergent validity (evidence of similarity between measures of theoretically related constructs) and those that exhibit Discriminant validity (the absence of correlation between measures of unrelated construct). Therefore, by exploring the nomological Network, we can operationally define and detect applicable measurement items for a construct even when the constructs appears in different theories.

Although the nomological network might seem to be the right tool in guiding researchers develop and validate construct, it is never used in large scale construct development. Integrat-

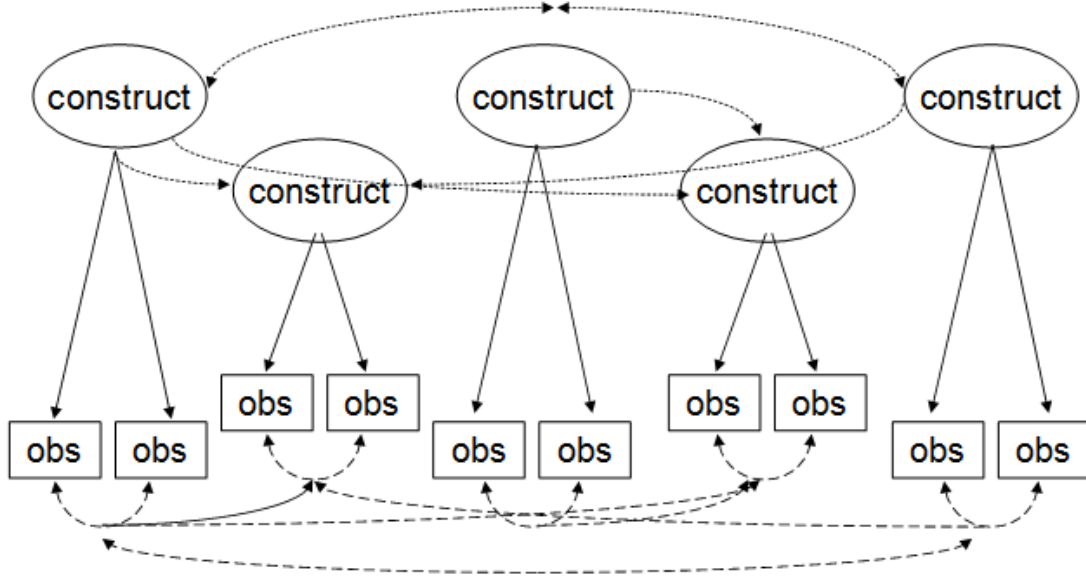


Figure 2.1: The Nomological Network developed by Cronbach and Meehl [14]. The figure shows the relationships of theoretical constructs to observables, observable properties or quantities to each other, and different theoretical constructs to each other.

ing constructs from multiple theories requires validation of every construct’s measurement item which needs huge investment of time and expertise, involving fusion of different theories which are developed under narrow areas and niche interests. This problem is further aggravated by the constructs nature of lack of concreteness, which are usually ambiguous and misinterpreted in the network. Constructs also vary in the nature of those close in description to those highly theoretical constructs. Finally, construct relationships cannot be expressed with a single simple coefficient and the integration of diverse constructs cannot be an entirely quantitative process.

## 2.9 From Nomological Network to ConstructNet

We can derive an interlocking system of constructs by loosening up the strict conditions imposed by a nomological network. Instead of cross-validating every construct measurement item through experiments, we can weave a theoretical construct network even when constructs have never been tested as part of the same theory or model, by computing the similarity of construct properties: name, definition and items. We call this network ConstructNet. Through various NLP

advancements, especially those that are used to compute text similarity, we can discover three types of construct relationships:

- (1) Correspondent constructs—constructs that are very similar in context, e.g. *Complexity* versus *Ease of Use*
- (2) Related constructs—constructs that are likely to be correlated, e.g. *Anxiety* versus *Depressed*
- (3) Independent constructs—Unrelated constructs, e.g. *Ease of Use* versus *Depressed*.

We intend to explore various similarity measure to predict construct relationships (see Chapter 4). The number of relationships between constructs grows exponentially, thus, 10,000 articles containing 50,000 constructs would have over 1.2 billion potential relationships. To restrict its scope, this study is limited to networks built upon correspondent and independent constructs.

## 2.10 Construct Collection

Any article published in a reputable journal containing at least one construct was a candidate for inclusion in the study.

Theories consist of two main elements: constructs and relationships. Each theory reported normally consists of a number of constructs representing perceptions, belief, motivation, attitudes, preferences, social acts and satisfactions. The second main element of the theories—relationships—are generally represented by hypotheses and the statistical findings of those relationships. The statistic reports the discrepancy of constructs and how they correlate empirically with one another.

We collected constructs that are reported between 1983-2009 in two information system journals: *Management of Information System Quarterly* (MISQ) <sup>2</sup> and *Information Systems Research* (ISR) <sup>3</sup>. Since these journals are perceived as reputable in IS, we operate on the assumption that

---

<sup>2</sup> <http://www.misq.org/>

<sup>3</sup> <http://isr.journal.informs.org/>

the constructs reported in the articles are validated. This assumption may be violated in some cases without major impact on the work.

Construct parsing is the process of extracting construct and their properties from the articles. We created a relational database to preserve the constructs that are of interest to our study. The data were constituted by:

- (1) All individual paragraphs in the originating paper (to develop language understanding);
- (2) Construct name;
- (3) Construct definition; and
- (4) Construct measurement items.

For each construct, its name, definition and set of measurement items was stored. The number of construct items varies between 2-20, depending on the complexity of constructs. For reference, every paragraph (except appendices and bibliographies) of the articles the constructs were extracted from was stored, and associated to the constructs in the database. In most cases, extracting constructs from articles can be a straight forward process if constructs are nicely presented inside tables (see Figure 2.2). Construct relationships were not stored, and they were left to be discovered with the proposed tools.

Instead of using computational approaches, the constructs are extracted by trained research associates (RAs). This is to ensure the data collected are of the highest quality as we are not aware of any automated extraction method that is able to do the level of accuracy we wished to have.

The parsing begins with careful selection articles from top journals in social and behavioral disciplines by the project investigator (PI). A set of articles, usually 10, are then examined by the senior research associates. While future considerations may require careful selection of articles that are highly related to the project and also warrant that the selected articles have a fair chance of referring to one another based on the articles citation. The selection can also based on the citation to the constructs already held in the database. For this project, any article that contains at least



## Empirical Validation of UTAUT

### Preliminary Test of UTAUT

Using the post-training data (T1) pooled across studies (N = 215), a measurement model of the seven direct determinants of intention (using all items that related to each of the constructs) was estimated. All constructs, with the exception of use, were modeled using reflective indicators. All internal consistency reliabilities (ICRs) were greater than .70. The square roots of the shared variance between the constructs and their measures were higher than the correlations across constructs, supporting convergent and discriminant validity—see Table 14(a). The reverse-coded affect items of Compeau and Higgins (1995b) had loadings lower than .60 and were dropped and the model was reestimated. With the exception of eight loadings, all others were greater than .70, the level that is generally considered acceptable (Fornell and Larcker 1981; see also Compeau and Higgins 1995a, 1995b; Compeau et al. 1999)—see Table 15. Inter-item correlation matrices (details not shown here due to space constraints) confirmed that intra-construct item correlations were very high while inter-construct item correlations were low. Results of similar analyses from subsequent time periods (T2 and T3) also indicated an identical pattern and are shown in Tables 14(b) and 14(c).

correlations. The results from the measurement model estimations (with reduced items) are not shown here in the interest of space.

An examination of these highest loading items suggested that they adequately represented the conceptual underpinnings of the constructs—this preliminary content validity notwithstanding, we will return to this issue later in our discussion of the limitations of this work. Selection based on item loadings or corrected item-total correlations are often recommended in the psychometric literature (e.g., Nunnally and Bernstein 1994). This approach favors building a homogenous instrument with high internal consistency, but could sacrifice content validity by narrowing domain coverage.<sup>11</sup> The items selected for further analysis are indicated via an asterisk in Table 15, and the actual items are shown in Table 16.

Tables 17(a) and 17(b) show the detailed model test results at each time period for intention and usage, respectively, including all lower-level interaction terms. Tables 17(a) and 17(b) also show the model with direct effects only so the reader can compare that to a model that includes the moderating influences. The variance explained at various points in time by a direct effects-only model and the full model including interaction terms are shown in Tables 17(a) and 17(b) for intention and usage behavior, respectively.<sup>12</sup> We pooled the data across the different

Figure 2.2: Excerpt of article content published in MISQ. Black boxes represent boundary of paragraphs to be extracted.

Table 10. Effort Expectancy: Root Constructs, Definitions, and Scales		
Construct	Definition	Items
Perceived Ease of Use (Davis 1989; Davis et al. 1989)	The degree to which a person believes that using a system would be free of effort.	<ol style="list-style-type: none"> <li>1. Learning to operate the system would be easy for me.</li> <li>2. I would find it easy to get the system to do what I want it to do.</li> <li>3. My interaction with the system would be clear and understandable.</li> <li>4. I would find the system to be flexible to interact with.</li> <li>5. It would be easy for me to become skillful at using the system.</li> <li>6. I would find the system easy to use.</li> </ol>
Complexity (Thompson et al. 1991)	The degree to which a system is perceived as relatively difficult to understand and use.	<ol style="list-style-type: none"> <li>1. Using the system takes too much time from my normal duties.</li> <li>2. Working with the system is so complicated, it is difficult to understand what is going on.</li> <li>3. Using the system involves too much time doing mechanical operations (e.g., data input).</li> <li>4. It takes too long to learn how to use the system to make it worth the effort.</li> </ol>
Ease of Use (Moore and Benbasat 1991)	The degree to which using an innovation is perceived as being difficult to use.	<ol style="list-style-type: none"> <li>1. My interaction with the system is clear and understandable.</li> <li>2. I believe that it is easy to get the system to do what I want it to do.</li> <li>3. Overall, I believe that the system is easy to use.</li> <li>4. Learning to operate the system is easy for me.</li> </ol>

Figure 2.3: In some cases, the constructs and their corresponding properties are contained nicely in a structured table.

one construct is included. Inclusion of construct articles is crucial because the exclusion of an

important article would have far-reaching consequences. Due to the importance of this selection, articles were only selected by PI and experienced research associates on the project. Each article was thoroughly read and examined by two research associates and the results compared.

Once articles fitting the definitions are identified and selected, they enter the parsing workflow. The parsing involved reading the articles, extracting constructs, and saving them into the database. This included paragraphs, construct names, definitions, measurement items, and references indicating the origin of the construct. The parsing task was usually carried out by junior research associates, who are well-trained by senior research associates to perform the task. In addition, only the top 2.5 percent of student applicants were hired based on the interview and a three-hour training and testing exercise. Those hired underwent extensive training and worked with experienced RAs until their performance met team expectations. A supportive team environment with regular team meetings, good salaries and opportunities for advancement ensures that interest and motivation to carry out the task is high. In case any article proves too hard to parse, it was escalated to senior researcher and then the PI, who has 13 years of experience extracting variables from academic papers.

Any construct that was parsed and stored in the database had to be verified by the PI and senior RAs. A web-based application was built for this purpose. The senior research associates could use the auditing functionality of the application to examine information about a paper and correct errors. This allowed the senior RAs who have more experience to audit and improve the quality of the work done by junior RAs. The work flow system routed errors and comments back to the parsing RAs for their learning. After the paper had been parsed and added to the system, an auto-generated email went out to the first author of the parsed article, who could dispute error corrections, if any.

A total of 9183 articles and 39215 constructs were parsed over two years covering IS, education and nursing disciplines. The study only uses 653 articles which are from IS discipline. All articles were published in 26-year period, and totalled to 1054 unique constructs. There were in total 5,526 measurement items and 19,646 paragraphs.

**Science, math, and engineering salience (21505)**

**Variable Information**

Delete Variable ☐

Name

Parent

Definition

Category

Dependency

**Variable Items**

# Text

Item	Text	Reversed
1	Thinking about how you would normally act, how likely are you to actually tell a new student in your school about your science, math, and technology activities?	<input type="checkbox"/>
2	Thinking about how you would normally act, how likely are you to actually tell a person you are attracted to about your science, math, and technology activities?	<input type="checkbox"/>
3	Thinking about how you would normally act, how likely are you to actually tell someone striking up a conversation with you in a waiting room about your science, math, and technology activities?	<input type="checkbox"/>

Add More Full set of items

**Changes**

User	Date	Changes	Compare Record
Anne Housman	5/12/2010 1:16:23 PM	Add Variable Item	<input type="button" value="Compare Record"/>
Anne Housman	5/12/2010 1:16:23 PM	Add Variable Item	<input type="button" value="Compare Record"/>
Anne Housman	5/12/2010 1:16:23 PM	Add Variable Item	<input type="button" value="Compare Record"/>
Anne Housman	5/12/2010 1:16:23 PM	Insert Variable Reference	<input type="button" value="Compare Record"/>
Anne Housman	5/12/2010 1:16:23 PM	Insert Variable	<input type="button" value="Compare Record"/>

© 2010 Leeds School of Business & Regents of the University of Colorado | Help | Report Problem

Figure 2.4: Auditing screen.

## 2.11 Conclusion

We described and highlighted the importance of constructs. Constructs reported in research publications usually are constituted of three features: name, definition and a set of measurement item, and are used mainly to represent unit of knowledge and commonly language among researchers. Current development of constructs in various discipline has led to construct proliferation, linguistic ambiguity and disconnected constructs. The problems pose hindrances to the progression of science in multiple disciplines.

We also discussed the data collection process. We relied on human experts to parse the constructs from articles published in reputable journal to ensure the highest quality of data,. We also built an interactive work-flow system to easily manage and maintain the quality of construct in our database.

Knowledge integration is a challenge to scientists due to the volume of published studies, theories and constructs. Cronbachs proposed approach of finding similar constructs through a nomological network proposed has never been implemented. With the extracted constructs , we

attempted to carry out experiments to solve the problems using NLP approaches by measuring the semantic similarity of constructs reflected in their features.

## **Chapter 3**

### **Background**

#### **3.1 Overview**

In this chapter we introduce the goal of the project that prompted the study. Then, it is followed by a discussion with an example about theories in order to inform its relationship to constructs. We end the chapter with the limited solutions that are used to overcome the hindrances when developing theories.

#### **3.2 Human Behavior Project (HBP)**

Scientists realize that it is impossible to find and incorporate all related disciplinary knowledge. The Human Behavior Project (HBP) aims to integrate scattered behavioral science knowledge. This can be accomplished through extracting constructs from top social and behavioral journals across the disciplines which are produced using psychometric method. It is also the intention of the project to produce a one-stop human behavior search engine portal to the search of high-quality knowledge from multiple disciplines.

#### **3.3 Human Decision Making through HBP**

Humans are complex organisms and their behaviors are hard to foresee. However scientists have discovered that 93% of human behavior is actually predictable [75]. Behavioral science is a body of science aiming to understand human behavior and its motivators reflected by cognition [36]. It is about studying the nonuniformities of human behavior at the same time identifying

factors that influence and determine human conduct with scientific evidence. Behavioral science provides answers to the question of why people act in specific ways and how to induce that action. It encompasses multiple disciplines such as psychology, sociology, and anthropology, that explore the activities of interaction and cognition among humans through systematic experiments such as field surveys and experimental studies. Through constant development of new theories, behavioral science is capable of offering cost-effective decision support through predicting human understanding or initiating behavior change and change the course of the outcome [92]. Theory is a master plan for decision-making. Researchers use theory to find the answers to the questions of “what”, “why” and “how”. These questions explain the nature of behavior; why individuals invoke certain actions or do not engage activities [26].

### 3.4 Theory

A theory is knowledge that comprises facts, assumptions and hypotheses. The general goal of theory is to understand reality systematically. The definition of theory according to Glanz et al [26] is “a set of interrelated concepts, definitions, and propositions that presents a systematic view of events or situations by specifying relations among variables in order to explain and predict events or situations.” The definition clearly pinpoints two very important aspects, (a) variables (or constructs, in our context), and (b) interrelationships. Theory allows us to visualize are the most important construct and how the constructs react and interact with one another. Without theory, when interventions take place, we might address the wrong issues or variables or hit just a small proportion of constructs required to have the desired effect. In the following discussion, Technology Acceptance Model will be used as an example of theory to reinforce our discussion.

The acceptance and effective utilization of IS by individuals and organizations are areas of research that have gained importance in recent years. User acceptance modeling has been an important field of study in IS in the past 20 years. Davis’ [16] Technology Acceptance Model (TAM) introduced the most widely used research constructs in information systems theory and models how users come to accept and use a technology. The model is rooted in the prior work of the Theory

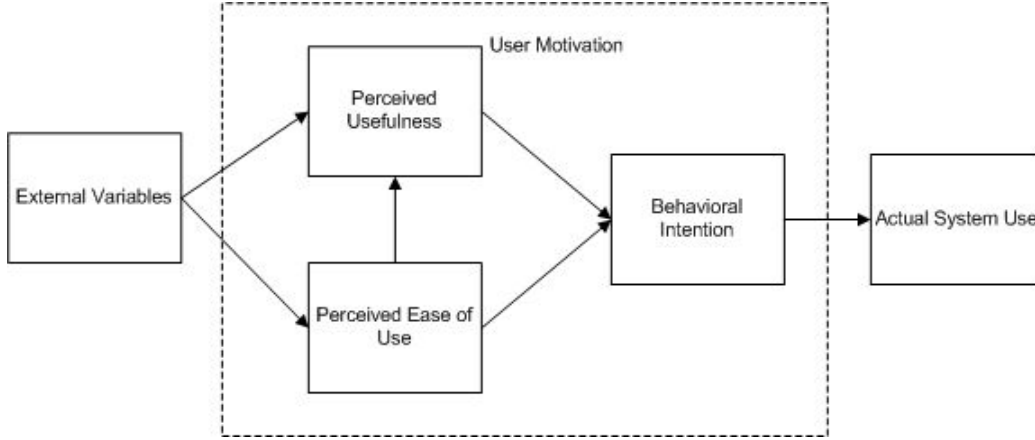


Figure 3.1: Technology Acceptance Model [18].

of Reasoned Action (TRA), a social psychology model by Fishbein and Ajzen [22]. The model proposed that when users are introduced to a technology, a number of factors affect their decision about when and how they will use it. According to Davis [17], there were three factors (constructs) that can be the predictor of users' motivation: *Perceived Ease of Use*, *Perceived Usefulness*, and *Behavioral Intention*. He hypothesized that the major determinant of whether the user will actually use or reject the system is mainly driven by user's behavioral intention using the system. *Behavioral intention*, in turn, was influenced by two major constructs:

- (1) *Perceived Usefulness* - "the degree to which a person believes that using a particular system would enhance his or her job performance" [18](p. 320); and
- (2) *Perceived Ease of Use* - "the degree to which a person believes that using a particular system would be free from effort" [18](p. 320).

Accordingly, *Perceived Ease of Use* directly influences *Perceived Usefulness*. Both of the two constructs were believed to be influenced by the *External Variables*, such as system design characteristics user training, and the nature of the implementation process.

TAM, which is a theory, can be a valuable and cost-effective tool for screening and evaluating systems or applications, and reliably predicting whether they will be accepted by users before the users get heavily involved in the technology. Prior research has shown that TAM parsimoniously

predicts user acceptance of technologies usage [44]. Though TAM is not a descriptive model, that is, it does not provide diagnostic capability for specific flaws in technology or systems, it can serve the purpose of screening a new system and predict product acceptability. TAM suggests that such evaluations can be made very cost-effectively because user perceptions of a technology are formed very early just after his or her initial exposure to the system [18]. It can thus provide a valuable tool to practitioners who aim to support organization functions through information systems. TAM has been applied in numerous studies testing user acceptance of information technology, for example, word processors [17], spreadsheet applications [53], e-mail [78], and web browser [58], and has evolved into such theories as the Unified Theory of Acceptance and Use of Technology (UTAUT) [84].

### **3.5 Solutions**

Many researchers are aware of the problems that hinder theory development. The following sections are some of the attempts to integrate knowledge that are taken at a smaller scale, normally just within one discipline.

#### **3.5.1 Search Engines**

Because the evolution of the information age has led to the development of numerous search engines such as Google, Bing, and Yahoo! on the World Wide Web (WWW), offering search facilities on a wide spectrum of subjects, we would expect to find anything on the WWW through these search engines. However, current search engines fall short of ability to find specific theories for decision-making. Imagine a questionnaire researcher who needs to investigate the acceptance of a new technology to an organization, a search for "ease of use" will return hundreds of thousands of documents which contain those words from any site. At the time of writing, the search on Google using the keywords mentioned returned about 220,000 results, which covered areas from medical biology to pedagogy. The abundance of information obtained this way requires a huge effort of filtering and assimilation to be useful and meaningful. In other words, it is difficult to perform a



specific search with precise results. One concern about WWW search is that results are displayed in relation to the number of visits from users or links from other articles or web pages, not in relation to its content coverage or another index of quality of the construct.

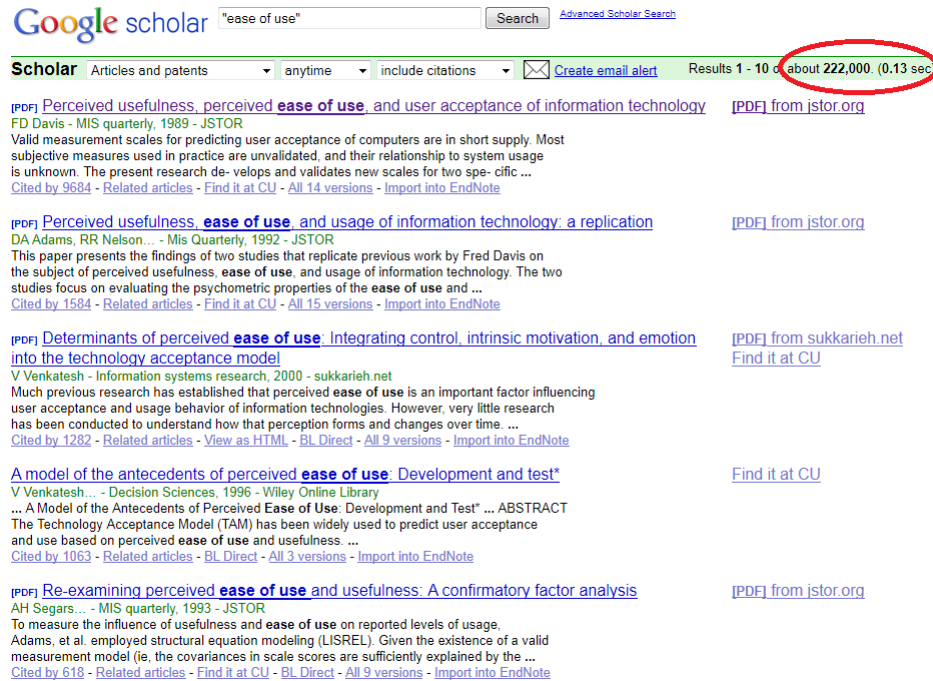


Figure 3.2: The returned results with Google Scholar using Ease of Use. It returns 222,000 hits, which is beyond human capability of examination. The returned results also include documents that have any of the keywords.

### 3.5.2 Reference Resources

Currently, research theories can only be traced by exploring outdated wikis <sup>1</sup> (see Figure 3.3), topic dependent databases <sup>2</sup> (see Figure 3.4), collaborative repositories <sup>3</sup> (see Figure 3.5) and human compiled encyclopedias [20, 74]. Researchers may overlook meaningful information and closely related constructs from different studies, whether old or new, if there is little effort made to integrate knowledge. Pinsonneault and Kraemer [66] describe this scenario by stating that "survey questions have accumulated in a truly mountainous supply" which complicates the matter over

<sup>1</sup> [http://www.fsc.yorku.ca/york/istheory/wiki/index.php/Main\\_Page](http://www.fsc.yorku.ca/york/istheory/wiki/index.php/Main_Page)

<sup>2</sup> <https://www.gem-beta.org/Public/ConstructList.aspx?cat=1>

<sup>3</sup> <http://www.theorymaps.com/>

time. Although it is still possible to find a specific construct in a very time-consuming manner, it is difficult to identify and connect semantically related constructs that appear under different names. Furthermore, this requires adept pre-existing knowledge of a niche area, abundant experience in article review, survey compilations and article level searches in commercial databases.

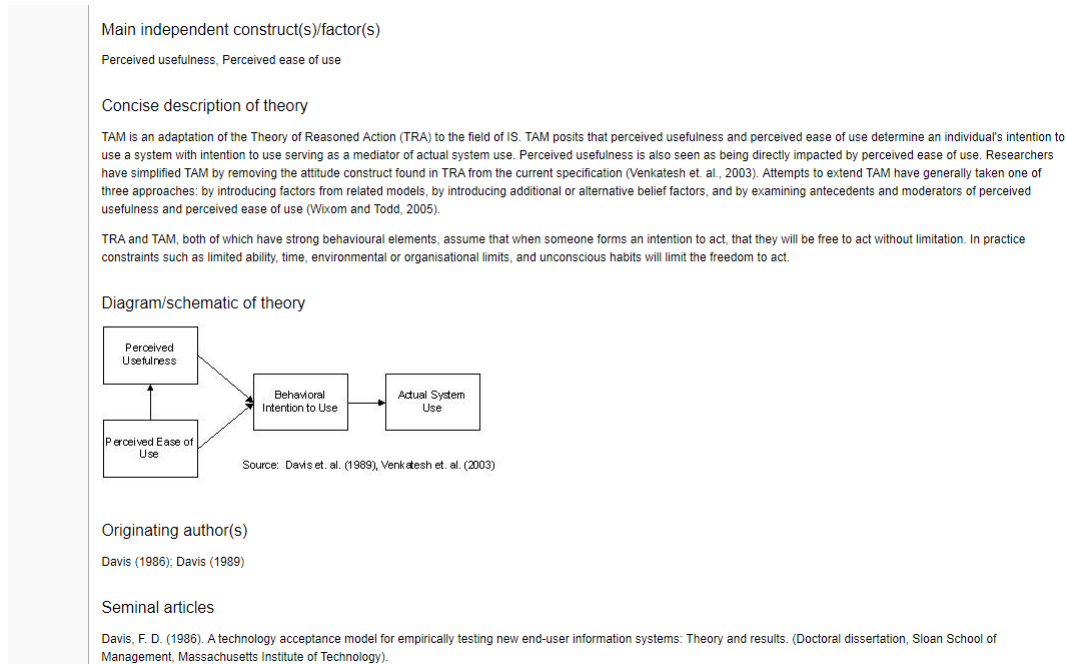
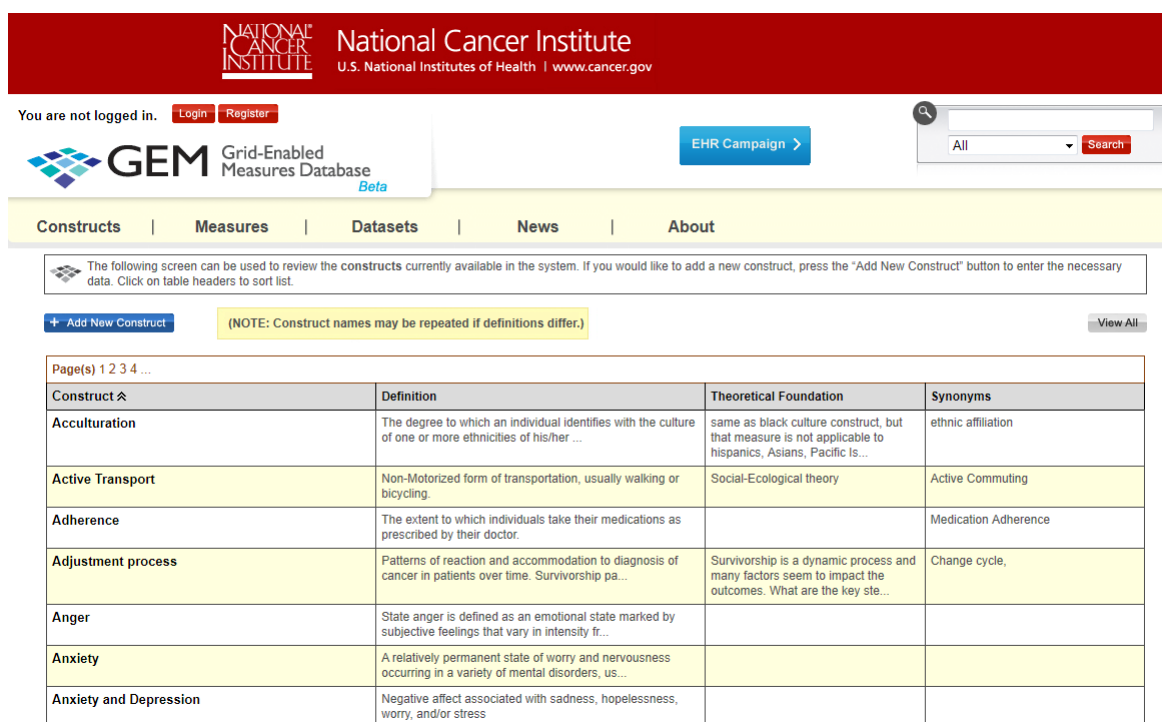


Figure 3.3: The wiki-based resource from York University is a site that provides researchers with summarized information on theories widely used in information systems (IS) research. Visitors can find detailed information of IS theories, which are created and edited by visiting researchers.

### 3.5.3 Citation Analysis

Citation analysis [55] is the examination of the frequency a paper is cited, assuming that influential scientists and high impact works are cited more often. Currently, it is considered one of the most effective approaches to discovering related knowledge. Citation analyses explore shallow connections among cited papers but are not able to uncover the links where they should exist but do not. The connections discovered through citation analysis relate papers rather than units of knowledge and these connections should not be regarded as existing knowledge. It excels at exposing general structures and clusters of papers but cannot really connect specific research findings from



The following screen can be used to review the constructs currently available in the system. If you would like to add a new construct, press the "Add New Construct" button to enter the necessary data. Click on table headers to sort list.

[+ Add New Construct](#) (NOTE: Construct names may be repeated if definitions differ.) [View All](#)

Construct	Definition	Theoretical Foundation	Synonyms
Acculturation	The degree to which an individual identifies with the culture of one or more ethnicities of his/her ...	same as black culture construct, but that measure is not applicable to hispanics, Asians, Pacific Is...	ethnic affiliation
Active Transport	Non-Motorized form of transportation, usually walking or bicycling.	Social-Ecological theory	Active Commuting
Adherence	The extent to which individuals take their medications as prescribed by their doctor.		Medication Adherence
Adjustment process	Patterns of reaction and accommodation to diagnosis of cancer in patients over time. Survivorship pa...	Survivorship is a dynamic process and many factors seem to impact the outcomes. What are the key ste...	Change cycle,
Anger	State anger is defined as an emotional state marked by subjective feelings that vary in intensity fr...		
Anxiety	A relatively permanent state of worry and nervousness occurring in a variety of mental disorders, us...		
Anxiety and Depression	Negative affect associated with sadness, hopelessness, worry, and/or stress		

Figure 3.4: Grid-Enabled Measures(GEM) is an online repository of behavioral and social science measures. On this website, researchers have the ability to search, download, submit and provide feedback on measures. As of 7/20/11, the site contained a total of 147 constructs.

different papers. Citation analysis operates at the article level which makes it a poor choice for construct discovery. Theory development without careful examination of the existing literature and constructs is one of the sources of construct proliferation. Researchers unaware of the synonymous constructs or scales are likely to create constructs that will never be cited [13]. It is a sobering fact that 90% of papers published in academic journals are never cited [55].

### 3.5.4 Meta Analysis

Meta-analysis is a statistical approach which combines the results of studies that address a shared research hypothesis. It is able to test the relationship between two variables such as X affects Y. Meta-analysis presents a possible alternative that allows the analysis of the relationship between two pools of constructs. However, such analysis requires intensive resources and for this reason, it is not a fit for cross-discipline knowledge integration, but works well for narrow problems, such as the

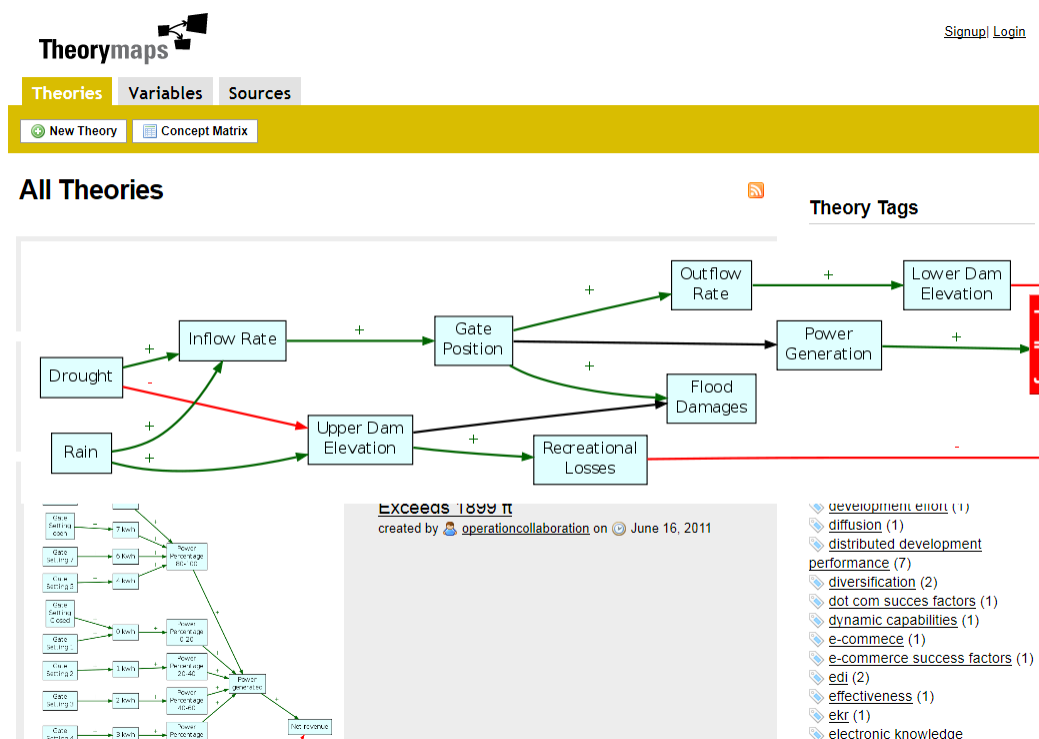


Figure 3.5: Theory Maps is another online repository for variables and measurement items. It is a collaborative project where the participant can input theories from publication and see them as graph. The main purpose of the project is to allow researchers to manage bibliographic sources.

relationship between two variables. Integrating knowledge using a meta-analytic approach requires the knowledge be validated as it depends heavily on qualitative evaluations of research constructed by the researchers. If the researcher treats dissimilar constructs as the same, the results become unreliable. In addition, if the research fails to identify identical constructs, the result may not be meaningful or representative of current research knowledge

### 3.6 Conclusion

Constructs are behavioral variables. In this chapter, we discussed the role of constructs in theories from social and behavioral sciences. It is clear that the constructs are the factors that influence human conduct as they are reflected through human behavior. HBP aims to integrate behavioral science knowledge from multiple disciplines to understand human decision-making. The chapter also showed that the available solutions are not suitable to perform knowledge integration

at large scale.

## **Chapter 4**

### **Computational Approaches to Detect Similar Constructs**

#### **4.1 Introduction**

In this chapter, I present and discuss various computational approaches from the past literature that are used for measuring text or sentence similarity. I adopt and adapt existing automated computational approaches to expedite the process of discovering construct relationships with minimal human supervision. Recognizing that I am judging the construct similarities based on the construct properties which are made up of short natural text, I am in fact dealing with the problem of semantic analysis.

This chapter begins with the discussion of past and related works in sentence similarity measures. Next, it describes in detail the approaches adopted in this study, followed by a description of how to use the gold standard to evaluate the efficiency of the approaches on different construct properties.

#### **4.2 Text Similarity Measurement**

The discussion on constructs in the previous chapters has formed a conjecture that construct similarities can be computed then predicted through the construct properties of name, definition and items. These properties, which are directly extracted from various renowned journals, are basically natural sentences, with an average of 11 words. We hypothesize that the semantic similarity in these properties can be seen as a significant indicator to predict the relatedness degree of construct relationship. The study surrounding such measures has been commonly known as text or sentence

similarity measure.

There is no shortage of literature on sentence similarity measures. It has increasingly gained ground since the 1970 in Natural Processing Procrssing (NLP) especially in the application of information retrieval, question answering applications, relevance feedback, text classification, word sense disambiguation, machine translation and text summarization. The wide deployment of sentence similarity measures in numerous areas has made it a generic component in text-related knowledge representation and discovery.

In general, works related to sentence similarity can be classified into four categories: Document Vector and word-occurrence methods; Corpus-based methods; Hybrid method; and Descriptive method.

#### **4.2.1 Document Vector and word-occurrence methods**

The Vector model is regarded as one of the earliest application of sentence similarity measure. The vector model, which was first proposed by Salton and Lesk [73], is used widely for information retrieval where the model is used to rank the most relevant documents given a query document. The model is also known as the “bag of words” method where it only retains word occurrences but discards function words and syntactic information. The model consists of a  $n$ -dimensional space which representing  $n$  words in the model. Documents in the model are represented as  $n$ -dimensional vectors where  $n$  can span from thousands to hundreds of thousands which are subjected to the availability of the vocabulary in a corpus. The values of the vector are usually the occurrences of the corresponding word in the document. To compute the similarity of the two given documents, each document is turned into an  $n$ -dimensional vector and the similarity score is computed using a similarity metric, which is normally a Euclidean distance or cosine of an angle. To date, this approach has been proven to be one of the best techniques and remains as one of the most effective method for information retrieval [80, 57]. The vector model is simplistic. However, the representation of documents in the model can be very sparse if  $n$  is huge and documents are not associated well if different words but a similar context are used.

Another straight-forward method is counting the co-occurrence of words of two documents [79]. Word-occurrence is basically rooted on the hypothesis that more similar documents have more words in common. Such a method works effectively when dealing with long documents due to the high degree of word co-occurrence but is rendered inoperative on short texts where word co-occurrence is basically null. Similarly to the vector model, the method is not able to detect any degree of similarity if different words are used to express the same meaning, and may in fact be misleading due to use of pronouns, prepositions, conjunctions and articles.

#### 4.2.2 Corpus-based methods

The corpus-based model[31], sometimes also known as the distributional model [54, 23, 35], is an extension to the vector model. It attempts to overcome the incorrect assumption that documents and words represented in a vector space are orthogonal to others, given the fact that a term, in human linguistics, can always be replaced with other or combined with other terms. For this reason, matrix factorization and other compression algorithms have been used on term-document matrices to analyze relationships among words in a corpus. One popular example of such compression techniques is Singular Value Decomposition (SVD). The singular vectors and corresponding singular values resulting from SVD allow words and documents to be mapped into the same "semantic space" based on the assumption that all words are presumably related to each document either due to synonymy or different meaning. The resulting semantic space places similar words and documents as measured by co-occurrence near to one another even if their words never co-occurred in the documents. One of the most well-studied corpus models is Latent Semantic Analysis (LSA) [19, 40], which is described in detail in the following section. Hyperspace Analogues to Language (HAL) [10] is also another important corpus model which works similarly to LSA. It assumes that two words are semantically related if they are seen to appear with the same word.



### 4.2.3 Hybrid Method

Work on developing promising word similarity measures based on the synonymy and similarity of words has catalyzed the discovery of various sentence similarity computations. Linguistic measures such as Lesk Algorithm [46], Resnik similarity[70] and Jiang Conrath distance [32] use proposed knowledge-based similarity measures based on the semantic distance of words in readily available thesauri like Wordnet [56] and MeSH [49]. The basic idea is to define a looser metric to indicate a synonymy relationship and semantic distance between words using word senses which are encoded as hierarchical structure in these knowledge bases. Figure 4.1 shows a fragment of the Wordnet hierarchy.

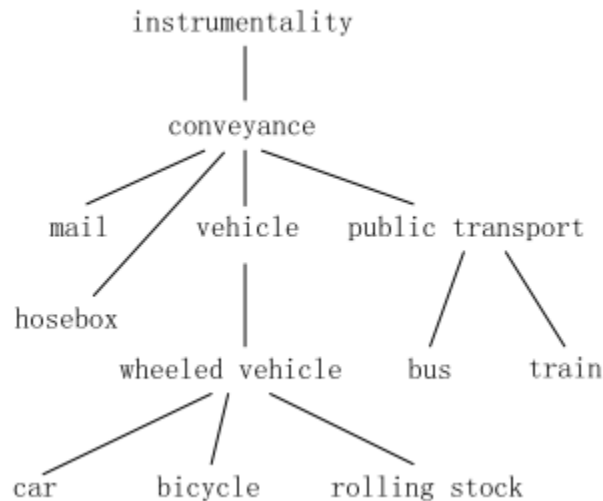


Figure 4.1: A fragment of WordNet 2.1 .

Two words are very similar if they share more semantic features or are close to each other; two words are deemed less similar if they have very few meaning elements or greater semantic distance. The similarity of words in this context is a function of word sense distance in the hierarchy of the knowledge bases. The semantic distance can be measured by taking horizontal length path, or vertical length path, or a combination of both. The prominent advantage of the knowledge base model over the previously discussed vector and corpus models is that it has a richer vocabulary

which is not bound to the available words in a corpus. The downside of the knowledge base models are their lack of up-to-date words and inability to work well within new or domain-specific words.

Hybrid methods facilitate and combine both the corpus based methods and the knowledge encoded in a thesaurus methods to determine text similarity. Commonly, words from a text are compared to other texts using the thesaurus model to drive a score to indicate similarity degree, combined with the information that is obtained through the corpus, which in turn is mapped into a combined function to indicate sentence similarity. The hybrid model also includes work that includes the corpus-based model such as LSA and HAL, which is used to replace Wordnet especially for the works that focus on the very specific area.

Apart from the corpus-based model, corpus information such as the language models, specifically Inverse Document Frequency (IDF) [33] and Pointwise Mutual Information [81], and Information Content [47, 31] have been reported as being used to derive text to text similarity metric.

#### **4.2.4 Descriptive method**

The Descriptive method [47] is a supervised learning approach that attempts to represent the constituents of a sentence using features such as their relation, lemma, part-of-speech (POS), voice and semantic roles. The descriptive method models each text as a vector and the values of the element are the corresponding features. The vectors are then generalized with machine learning schemes to produce a trained classifier. Similarity between texts is computed by projecting them into the classifier to obtain a similarity score.

### **4.3 Review of Selected Sentence Similarity Approaches**

As discussed above, the existence of a wide variety of sentence similarity measures rooted in different linguistic theories is evidence that computing the similarity of the sentences is not a trivial task. Nearly all measures have the difficulty of correct capture of the semantics or meaning in the sentences. For this reason, we have examined measures that integrate knowledge bases, measures that incorporate semantic roles, measures that compute corpus statistics, and the combination of

these. The main reason behind these different methods is to capture sentence semantics through different aspects and characteristics. We also intend to predict construct relationships through their properties with different text similarity measures.

In this section, we are providing an in-depth review and discussion of three sentence similarity measures that are rooted in different frameworks and that have gained noteworthy attention in the field of study. The three measures to be discussed in the following section are: LSA [19]; Li et al. sentence similarity measure (aka Li et al.) [47]; and Mihalcea et al. Sentence similarity measure (aka Mihalcea et al.) [57].

#### 4.3.1 Latent Semantic Indexing (LSA)

Latent Semantic Analysis (LSA), also known as Latent Semantic Indexing (LSI) when it is used in document retrieval, is a popular natural language processing technique created by Deerwester et al. [19]. It is regarded as a theory and also a computational method for extracting and representing the meaning of words. A “bag of words method”, the underlying idea of LSA is the assumption that a pool of documents pose a set of mutual constraints which determine the semantic similarity of words and set of words. The constraints could be thought of as latent links between the words and their context and mutual dependencies on each other. Thus, when two terms have occurred in the same context, even though they do not occurred in the same text, LSA is able to regard them as having similar meaning.

The mathematical foundation of LSA is the linear algebraic theorem known as the Singular Value Decomposition (SVD). Documents are represented as the  $A$  matrix, which is a term-by-document sparse matrix and after optional weighting and normalization, SVD is used to decompose the matrix into three matrices:  $U$ , a term by dimension matrix representings words;  $S$ , a singular value matrix; and  $V$ , a document by dimension matrix representing documents. The equation can be written as

$$A = USV^T \tag{4.1}$$

$U$  and  $V$  are orthogonal matrices whereas  $S$  is a diagonal matrix with main diagonal entries sorted in decreasing order. In practice,  $A$  could be approximated with  $A_k$  by preserving the first  $k$  singular values and the corresponding first  $k$  columns in  $U$  and  $V$ . The approximation can be written as

$$A_k \approx U_k S_k V_k^T, \quad (4.2)$$

where  $U_k$  is a term-by- $k$  matrix,  $S_k$  is a  $k$ -by- $k$  matrix and  $V_k$  is a document-by- $k$  matrix. The interesting thing about this approximation is not only that  $A_k$  has minimal error, but it translates the term-by-document matrix into a correlated topic space. In consequence, each row vector of  $U_k S_k$  represents a word in the topic space, and has  $k$  columns which give the occurrence of the word in the topic space. Likewise, each row vector of  $V_k S_k$  represents a document vector which correlates topics in the topic space.

In short, the SVD process on term-document matrix  $A$ , generated  $U$  and  $V$  to represent words and document respectively in the reduced space. The low-rank approximation through preserving the first  $k$  diagonal elements in  $S$  has produced the mutual constraint among words in different documents.

#### 4.3.1.1 Finding similar meaning words

Given a word,  $w$ , to find similar words in the topic space,  $w$  is projected into the reduced space to become a word vector,  $\vec{w}$ . This is done through finding the  $w$  corresponding vector in  $U$ . It is then compared with each row of  $U_k S_k$  using cosine similarity measurement. Words that have higher cosine are deemed more relevant than those that have lower cosine. To compute similarity of two words, the words are projected into the space to obtain their corresponding vectors, and their similarity is the cosine value.

$$Sim_{LSA}(w_1, w_2) = \cos(U_{(w_1,s)} S_k, U_{(w_2,s)} S_k) \quad (4.3)$$

#### 4.3.1.2 Finding similar documents

Finding similar documents in the reduced space is known as Latent Semantic Indexing (LSI). Given a query,  $q$ , which is either a word or document. Query vector,  $\vec{q}$ , is obtained through  $U$  if the query is a word, otherwise, is an aggregation of word vector in the sentence. To find a cluster of documents that are semantically closed to  $\vec{q}$  in the topic space,  $\vec{q}$  is compared against  $V_k S_k$ . Documents that are related are those with high cosine score.

#### 4.3.2 Li et al. Similarity Measures

The sentence similarity measure proposed by Li et al. [47] is conspicuous among others because it is reported to work well on short texts[47]. One of the main reasons is it useful is it attempts to preserve the word order information in sentences which is normally lost in the bag of words method. The word order is regarded as crucial information because it represents the syntactic information, which when combined with semantic information conveys the precise meaning of sentences. The Li et al. similarity score is basically derived with the following information: Semantic similarity between words; Word order similarity; and Statistic of corpus.

Clearly, the sentence similarity is an aggregated function of word similarity, word order in sentences and corpus information content. The word similarity is computed by taking the path length and the depth of two words in the hierarchical semantic knowledge base, i.e. Wordnet [56]. Word order similarity is then factored in, which is obtained by turning each sentence into a vector by ordering the words as they appear and computing the difference of word order. Finally, in order to separate the informative words from those that are not, information content of word is derived statistically from the Brown corpus [52] and is normalized onto each word.

Li et al.'s argument for using Wordnet to measure word similarity is because it is one of the richest and most accurate lexical dictionaries ever to have been crafted yet is readily available and does not adhere to a specific domain. It is reported in Miller [56] that similarity measures based on Wordnet correlate well with human judges. In addition, Li et al. report that it is useful to

incorporate the word order into the approach to compromise the loss of information with the bag of words method, especially with short texts. Finally, corpus statistics is used to reflect the actual usage of words which gives the ability to adapt the application to a specific application.

#### 4.3.2.1 Measuring Sentence Similarity

The following illustrates the procedures for computing the sentence similarity between two candidate sentences. Given two sentences,

$$S_1 = \{\text{RAM keeps things}\},$$

$$S_2 = \{\text{The CPU uses RAM}\}.$$

The process begins with forming a joint word set from the sentences. The joint word set,  $J$ , is basically all the unique words from the sentences. Note that words in  $J$  are not preprocessed or stemmed and they remain as they appear in the sentences.

$$J = \{\text{RAM keeps things The CPU uses}\}$$

Once the joint word set is formed, each candidate sentence is mapped to  $J$  to produce a lexical semantic vector. The elements of the lexical semantic vector represent words in the joint word set and their values are the highest similarity of words from the candidate sentence.

	RAM	keeps	things	The	CPU	uses
RAM	1	0	0	0	0	0
keeps	0	1	0	0	0	0
things	0	0	1	0	0.2802	0.4433
<i>Sim</i>	1	1	1	0	0.2802	0.4433
Weight	I(RAM) I(RAM)	I(keeps) I(keeps)	I(things) I(things)		I(CPU) I(things)	I(uses) I(things)
$\vec{S}_1$	0.390	0.330	0.179	0	0.074	0.08

Table 4.1: Deriving semantic vector for  $S_1$ . The similarity scores were computed using Equation 4.4

	RAM	keeps	things	The	CPU	uses
The	0	0	0	1	0	0
CPU	0	0	0	0	1	0
uses	0	0	0.25	0	0	1
RAM	1	0	0	0	0.33	0
<i>Sim</i>	1	0	0.25	1	1	1
Weight	I(RAM) I(RAM)		I(things) I(uses)	I(The) I(The)	I(CPU) I(RAM)	I(uses) I(uses)
$\vec{S}_2$	0.190	0	0.16	0.03	0.389	0.04

Table 4.2: Deriving a semantic vector for  $S_2$ . The similarity scores were computed using Equation 4.4

Table 4.1 shows the process of deriving the lexical semantic vector of  $S_1$  from the joint sentence. The first row in the table represents words from joint word set  $J$ , and the first column represents words in sentence  $S_1$ . All words are listed in the order they appeared in both  $J$  and  $S_1$ . For the words that co-occur in both  $J$  and  $S_1$ , the value is set to 1 at the cell of cross point to represent exact match (the first three diagonal cells of “RAM”, “keeps”, “things”). Otherwise, the cell at different words’ cross point (e.g RAM-keeps) is corresponded to the highest similarity score, which is computed with the Equation 4.4, and if and only if the score exceeds the preset threshold, 0.2. For example, the word “CPU” is not in  $S_1$  but the most similar word is “things”, with a similarity of 0.2802. Thus, the cell at the cross point of “CPU” and “things” is set to 0.2802, as it exceeds the threshold of 0.2. Note that most cells that have 0 as their similarity scores are less than 0.2.

$$s(w_1, w_2) = e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} \quad (4.4)$$

Equation 4.4 is a word-word similarity function, where  $\alpha$  and  $\beta$  represent the scaling parameters contributed by the shortest path length and depth in the Wordnet, and  $h$  is the length of depth of words in Wordnet hierarchy. Li et al. have reported  $\alpha = 0.2$  and  $\beta = 0.45$  as the optimal parameters when working with Wordnet in particular.

Word similarity is computed from the word information obtained from Wordnet. Words in Wordnet are organized in a hierarchical structure and the similarity can be determined by computing minimum length of path and depth of the given words. Two words are more similar if they have shorter path length. For example in Figure 4.1, it is shown that “car” is more similar to “bicycle” than “bus” because its travel path is shorter. Furthermore, words at the upper level of the hierarchy are perceived as more general semantically and less similar to one another whereas words at the lower layer of the hierarchy are more concrete and more similar to one another. Thus, word depth, which is known as the scaling depth effect, is taken into account.

The lexical semantic vector,  $S_1$ , is then obtained by selecting the largest value in each column (see the third last row in Table 4.1). The reasons for setting the threshold are to eliminate noise, and to make it less vulnerable when working with function words, since there is not preprocessing involved.

The approach also takes into consideration the occurrence frequency of the words. It is rather intuitive that higher frequency words contain less information than those with lower frequencies [34] and information content of a word can be derived from the corpus with the following equation,

$$s_i = s.I(w_i).I(w_i) \quad (4.5)$$

Each element in the semantic vector is weighted by multiplying with  $I(w_i)$  and  $I(w_j)$  (see the second last row in Table 4.1) which is denoted by

$$I(w) = -\frac{\log p(w)}{\log(N+1)} \quad (4.6)$$

where  $N$  is the total number of words in the corpus. The derivation ends with lexical semantic vector,  $\vec{S}_1$ , which is shown in the last column in Table 4.1 The second sentence is also derived in the same way to produce  $\vec{S}_2$ , see Table 4.2. The process yields

$$\vec{S}_1 = [0.39, 0.33, 0.179, 0, 0.074, 0.008]$$

$$\vec{S}_2 = [0.19, 0, 0.16, 0, 0.389, 0.04].$$



Then, the similarity of two sentences is obtained through

$$Sim_{semantic}(S_1, S_2) = \frac{\vec{S}_1 \cdot \vec{S}_2}{\|\vec{S}_1\| \cdot \|\vec{S}_2\|} \quad (4.7)$$

Next, the process to derive the word order vectors is described. The word order vectors are also derived using the joint word set,  $J$ . For that purpose, words in  $J$  are assigned with unique indices according to the order they appear in candidate sentences.  $O_j$  is a word order index container for  $J$ .

$$J = \{\text{RAM keeps things The CPU uses}\}$$

$$O_j = \{1, 2, 3, 4, 5, 6\},$$

To derive the word order vector for  $S_1$ , we assign the order index in  $J$  to the corresponding words in the candidate sentence. The first word “RAM” in  $S_1$  has an index of 1 and “keeps” has 2, etc.  $J$  normally has more words than the candidate sentences. For instance,  $S_1$  does not have the words “The”, “CPU”, and “uses”. For such cases, the index is the most similar word index in  $J$ . For example, according to the Table 4.2, “The” is similar to no other word and thus it has index value of 0, whereas “CPU” and “uses” are most similar to “things” in  $S_1$ , which has index of 3. Thus the word order vector for  $S_1$  is

$$\vec{O}_1 = [1, 2, 3, 0, 3, 3]$$

The minimum similarity threshold also applies here. For any word that have similarity less than the preset threshold, the value is zero.  $\vec{O}_2$  is also derived similarly,

$$\vec{O}_2 = [1, 0, 3, 4, 5, 6].$$

The word order similarity score is then

$$Sim_{order}(S_1, S_2) = 1 - \frac{\|\vec{O}_1 - \vec{O}_2\|}{\|\vec{O}_1 + \vec{O}_2\|} \quad (4.8)$$

The overall sentence similarity is defined as the combination of lexical semantic similarity and word order similarity,

$$Sim_{Li}(S_1, S_2) = \gamma Sim_{semantic} + (1 - \gamma) Sim_{order} \quad (4.9)$$

where  $\gamma$  is a relative contribution of semantic similarity and word order similarity to the overall similarity. According to Li et al,  $\gamma$  is a value greater than 0.5.

### 4.3.3 Mihalcea et al. Similarity Measure

The similarity measurement proposed by Mihalcea et al.[57] also utilizes corpus- and knowledge-based measures of similarity. However, the similarity score is computed with the combination of the following conceptual frameworks: word similarity, and word specificity.

Borrowing similarity metrics from applications such as malapropism detection and word sense disambiguation, Mihcalcea et al. use six different word similarity metrics in their study: Leacock and Chodorow [43], Lesk [46], Wu and Palmer [89], Resnik [70], Lin [48], and Jiang and Conrath [32]. These metrics have originally been created to measure concept likeness, rather than word, but they can be easily adapted to compute word similarity by computing the shortest distance of given words' synsets in the Wordnet hierarchy.

Mihalcea et al.'s similarity measure recognizes entailment. Given two texts,  $T_1$  and  $T_2$ , the entailment is thus defined as the directional semantic similarity of a text segment  $T_1$  with respect to text segment  $T_2$ . Pragmatically, it is a set of words from  $T_1$  with the maximal similarity to word in  $T_2$ , or vice versa. For that reason, word similarity is considered a directional measure. Such definition provides the flexibility to handle applications that require entailment and is easily converted to a bidirectional measure by taking the average of two unidirectional measures.

Word specificity refers to the specific meaning words (e.g. collie and sheepdog) versus generic concept words (e.g. animal and mammal). The similarity measure gives specific meaning to words with higher weight than generic concept words because specific meaning words are more precise and concrete compared to generic concepts words which are abstract and intangible.

Word specificity in the measure is computed using the inverse document frequency (IDF) proposed in [34]. IDF assumes rare words, which occur in fewer documents have greater discriminatory weight than common words, which inherently appear in almost every documents. Mihalcea et al. use the British National Corpus to derive each word IDF.

#### 4.3.4 Measuring Sentence Similarity

Mihalcea et al.’s measure was originally used to measure text similarity. It was reported used to measure similarity at the sentence level in [57]. Given two candidate sentences,  $S_1$  and  $S_2$ , the measurement begins with tokenization and part-of-speech tagging all the words in the sentence into respective word classes (noun, verb, adverb, adjective and cardinal). For each word in the sentence, it is measured against all the words from the other sentence to find the highest semantic similarity ( $maxSim$ ) with the six word-word metrics. The word-word similarity is computed only on the words from the same word class which are either from noun or verb word classes. The reason for this is that noun and verb semantic trees in Wordnet are separated and it is not possible to obtain similarity between nouns and verbs. For word classes that do not have readily available knowledge bases (e.g. adverb, adjective and cardinal) lexical or word matching is used instead. The equation used to compute the similarity of two words is thus

$$Sim_{Mihalcea}(S_1, S_2) = \frac{1}{2} \times \frac{\sum_{(w \in S_1)} maxSim(w, S_2) \times IDF(w)}{\sum_{(w \in S_1)} IDF(w)} + \frac{\sum_{(w \in S_2)} maxSim(w, S_1) \times IDF(w)}{\sum_{(w \in S_2)} IDF(w)} \quad (4.10)$$

#### 4.4 Limitations of the Reviewed Sentence Similarity Measures

We have presented three candidate sentence similarity measures and we are keen to apply them in our study to measure construct relationships. Although these measurements have proven their effectiveness in various applications, they do come with limitations. We discuss each measure’s limitations and then provide workarounds in the upcoming section.

#### 4.4.1 LSA Limitations

LSA may be used for examination of a variety of text units even very small ones such as individual words. However, LSA works through the creation of a semantic space from a large set of paragraph size or larger texts. If such a set of texts is not available, or primarily contains short texts, LSA will not work properly. This shortcoming has been reported extensively in [87, 69, 85]. We cannot build a semantic space with construct items, name and definition, because we mostly have only short sentences averaging 11 words, thus limiting our ability to use LSA.

#### 4.4.2 Limitations in Li et al. and Mihalcea et al. Similarity Measures

Both measurements proposed by Li et al. and Mihalcea et al. utilize Wordnet as the lexical knowledge base. Although the latest version of Wordnet does contain a substantially wide range of common and general words, it does not cover specific domain vocabulary such as those from the IS, education and psychology disciplines. Wordnet is primarily designed to act as an underlying database for different applications, and cannot be used in specific domains that it does not cover. To prove that, we carried out a pilot study on judging the effectiveness of similarity on a number of construct items.

Table 4.3 shows the similarity score of two sentences with Wordnet-based sentence similarity measures and LSA. The Wordnet-based sentence similarity measure works reasonably well on sentences that consist of common words (see first and second rows). The table also highlights that for a few particular cases, the Wordnet-based sentence similarity measure is rendered null for two sentences that are somewhat related (see third, fourth and fifth rows) because the other sentences contain IS specific terms e.g. electronic mail, enterprise resource planning and e-commerce. It can be seen that LSA manages to capture the relationship, though given low similarities. Supposedly these terms are associated with technologies, and should be assigned a certain degree of similarity, but since they are not contained in Wordnet, the sentence similarity measure is not able to recognize the context.

In the coming section, the necessary adaptations to the measures are addressed.

## 4.5 Proposed Similarity Measures for Construct Relationships

### 4.5.1 Construct Similarity Measure derived from Li et al. (ConSimLi)

The result of the pilot study in Table 4.3 has clearly shown that sentence similarity measures built upon the Wordnet-based word-word similarity measure does not suit this study well, as it requires a domain-specific corpus. This shortcoming can be overcome by replacing Wordnet with LSA, which is built on the specific corpus being studied. Instead of computing words' semantic distance in hierarchical structure of Wordnet, the cosine angle of row vector representing words in LSA is computed.

Li et al. incorporated word order vector which is derived from sentence to capture sentence semantic and syntactic information. However, recent findings suggest that the word order vectors do not significantly improve the similarity measure [29]. Thus, in this study, the operator that uses word order vector as part of the similarity computation is discarded. This, in turn, yields the final equation which is:

$$s(w_1, w_2)' = Sim_{LSA}(w_1, w_2) \quad (4.11)$$

### 4.5.2 Construct Similarity Measure derived from Mihalcea et al. (ConSimMi)

For the same reason as with Li et al., the term-term similarity measure was replaced in Mihalcea et al.[57] with LSA which yields the following equation,

$$Sim_{Mihalcea}(S_1, S_2) = \frac{1}{2} \times \frac{\sum_{(w \in S_1)} Sim_{LSA}(w, S_2) \times IDF(w)}{\sum_{(w \in S_1)} IDF(w)} + \frac{\sum_{(w \in S_2)} Sim_{LSA}(w, S_1) \times IDF(w)}{\sum_{(w \in S_2)} IDF(w)} \quad (4.12)$$

## 4.6 Predicting Construct Similarity with the Proposed Approaches

The three proposed similarity measures discussed above are used in construct properties to reveal construct relationship. The relationships can be predicted through the semantic context embedded in construct properties.

Construct relationships can be predicted by comparing the same construct property: name to name, definition to definition, where each property is treated as natural text and a similarity score is produced to indicate its degree of semantic similarity. To transform the score into binary relationships, a cutoff threshold is preset where any score above the threshold renders the relationship as correspondent, otherwise, independent.

Item similarities are computed for all items from two constructs and then the item scores are subsumed to indicate a construct relationship. Two functions are formulated to subsume item similarities into the construct relationship.

The first function is to use the second highest score among all inter-construct item relationships. The function hypothesizes that two constructs relationship can be predicted from the measurement items and when two constructs share two highly similar measurement items, it indicates that they are related. Taking the minimum score of the two reduces false positive which results from high score of just one pair of items that share high similarity.

The second function to represent the construct similarity is by taking the average of the two most similar item scores. This function tends to compensate items pairs that are made up of highly similar and very dissimilar item pairs. It was anticipated that the average score can better represent the construct relationship without the skewness demonstrated by the first function.

To sum up, with three derived similarity measures, LSA, ConSimLi, ConSimMi, on three different construct properties, name, definition, items, where two functions are used on items, results in building 12 computational models.

## 4.7 Evaluation

In the next chapter, the process of gold standard creation, which will be used to benchmark the 12 models is described. The goal is to identify one model which most accurately infers construct relationships based on the properties, as compared with human judgment.

## 4.8 Conclusion

This chapter began with a presentation of a variety of related literature, specifically in the area of text similarity measures, and detecting the semantics and meaning of a sentence. The problem of measuring construct similarity based on the textual properties, which mostly consists of short natural language text, is no different than the text similarity detection. For that reason, three sentence similarity measures that are well studied and well received were selected and discussed in detail. We presented the limitations of these approaches and followed by presenting solutions to the limitations. The discussion ends with a brief explanation of how the adapted approaches are used as an automated, unsupervised solution to predict construct relationships.

Construct item n	Construct item m	LSA	Li	Mihalcea
“I know the features of the technologies.”	“I know the cost of deploying the technologies.”	0.75	0.40	0.44
“I know the features of the technologies.”	“I don’t know the type of business activities in which these technologies have been can be deployed.”	0.60	0.74	0.63
“I know the features of the technologies.”	“I don’t know the type of business activities in which these technologies have been can be deployed.”	0.15	0.00	0.00
“I know the features of the technologies.”	“How knowledgeable are you on using enterprise resource planning.”	0.10	0.00	0.00
“I know the features of the technologies.”	“What is your general knowledge of e commerce.”	0.24	0.00	0.00

Table 4.3: Pilot study: How Li et al. and Mihalcea et al. similarity measures operating on Wordnet are not able to capture domain-specific words.



## Chapter 5

### Building a Gold Standard

#### 5.1 Overview

The main goal of the study is to devise automated approaches appropriate for weaving of related constructs into a network here termed ConstructNet. We conjecture ConstructNet as a metatheory—a theory which is built on existing theories—which can be visualized as a map that explain the interplay of constructs. Instead of finding the construct similarities operationally, we use Natural Language Language (NLP) techniques to reveal construct semantic relationship based on their properties

But how well does ConstructNet reflects the real scenario? We can measure the accuracy of ConstructNet by comparing it with a construct network that has been created by the experts. This human-crafted network is known as a gold standard (or ground truth or hypothesis). The outcome of each of the computational approaches is compared against the gold standard with evaluation metrics to determine which computational approaches are able to produce a ConstructNet close to the gold standard.

#### 5.2 Objective

For this study, we require a gold standard where the relationships of constructs are annotated either as correspondent or independent. One way of doing this was to place the correspondent constructs in the same cluster and the independent constructs in different clusters. With that, the construct relationship can be deduced automatically from these clusters: correspondent if they are

from the same cluster, or independent if they come from different clusters.

### 5.3 Related Works

Evaluation and error analysis with gold standard are not new in NLP. Works [64, 27] in computational linguistic have suggested the standards to design a gold standard. One such standard is performed through double blind annotation followed by the adjudication of disagreement. The blind annotation is proposed to eliminate errors or biases that are introduced by a single annotator. To improve the quality of annotation, the approach normally calls for more than one annotator to annotate the same instance independently a number of times. If there is disagreement, the annotators adjudicate among themselves to reach a consensus so that the gold standard produced is free of bias and error. Creating a gold standard is an expensive process. In order to alleviate the tedious process, many of the annotation tasks in NLP make use of readily available knowledge bases such as dictionaries and thesauri, as well as specific resources like Wordnet [56], FrameNet [5] and The Proposition Bank [65].

### 5.4 Our Challenges

We are working with 1054 constructs. With this number of constructs, creating a gold standard through blind annotation is far from possible. Note that this unprecedented task does not have a reference or a knowledge base to begin with. If we were performing blind annotation of every pair of constructs, taking into account the possibility of clusters that can be created and the number of constructs a cluster can contain, we are facing with  $n!$  ( $n=1054!$ ) possibilities. Therefore, the blind annotation does not seem fit as it requires enormous resources in terms of cost, time and effort.

### 5.5 Our Approach

In this section, we discussed our unique approach for creating a gold standard. The process of creating the gold standard was separated into seven steps:

- Step 1: Defining construct relationships;
- Step 2: Annotator training;
- Step 3: Colocating identical constructs;
- Step 4: Rough categorization;
- Step 5: Refined categorization;
- Step 6: Reconciling changes; and
- Step 7: Evaluating categories.

### **Step 1: Defining Construct Relationships**

The first step to the gold standard creation is to learn how experts, who have prominent experience in theory development measure the commonality of given constructs. Since the study predicted two types of construct relationships: correspondent and independent relationships, the experts are interviewed about the correct steps to judge the construct relationships based on the construct properties. The experts feedback was collected and generalized as:

A construct,  $C'$ , is defined to be correspondent to another construct,  $C$ , if some construct measurement items for  $C'$  could also be used to measure the latent construct measured by  $C$ . Operationally, a construct  $C'$  will be judged as correspondent to another construct  $C$  if the domain experts determine that two or more potential construct measurement items for  $C'$  could also be used to measure the latent construct measured by  $C$ .

The basis for such determination might include the similarity between construct measurement items, definitions, names, citations, unit of analysis, and other evidences for the two constructs.

### **Step 2: Annotator Training**

Once the expert feedback was collected and generalized, it was used to develop a training document (see Appendix A). The training document contained specific examples of construct relationship that were collected from the construct collection. They are used to help annotators

make correct judgements. The training document was distributed to all the annotators involved in the categorization process. The document was an important guideline and had to be read and understood before participation in the categorization process. After all annotators had read the document, it was discussed with emphasis on its examples.

The annotators involved in the categorization process were project investigator (PI)<sup>1</sup>, one chief research assistant, one senior research assistant (RA)<sup>2</sup>, five experienced RAs, and three PhD students<sup>3</sup>. All of the annotators were carefully selected by the PI as it was important to include only experienced researchers to ensure the high quality of the gold standard. The PI has more than 15 years of research experience in the IS discipline with multiple literature review and categorization projects completed and the chief, senior and experienced RAs had on average half to two years of research experience in extracting constructs from articles. All the RAs hired undergo extensive training and familiar with construct extraction.

The annotators worked as teams. Each team that consisted of two or three annotators from a mixed group of research assistants and PhD students designed to always pair the least RAs with a Ph.D. student. In addition, another special team, known as the adjudicator team, was made up of the PI and one RA.

### **Step 3: Colocating Identical Constructs**

To slightly simplify the categorization tasks all identical constructs (defined as those constructs with identical names and at least two identical construct items) were identified. The rough categorization was carried out by the PI alone though use of a special-purpose application allowing the detection of identical constructs.

---

<sup>1</sup> Kai Larsen.

<sup>2</sup> Heather Witte and Leslie Grush.

<sup>3</sup> Chih How Bong (Computer Science), JingJing Li (Business), and Jeffrey Ryan Sweeney (Business).

#### Step 4: Rough Categorization

Because there are 1054 constructs, finding pairwise relationships for slightly over a thousand constructs is indeed a complex problem and can be tedious. The possibility of relationships generated from this number is  $\frac{1}{2} \times 1054 \times 1053 = 650,000$ . With over half of a million relationships, it is almost impossible to annotate every construct pair. Fortunately, the study focuses on discovering correspondent constructs, it is easier to begin the categorization process by grouping highly related constructs instead of choosing any random construct pair. The rough categorization is carried out by whole team. Overall, the categorization task was divided into two stages: rough categorization and refined categorization with repeated iterations through the refined categorization. The purpose of rough categorization is to group constructs that have highly similar properties into the same cluster to simplify the task enough to make it cognitively tractable.

#### Step 5: Refined Categorization

The purpose of refined categorization is to turn the clusters generated through the rough categorization into subcategories. The outcome is a hierarchical taxonomy of constructs, which is also a gold standard to evaluate the proposed computational approaches.

The refined categorization process begins by assigning the hierarchical clusters labels. This is done by putting a blue triangle paper on top of the stack of construct groups from the same cluster. The taxonomy in this study to is restricted to two hierarchical levels. For example, main category “Trust” has subcategories like “Trust in Benevolence”, “Credibility” and “Trust”. The subcategories are known as correspondence in the thesis. For this exercise, the individual constructs were printed on a white rectangle of paper, and constructs were stacked together under a green rectangle sheet if they were believed to fit the definition of correspondence.

For example, Figure 5.1 shows the “Group” and “Academic” main categories (blue triangle), some correspondences (green rectangle) and constructs (white rectangle).

The labels for main categories are usually the general area descriptions: organization, com-

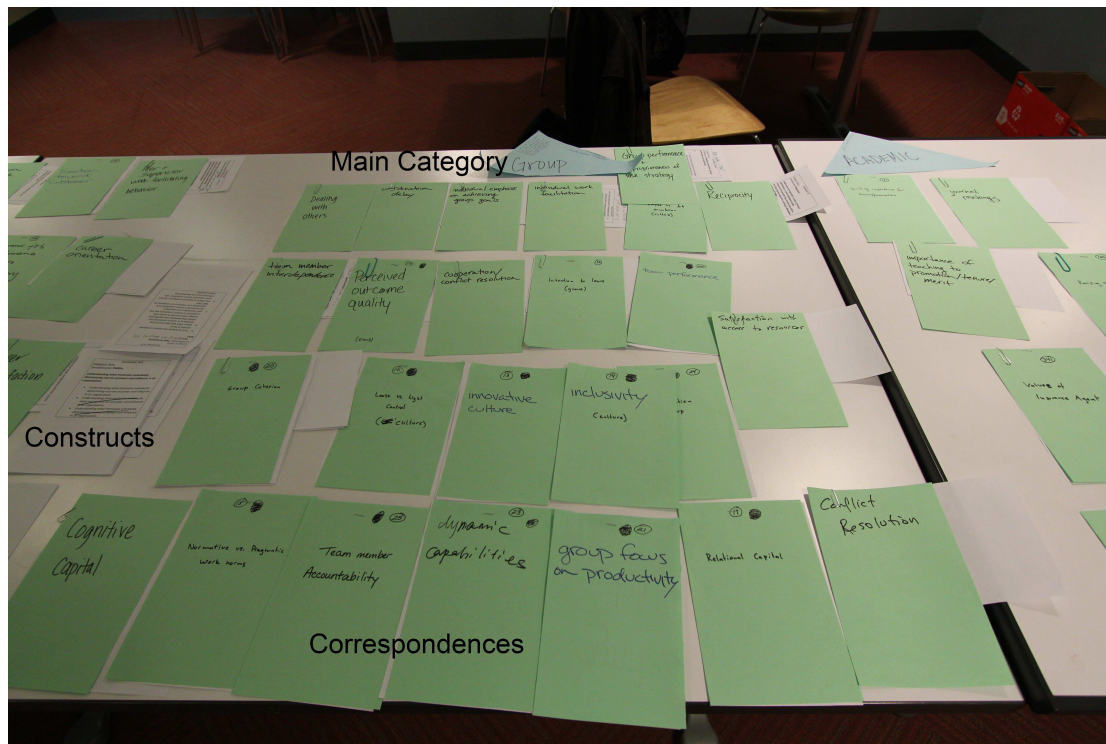


Figure 5.1: Main categories, correspondences and constructs.

munication, task, leadership, etc. The purpose of labeling is an attempt to group the constructs pertaining to the same field of study under one main category. The grouping eases the classification process by breaking up about a thousand constructs into a number of main category and it is especially helpful in providing a cue for annotators, allowing an ambiguous construct to be moved into the general area it belongs, if not to the correct construct group. With that, the ambiguities can be first resolved by identifying the area of study, based on the construct description. This normally is done by judging the properties of constructs such as name, definition and items. For some rare cases, the adjudicator was called in to resolve the ambiguities.

Once every cluster is labeled and every construct is clustered, annotator-teams take turn in examining each construct in the clusters. Examination involves cross-checking whether constructs have similar contexts and whether all constructs in a group truly fit the definition of correspondence. Each team is required to examine the clusters at least one time. Some clusters—usually those with a high number of constructs—require several examinations by different teams. The members of

the team cross-examine the constructs to determine if they can be categorized together based on their definition, items and name with primary focus on items because these have been rigorously developed and statistically tested. Once a team has completed a correspondence, the team can move on to work on a new cluster or main category, which has been already reviewed by other teams. However, there are cases when the properties do not reveal any clue. For cases like these, the sources (i.e. articles) of the construct are normally re-examined in order to better grasp its context.

The reviewing process is designed to examine the work carried out by other teams. The process can be carried out in the presence or absence of the teams who have previously created or reviewed them. Each team takes the opportunity to review every main category and correspondence and the process continues for a number of rounds before being finalized.



Figure 5.2: Reconciliation in progress.

### **Step 6: Reconciling Changes**

The annotators who review the clusters or main categories were allowed to create new correspondences to group similar constructs. They were allowed to rename the existing correspondences. They were also permitted to split a correspondence or merge correspondences if it seems appropriate. They were also free to relocate the construct from one correspondence to other, and this occurred frequently after annotators had seen more correspondences. In fact, experience from the process showed that a construct group could easily belong to multiple categories. The challenge, then, was to make sure multiple correspondent construct groups do not exist in separate categories.

Any change made to the correspondences has to be reviewed and consented to by other annotators, followed by the adjudicator. Sometimes, reconciliation has to be done several times before reaching a decision. Also, the correspondences can be reviewed multiple times by the same annotators. Changes can only be accepted with major votes. If consent cannot be reached among the annotators, the adjudicator is normally brought in to resolve the independent. In fact, for any change to be made, it is compulsory to obtain approval from the adjudicator.

Both the rough and refined categorization tasks were carried out in stages across six days and consumed about 200 person-hours. The categorization task resulted in two hierarchies taxonomy of category and correspondences, and all constructs were grouped and resided in the correspondences. The gold standard has 28 main categories and 343 correspondences and the number of constructs in each construct group was between one and 69. Figure 5.3 shows examples of main categories and correspondence found in the gold standard. IS Development: Risk Factor and IS Development: Etc are the smaller categories which are broken down from parent category, IS Development.

When carrying out the annotation process, in order to reduce the complexity of the problem, only one construct was assigned to one correspondence. In the case where there was a possibility that a construct can belong to a number of correspondences, the most likely category was selected. However, for the purposes of this dissertation, it bears repeating that the hierarchical structure represented only temporal cognitive scaffolding.



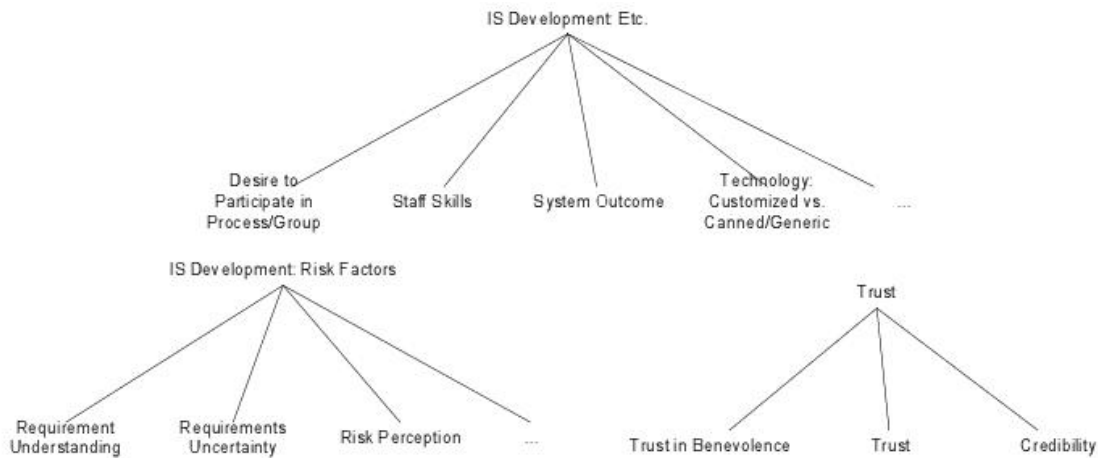


Figure 5.3: Example of hierarchical taxonomy in the gold standard. IS Development: Risk Factor and IS Development: Etc are the smaller categories which are broken down from IS Development.

### Step 7: Evaluating Categories

The purpose of validating a gold standard is to ensure that the gold standard has good inter-rater agreement between annotators and experts. Statistical measures are used to compute the agreement reflecting the quality of the gold standard. A high quality gold standard should have high inter-rater agreement between experts. In this exercise, two IS experts were appointed from different institutes who have proficient knowledge on most of the constructs being worked with. Both experts were tenured professors, had more than a decade of experience with IS research, and had extensive experience with review work, and one expert had received a best-paper award for a review article in a top IS journal. To compute the inter-agreement, each expert is expected to classify the relationships of the same set of constructs. Instead of classifying every construct in the gold standard, which is too laborious, the task was simplified by requesting that the experts classify 300 relationships of construct pairs that were semi-randomly selected from the gold standard. Since the gold standard is made up of highly imbalanced data—less than 2 percent correspondent construct pairs and 98 percent independent construct pairs—an Independent Drawing mechanism was devised to randomly select construct pairs. The advantage of the mechanism is that it is able

to select a balanced sample which consists of equal proportions of correspondent and independent construct pairs, and is therefore not influenced by the highly skewed distribution of independent construct pairs. Algorithm 1 shows the pseudocode for the *Independent Drawing* mechanism.

---

**Algorithm 1:** Pseudocode for *Independent Drawing*

---

**input:**  $\alpha$ , probability likelihood

**input:**  $n$ , desired sample size

**input:**  $P$ , construct pairs

**output:** *samples*

*samples*  $\leftarrow \emptyset$

**while** *samples*  $\leq n$  **do**

*prob*  $\leftarrow \text{random}(0, 1)$

**if** *prob*  $\geq \alpha$  **then**

randomly select a correspondent construct pair

add the correspondent construct pair to *samples*

remove the correspondent construct pair from  $P$

**else**

randomly select a independent construct pair

add the independent construct pair to *samples*

remove the independent construct pair from  $P$

**return** *samples*

---

The mechanism uses a preset probability likelihood,  $\alpha$ . When  $\alpha = 1.0$ , the selection is made up of entirely correspondent constructs; when  $\alpha = 0$  the selection ends up entirely with independent constructs. The selection can be better illustrated with the following example: if we intend to randomly draw 300 samples having equal proportion of correspondent and independent constructs, we can set  $\alpha = 0.5$ . The sample drawing begins with the system randomly rolling the “dice”, which is a random number generator between 1 and 0. If the dice first rolls a 0.7 and since it is higher than  $\alpha$ , the construct pair to be chosen is a correspondent construct pair. The

selection is completely random and non-replacing. Next, if the dice rolls a 0.2, and since it is less than  $\alpha$ , a independent construct pair is randomly picked and included in the samples. The process is repeated until it contains enough sample pairs.

The decisions made by the experts are stored in the database and are used later to compute agreement with the gold standard. For this purpose, web-based application was created to collect expert annotations. The web-based application is depicted in Figure 5.4.

Variable 1	Variable 2
<b>Intrinsic Motivation (R,S) (18)</b> Definition: Deriving satisfaction that lies in the content of the activity itself.	<b>Attitude toward using technology (105)</b> Definition: An individual's overall affective reaction to using a system
<ul style="list-style-type: none"> <li>• I enjoy learning business and technical knowledge about (Purchasing) module.</li> <li>• The more difficult it is to understand business and technical knowledge about the (Purchasing) module, the more I enjoy learning it. (1)</li> <li>• I enjoy learning business and technical knowledge about the (Purchasing) module that are completely new to me.</li> <li>• I have to feel that I'm personally benefitting from learning business and technical knowledge about the (Purchasing) module. (1,2)</li> <li>• I want to find out how good I really can be at learning business and technical knowledge about the (Purchasing) module.</li> <li>• I'm more comfortable when I can set my own goals for learning business and technical knowledge about the (Purchasing) module. (2)</li> </ul>	<ul style="list-style-type: none"> <li>• Using the system is a bad/good idea.</li> <li>• The system makes work more interesting.</li> <li>• Working with the system is fun.</li> <li>• I like working with the system.</li> </ul>

[Hide](#)

Figure 5.4: Screen shot of the web-based application used to collect expert annotations.

For the example above, although  $\alpha = 0.5$ —a fair chance of drawing both types of construct pairs—an equal number of construct pairs is not always the result. Instead, the 300 construct pairs may make up either more correspondent pairs or fewer. As already discussed above, the core of selection is primarily decided by the random dice during the randomization of the samples.

## 5.6 Gold Standard Validity

Cohen's Kappa coefficient [12],  $\kappa$ , was used to compute inter-agreement. Once the experts had independently completed assigning relationships to the same 300 construct pairs, agreement was measured against the gold standard. In addition,  $\kappa$  was computed between experts to learn the

relative agreement between them. To allow meaningful interpretation, guidelines were employed from Landis and Koch [41] who have characterized  $\kappa$  less than 0 as indicating no agreement and 0 to 0.20 as slight agreement, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.811 as almost perfect agreement. Table 5.1 gives the results.

	Inter-agreement $\kappa$
Expert 1 vs Gold Standard	0.77
Expert 2 vs Gold Standard	0.68
Expert 1 vs Expert 2	0.66

Table 5.1: The degree of agreement between the experts and the gold standard.

Based on the guidelines given above, the results show that both experts have substantial agreement with the gold standard, 0.77 and 0.68 respectively. The results also show that the two experts yield substantial agreement between them, yet lower agreement than either expert had with the gold standard. Hence, the gold standard that was created is considered to be relatively accurate and appropriate for the task at hand. In the next chapters, the gold standard will be used to evaluate the robustness of the proposed computational approaches in finding construct relationships.

## 5.7 Conclusion

Exploring and labeling over a thousand construct relationships can be tedious and time consuming. As was explained in this chapter, an relatively standard approach was adopted to rapidly create a gold standard without involving enormous resources. First, a computational approach was used to find identical constructs. Once the constructs are clustered, they were reviewed by annotators and sub-categorized into more refined categories. This yielded the gold standard of construct relationships. To validate the gold standard, 300 construct pairs were semi-randomly drawn and provided to experts for labeling. The Kappa coefficient showed that there is substantial agreement between experts and the gold standard. In the following chapter, the gold standard will be used to benchmark the proposed computational approaches.

## **Chapter 6**

### **Evaluation**

#### **6.1 Overview**

This chapter describes a systematic methodology to evaluate the similarity measures used to predict construct relationships. It begins by finding the optimal dimensionality to perform Singular Vector Decomposition (SVD) for (Latent Semantic Analysis) LSA. Then, benchmarks all the 12 computational models with Receiver Operating Characteristic (ROC) plots. It then proceeds to select only the best generalized function based on construct items. This reduces the number of models to be investigated into nine. By comparing the nine models, it is intended to settle for a cutoff score that is able to produce a cutoff sample that has high accuracy. The chapter concludes by presenting the outcome of combined measures.

#### **6.2 LSA Dimensionality**

It is necessary to find an optimal dimensionality to perform Singular Value Decomposition (SVD) because some of the computational models are generated with Latent Semantic Analysis (LSA).

This is because the performance of the LSA is heavily dependent on the different dimensionality used during the SVD process. According to [19], lower dimensionality allows broader comparison of semantic concepts while a higher dimensionality allows more specific comparisons of concepts. Although higher dimensionality promise a more accurate result to a point, they require more computational resources when computing similarity. It is important to seek a dimensionality

that produces an appropriate compromise between accuracy and performance.

Past literature has reported that 300 dimensions will usually give the best results with tens of thousands of documents. However, recent studies [8, 38, 60] have reported that the range of 50-1000 dimensions are suitable, depending on the length of documents. Also, the dimensionality is sometimes limited by the number of documents in the collection.

To obtain the optimal LSA dimensionality, five semantic spaces were built with the dimensionality between 200-500. The gold standard was then used to evaluate how accurately each model retrieving relevant measurement items from the corresponding constructs. R-Precision is reported [50] for each model, and the dimensionality from the model which has the highest score is settled on. R-precision is defined as the precision at R-th position in the ranking of results for a query that has R relevant documents. This measure is highly correlated to average precision.

The semantic spaces are created with article paragraphs as a unit of document. All models use the exact settings except the number of dimensionality when generating the semantic spaces. For each model, every construct measurement item in the database was projected into the semantic spaces. The result of the projection is the item's specific semantic locations in the semantic space. The locations, which are represented as vectors, are then stored in the semantic subspace. This subspace is named meta-semantic space, and it contains the semantic location of each item. It is presumed that the identical items should be clustered together in the semantic spaces even though they are from different constructs.

Each unit of evaluation is performed by projecting a measurement item text into the main semantic space and then using the projected location vector stored in the meta-semantic space to find the relevant items by computing the cosine angle. An accurate model is the one which is able to return more relevant measurement items at the highest similarity score.

Table 6.1 summarizes the durations of each evaluation process and the R-Precision scores for each model. From the table, it is obvious that the high dimensionality models take more computational resources. It is also clear that higher number of dimensionality does not actually improve R-Precision, although the R-Precision score at dimensionality of 400 seems a little better.

Dimensionality	Duration(mins)	R-Precision
200	33.17	0.35
300	41.49	0.35
400	50.53	0.36
500	58.11	0.35
600	67.58	0.35

Table 6.1: Comparison of different dimensionality model in retrieving relevant measurement items.

This implies that different numbers of dimensionality do not have significant influence on the models.

For the upcoming experiments, the LSA dimensionality of 300 is set.

### 6.3 Comparison of Construct Similarity Measures

This section is comparing how well the three proposed similarity measures, LSA, ConSimLi and ConSimMi, perform in predicting construct relationships based on their name(N), definition(D), and the two generalized functions on construct items: minimum score of the two most similar items (Min) and average score of two the two most similar items (Avg). For ease of reference in the coming section, ConSimLi(N) is used to refer to the computation model that is built with ConSimLi measure on construct names.

The prediction task is to label whether the construct relationship is either correspondent or independent, by judging the semantic context embedded in the textual properties. A baseline is included in the comparisons. The baseline here is basically a simple algorithm that randomly assigns the construct relationships to either correspondent or independent relationship.

The performance of similarity measures using a very popular graphical data mining evaluation scheme known as Receiver Operating Characteristic (ROC) is reported. The performance of each measure in ROC curves is represented by plotting the true positive rates (TPR) on Y axis against the false positive rates (FPR) on X axis. TPR and FPR are defined as

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (6.1)$$

where TP is the true positive, FP is the false positive, TN is the true negative and FN is the false negative.

The values of both TPR and FPR are between 0 and 1. The ROC normally shows as a diagonal line from lower left to upper right. The diagonal line indicates random guess and a good classifier should have points above the line.

Traditionally, ROC is used to gauge the performance of binary classifiers which can be determined by the Area Under ROC (AUC). For a perfect and error free classifier, the increment of TPR does not entail the increment of FPR and the AUC is exactly one (the line goes up straight along the Y axis). However, in most computational classification tasks, when the sample size is



increased, the number of FP increases as well. In other words, the classifier miss-labels negative instances as positive. In general, a good curve is always the one that crawls to the upper left hand of the corner (0.0,1.0). This indicates that the ratio of TPR to FPR is high. On the other hand, researchers are less happy when the ROC curve follows a diagonal path from the lower left hand corner to the upper right hand corner. This means that every improvement in false positive rate is matched by a corresponding decline in the true positive rate.

One can quantify how quickly the ROC curve rises to the upper left hand corner by measuring the area under the curve. The larger the area, the better the classifier. To sum up, the ROC figures show a number of important characteristics:

- (1) ROC plot graphically depicts the compromise between TPR and FPR.
- (2) ROC curves start at (0.0,0.0) and end at (1.0,1.0).
- (3) A perfect classification corresponded to the point at (0.0,1.0). The ROC is then a vertical line with TPR of 1, which also has AUC of 1.
- (4) The plot normally shows a diagonal line from lower left to upper right indicating that for every TP, it is just as likely to encounter a FP. The diagonal line is also known as random guess.
- (5) A good classifier should have points far away from the diagonal line. Points below the diagonal line represents poor classification results.

The ROC plots are presented for the three different similarity measures in Figure 6.1–6.3. Each ROC plot includes the measure’s performance using different construct properties.

The ROCs were plotted based on the information collected at 21 cutoff thresholds: ranging from 0.0 to 1.0 with an increment of 0.05 at each iteration. At each cutoff score, any construct pair with a similarity score more than the cutoff is presumed to be correspondent, whereas the remaining are presumed to be independent. The value of the TP and FP were computed by comparing them to the gold standard. For ease of comparison, the ROC AUC for all models are also presented in

Table 6.2. The AUCs are computed using a trapezoidal approximations of the curves. The ROC for random guess is also plotted.

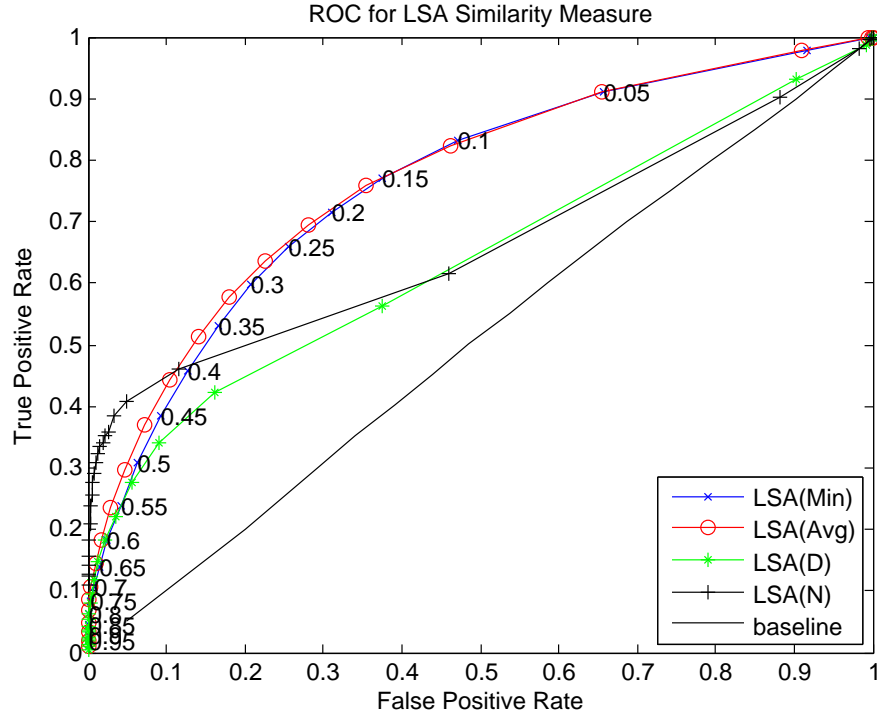


Figure 6.1: ROC with LSA in terms of cosine cutoffs between 0 and 1.

	<i>Baseline</i>	LSA	ConSim-Li	ConSim-Mi
Min score	<i>0.498</i>	0.763	<b>0.779</b>	0.612
Averaged score	<i>0.500</i>	0.770	<b>0.789</b>	0.623
Definition	<i>0.500</i>	0.642	0.634	<b>0.691</b>
Name	<i>0.493</i>	0.663	<b>0.691</b>	0.680

Table 6.2: Area Under Curve for 12 models.

The ROC for LSA is depicted in Figure 6.1. In the figure, it is obvious that the models based on the both construct items work better compared to the models based on construct name LSA(N) and definition LSA(D). When judging which generalized function based on items (LSA(Min) or LSA(Avg)) is better at predicting construct relationship, according to the plot, their differences are not significant as both curves are very close to each other. When compared to AUC in Table

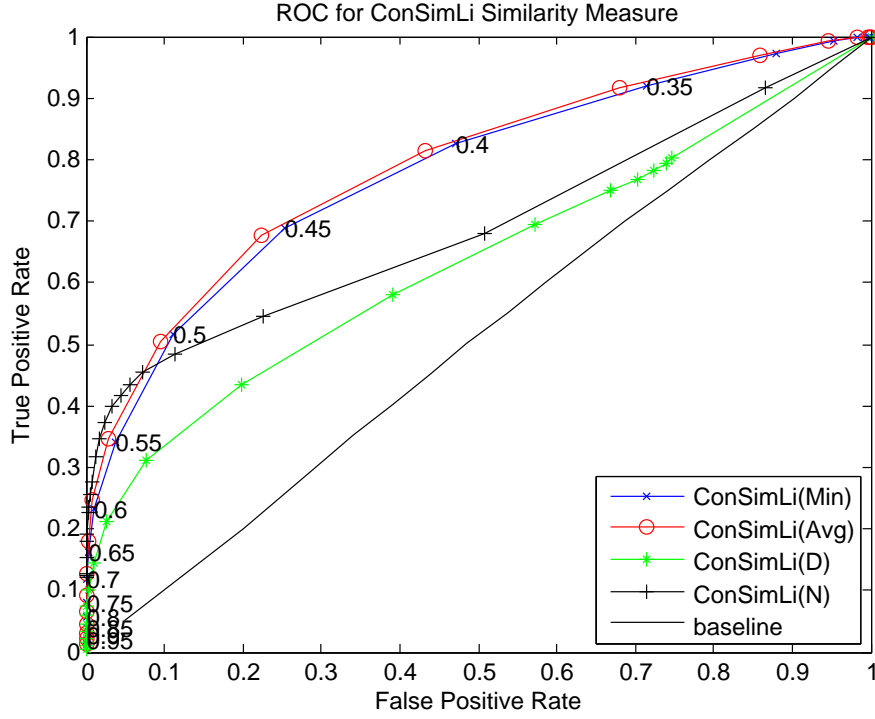


Figure 6.2: ROC with ConSimMi in terms of cosine cutoffs between 0 and 1.

6.2, they each have an AUC of 0.763 and 0.770 respectively, it is clear that LSA(Avg) edges out LSA(Min).

The curves are then followed by LSA(N). The figure shows that LSA(N) is able to pick up the corresponding construct at the early higher cutoff score, but become flattened quickly as the cutoff score decreases, simply implying that LSA(N) performs better than the other when high TPR is required.

The curve is also labelled (only on curves for averaged item model) with the cutoff scores which are used to compute the TPR and FPR. Its purpose to illustrate which cutoff scores produce the points close to (0.0, 1.0).

The ROC plot for ConSimLi, which is shown in Figure 6.2, displays a similar pattern to the LSA's ROC plot. The similarity measure based on the construct items are the best of all. It should be noted that ConSimLi(Avg) also performs marginally better than the ConSimLi(Min). They

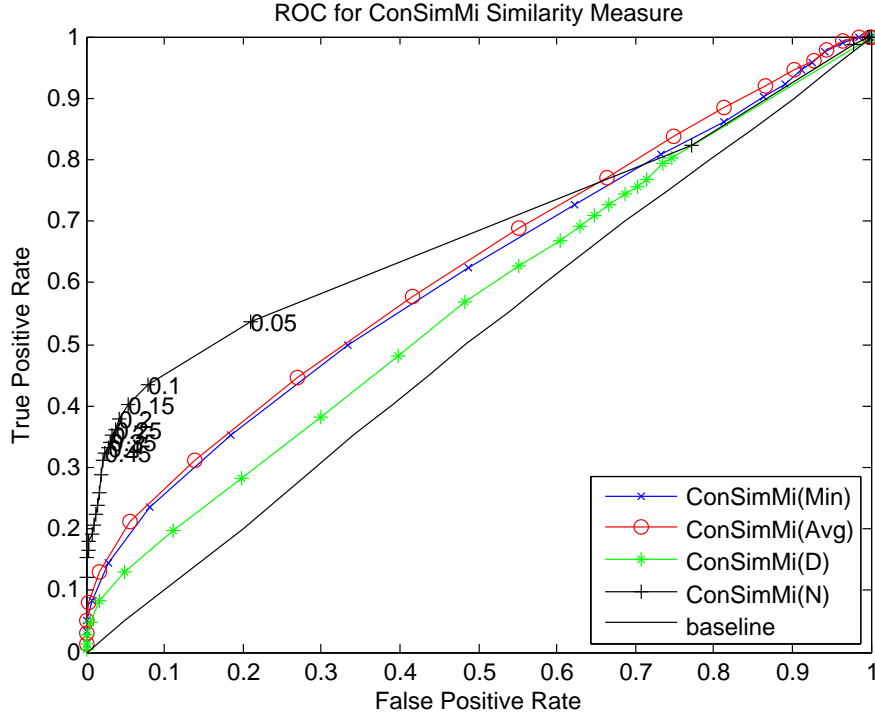


Figure 6.3: ROC with ConSimMi in terms of cosine cutoffs between 0 and 1.

are followed by the ConSimLi(N), and finally the ConSimLi(D). The four models also performed significantly better than the baseline. Differently from the LSA, the ConSimLi(N) flatten out at a higher FPR than LSA(N), which indicates that it is marginally better than the LSA. Overall, when comparing AUC of every model with LSA, the similarity measures based on ConSimLi competitively have an edge.

For ConSimMi, it was a surprise to see that the best model is achieved with ConSimMi(N). It has the largest AUC among three measure based on the construct name (see Table 6.2). However, unlike the two similarity measures just seen, ConSimMi(Min), ConSimMi(Avg) and ConSimMi(D) perform poorly here and their curves are close to the diagonal line. That implies that the ConSimMi does not perform as well as LSA and ConSimLi

In summary, LSA and ConSimLi work well with the construct items. The averaged score model performs better than the minimum score model, but the improvement is not significant. The

next best performer is the model based on construct name. All proposed measures with name are able to pick up a number of correspondent constructs at high TPR to FPR ratio, but it flattens out quickly. This is because there are a high number of similar name constructs in the database, which are also correspondent. The curves flatten out at some points from which it may be inferred that once the cutoff scores have passed the exact match score mark (similar name constructs have similarity score of 1.0), similarity based on name becomes less meaningful (harder to predict), and the number of FP is introduced at a much faster rate. Hence, we can conclude that the construct names and definition are not good predictors for construct relationships when compared to the construct items.

For the upcoming experiments, only averaged score, name and definition model of the three measures will be included for comparisons. Hence, this leaves in total nine models for comparison.

#### 6.4 Similarity Scores Distribution

The ROC curves assess the TPR against FPR and the overall performance of each model predicting construct relationships by computing their AUC. However, they do not clearly inform the accuracy of finding correspondent constructs at each cutoff sample. For example, in Figure 6.2, it is shown that 0.45 is the closest cutoff score located to the (0.0,1.0), but it does not tell the actual percentage of true correspondent pairs at this threshold. If an expert were asked to classify the construct relationships, where the cutoff sample is obtained, the expert should be fed with more correspondent construct pairs that are presumed to be similar in order to keep the expert constantly motivated. If the expert sees many independent pairs, the expert might terminate the task with a perception that upcoming construct pairs are likely to be independent. So, it is important to determine a cutoff score that results in more correspondent pairs than independent pairs.

Since the similarity measures use scores to represent construct similarities, it is useful to know which cutoff score of each model contributes the best accuracy. We recognize the fact that in real life it is not possible to correctly predict all correspondent construct, we strike a balance to settle for cutoff score which is able to produce the most correspondent constructs with the least

error (more true positive, less false positive). In other words, we want to find a cutoff score that maximize TP to FP ratio. With the gold standard, we evaluate each cutoff sample with accuracy — the percentage of correspondent construct pairs in the cutoff sample where the sample and the gold standard agree. Accuracy is formulated as

$$Accuracy_c = \frac{n_c}{N_c} \quad (6.2)$$

where  $n_c$  is the number of correspondent construct that the sample and the gold standard agree at the specified cutoff,  $c$ .  $N_c$  is cutoff sample size at  $c$ . Similarly to the previous section, every possible construct pair is computed using the similarity measures. Cutoff sample is obtained by selecting construct pairs that have similarity scores equal or greater than the specified cutoff score.

#### 6.4.1 Cumulative Frequency Distributions of Correspondents Constructs

Figure 6.4–6.6 present the cumulative frequency distributions of correspondent constructs at 21 different cutoff scores. The cumulative distributions are represented by histograms. The X-axis shows various intervals of scores (the interval labeled 0.50 includes any score from 0.500 to 0.549). Although the measures produce scores between -1 to 1, only distribution between cutoff scores 0 to 1 is shown because that includes almost all the correspondent construct pairs. The Y-axis of to the right shows the number of correspondent constructs in the interval or below the interval.

The accuracy of correspondent constructs obtained at different cutoff scores is plotted and overlayed over top of the histogram. The measures' accuracy is presented by line plots. The Y-axis to the left is the scale for the accuracy.

In the figures the accuracy of all models plummet as the cutoff scores move down. The decreasing accuracy is due to the increment of FP when increasing the cutoff samples. The accuracies of all models with ConSimMi (see 6.6 ) drop at a much faster rate, even at higher cutoff than LSA and ConSimLi. That implies that the model based ConSimMi does is not able to achieve high accuracy when compared to the other two similarity measures.

Comparing LSA to ConSimLi, it is obvious that both have similar trends of accuracy and

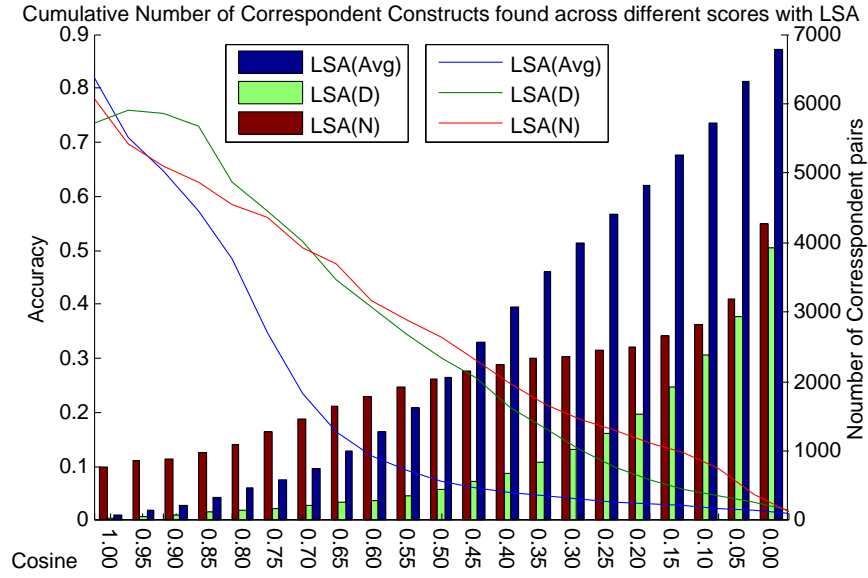


Figure 6.4: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff

the model based on the construct definition is more favorable for LSA while the model based on construct items is favorable for ConSimLi. Overall, models built with ConSimLi have a slight advantage over its rival as the accuracy rates drop at slower rates. This also complies with the previous finding in Section 6.3, which concludes that ConSimLi(Avg) is the the best performer among all.

Another fact worth noting is the similarity scores produced by each measure are not aligned with each other. For example, both ConSimLi(Avg) and ConSimMi(Avg) (see blue color histograms in Figure 6.5 and Figure 6.6) are able to find all the correspondent constructs at higher cutoff score compared to LSA (see Figure 6.4). ConSimLi and ConSimMi are able to find all the 6,944 correspondent constructs before they hit cutoff score 0, and LSA has to go lower in order to get all the correspondent constructs. Likewise, similarity based on name and definition require much lower cutoff scores in order to cover all the correspondent pairs.

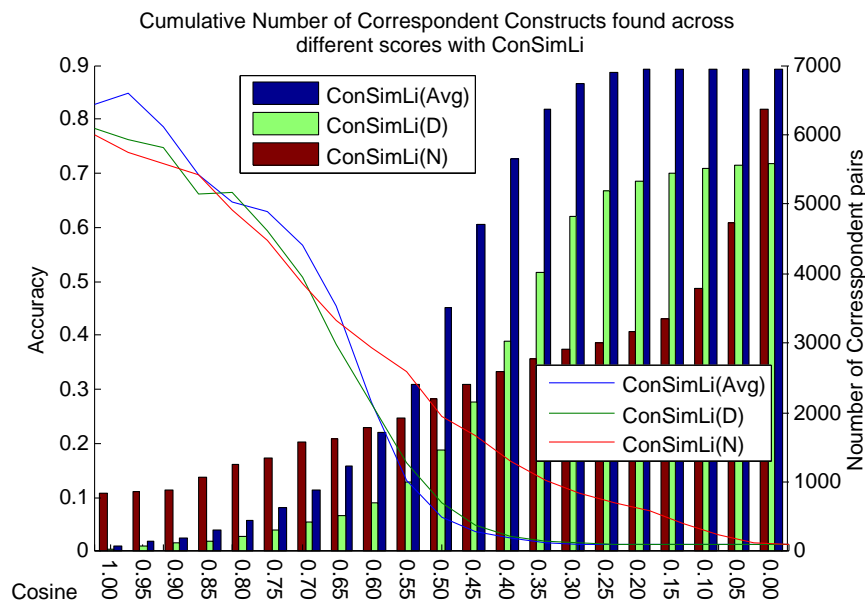


Figure 6.5: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff

#### 6.4.2 Frequency Distributions of New Correspondents Constructs

The following are the frequency distributions of new correspondent constructs found at 21 different cutoffs. The accuracy of the cutoff sample is plotted and overlayed over the distributions.

Most of the models' similarity scores are in normal distribution, see Figure 6.7-6.9, except LSA(D), LSA(N), ConSimLi(N) and ConSimMi(N).

Moreover, the distribution histograms for all name models show concave shapes, indicating a high number of true correspondent constructs at high and low cutoff scores. The high frequency to the left is because the database contains a high number of similar name constructs which are correspondent. Most of the similar name constructs have similarity scores of 1.0 and are easily picked up at first cutoff, 1.0. That implies that similarity based on name is critical if dealing with a database that has many similar name constructs. .

However, the higher frequency on the low end indicates that those construct relationships



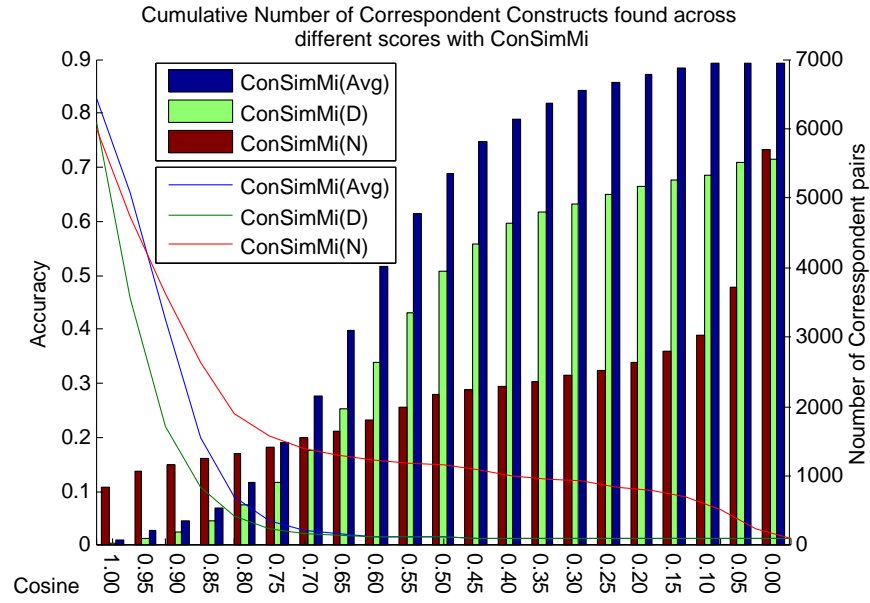


Figure 6.6: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff.

are deemed to be independent by the similarity measures. That the name models do not follow normal distribution is a strong indicator that construct names are not a good predictor, regardless of which similarity measures are used. However, they can be useful if dealing with a high number of similar name constructs.

The figures also show that the mean value of similarity scores in the frequency distributions varies from one measurement to another. To ensure optimal accuracy, the different cutoffs for each model need to be found. Based on the experiment, it is obvious that the ConSimLi based on construct item is favorable at cutoff score, 0.8.

## 6.5 Combined Measurement

This section is investigating whether combining any of the nine models improves the chances of discovering more correspondent constructs, accompanied with a minimum error rate. The previous

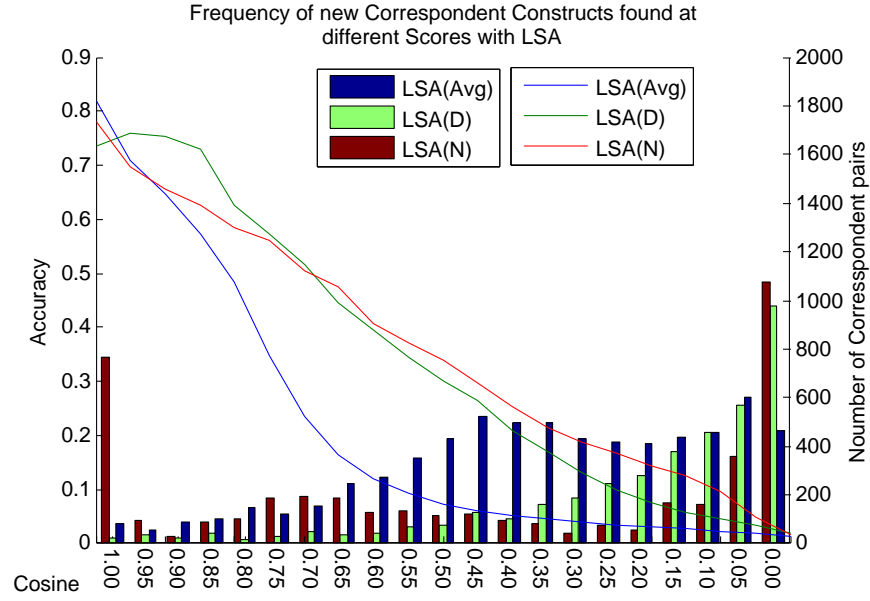


Figure 6.7: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlaid with precision (success rate) at each cutoff.

experiments concluded that the ConSimLi(Avg) is the best among the nine models, so in the following experiment it will be used it as a baseline with combined measures.

For the sake of brevity, new notation for each model is introduced. LSA, ConSimLi and ConSimMi continue to be used to refer to the similarity measures, which take I (item, averaged function), D (definition) and N (name). For example, the notation ConSimLi (N+D) refers to combined model of ConSimLi on construct name and definition.

A linear combination of the scores from the candidate models is performed to construct combined measures models. Each candidate model is given an equal weight and the combined model score is the average of the score that is added up from each candidate model. The combined model is made up of either two or three candidate models with different measures or different construct properties. This gives a total of 129 combined models (including the nine stand alone models). For the sake of brevity, only the results of combined models of all properties of the three

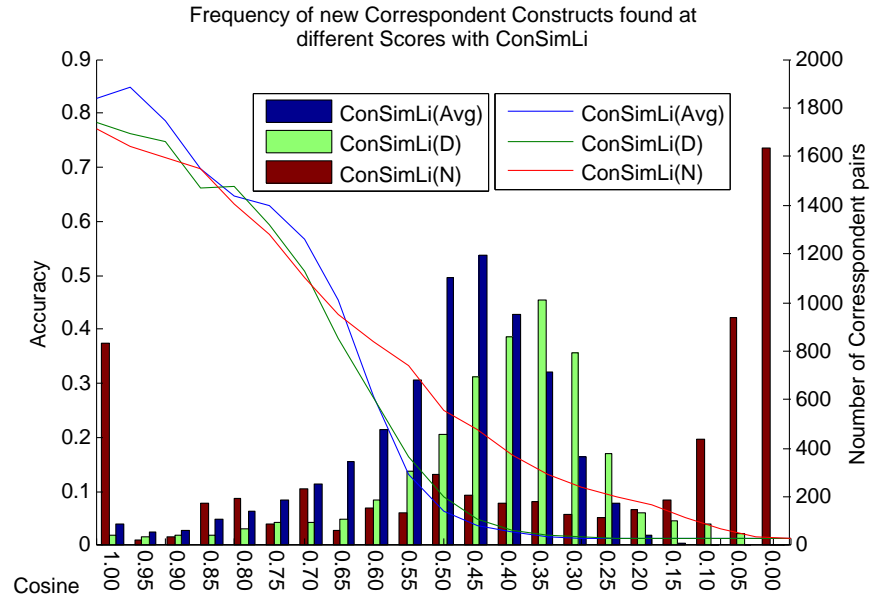


Figure 6.8: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlayed with precision (success rate) at each cutoff.

measures, the best and the worst combined models that are achieved with different measures are reported.

Similarly to the previous experiments, the similarity scores for all construct pairs are computed and the cutoff samples at the 21 cutoff scores are obtained. Then the success rates are computed and the true correspondent construct relationships in the cutoff samples are counted. The X-axis of the plot shows the error rate (which can be obtained by  $1 - accuracy$ ) and its values range between 0 to 1. The Y-axis shows the number of true correspondent constructs found at each cutoff sample. In this experiment, the performance result of ConSimLi(I) is plotted as a baseline for the comparison.

Figure 6.10 shows the results of this experiment. The ConSimLi(I+D+N) combined model is the best model because it is able to pick the most correspondent pairs with the lowest error rate. It is closely followed by LSA(I+D+N) combined model which is then overtaken by ConSimLi(I+D+N)

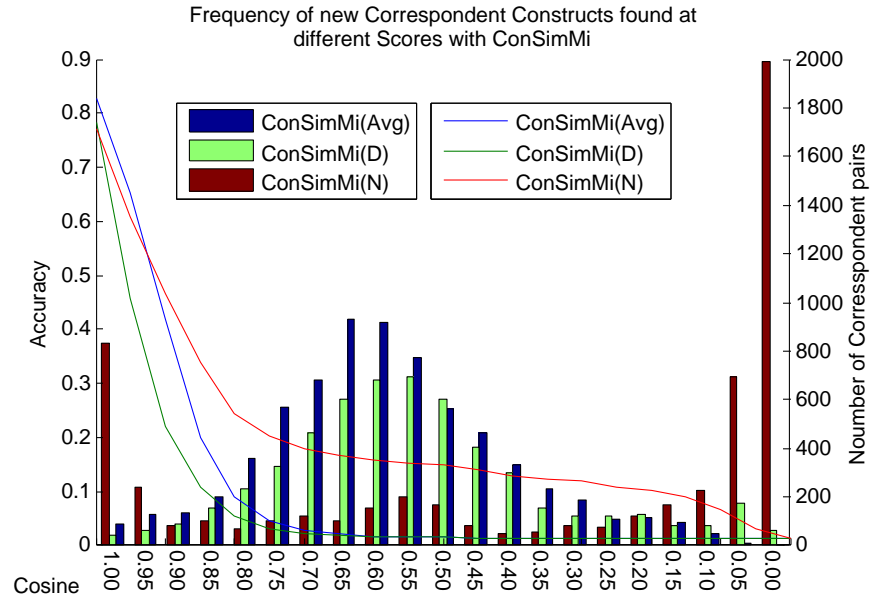


Figure 6.9: Frequency and cumulative frequency distribution of correspondent pairs across different cutoffs overlaid with precision (success rate) at each cutoff.

	Total Exact Matches	Correspondent Pairs	Accuracy
LSA(I)	99	81	82%
LSA(D)	83	63	76%
LSA(N)	989	769	78%
Li(I)	105	87	83%
Li(D)	97	74	76%
Li(N)	1079	883	82%
Mi(I)	105	87	83%
Mi(D)	221	101	46%
Mi(N)	1079	883	82%

Table 6.3: The number of total matches and the accuracy in the nine models.

at error rate = 0.43. In addition, the ConSimMi(I+D+N) combined model does not perform as well as the previous two.

The best combined model with different measures is attained by ConSimLi(I)+ConSimLi(D)+LSA(D), which only uses construct definition and item. Although it is the best heterogeneous measure model, it does not perform as well as homogeneous models

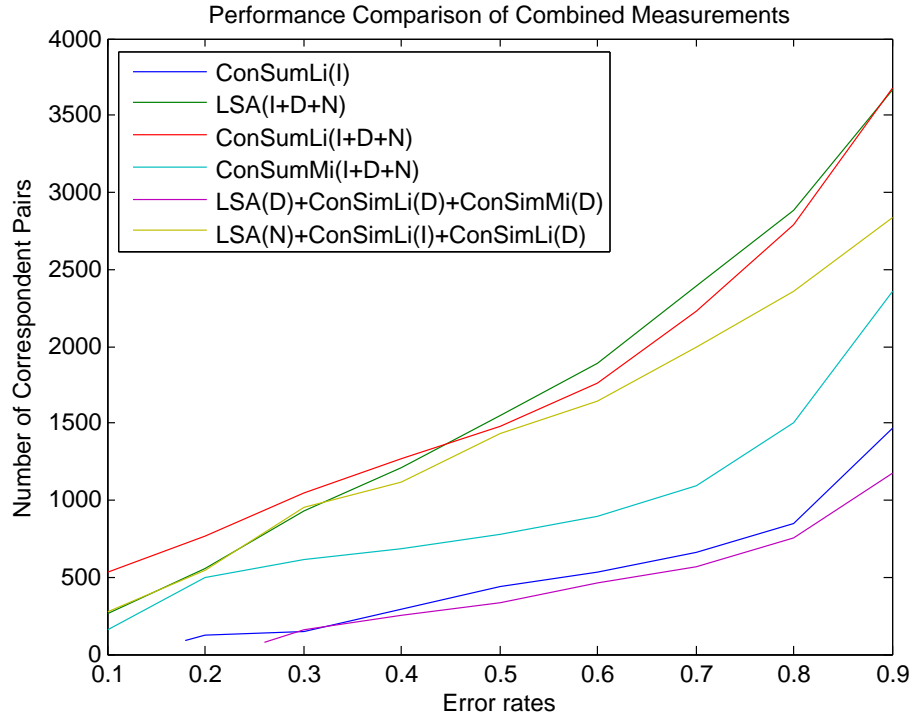


Figure 6.10: Frequency and cumulative frequency distribution of correspondent pairs across different cutoff overlayed with precision (success rate) at each cutoff.

(combining all construct properties with the same similarity measure) such as ConSimLi(I+D+N) and LSA(I+D+N). Hence, it is clear that the performance can be improved without the necessity of combining different measures.

The worst combined model is LSA(D)+ConSimLi(D)+ConSimMi(D) and it can be seen that this model has poor results than baseline. This is another piece of strong evidence clearly showing that the similarity based on definition alone (even combining it with different measures) is not able to predict construct relationship like others.

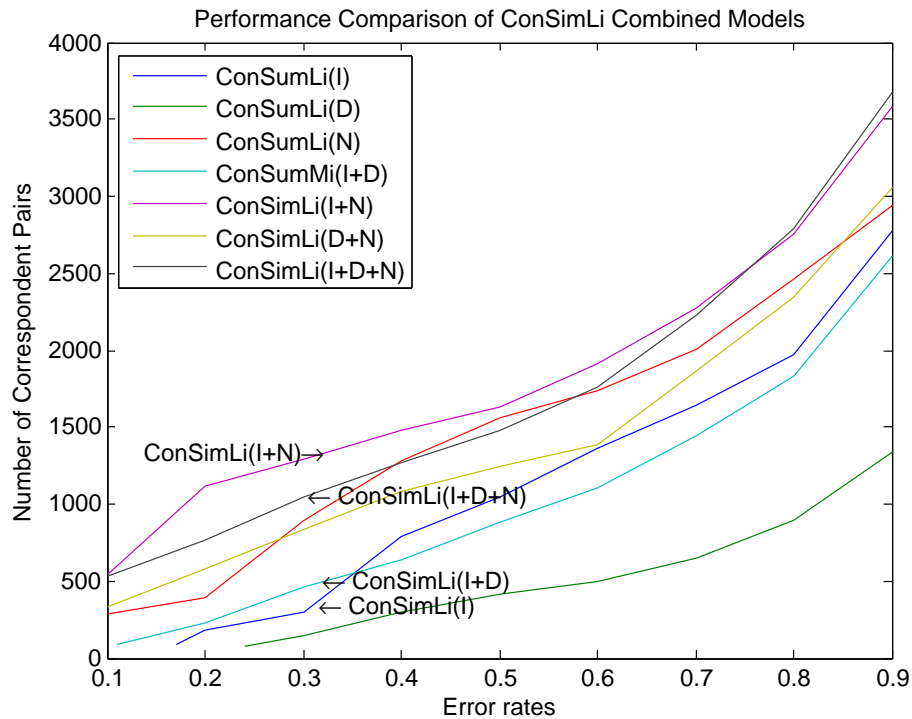


Figure 6.11: The performance of possible combined models generated by ConSimLi in finding new correspondent construct pairs.

One of the reasons that combined construct properties, especially name, is able to yield more correspondent constructs at a lower error rate is because there is a high number of similar name constructs in the database and the majority of them are correspondent constructs (see Table 6.3). Because of these exact matches, these constructs yield a similarity score of one, they inherently increase the averaged score and render them more similar (see Figure 6.7-6.9). In figure 6.11, the outcomes of all possible models generated with ConSimLi are shown, we can see that when combining item and name scores (see ConSimLi(I+N)), the additional number of new construct pairs found corresponds to the number of exact matches in the database, about 769. Likewise, when combining the score of the item with the definition in ConSimLi(I+D), the increment is seen to be due to the definition exact matches (see Table 6.3).

Detailed analysis reveals that the name property does have substantial impact on the overall score, if and only if the database contains a high number of similar name constructs. This also

can be seen in combined Mihalcea et al. measurement, which is considered less superior than other two, but able to boost its performance overall (compared to the baseline).

In summary, the experiments in this section have proven that combined measures are indeed able to find more correspondent constructs, especially when integrated with the name properties. However, this is subject to whether the database consists of sufficient similar name construct.

## 6.6 Conclusion

A number of experiments were carried out to evaluate the proposed similarity measures on construct name, definition and items to find construct relationships. Through ROC plots it was shown that measures based on ConSimLi with construct items perform best. A comparison of the two generalized functions on items, an averaged score function was settled on which has a small advantage over minimum score function. Although ROC is able to identify which approach works best, it does not really show which model produces the most accurate cutoff sample. Thus, the similarity score distribution of each model and accuracy at different cutoff scores were investigated. The results have shown that the ConSimLi with the construct items is the favorable measure.

The experiments with combined models have suggested that incorporating construct name, definition and items in the similarity measures increases the chance of discovering more true correspondent constructs, particularly with ConSimLi.

In general, the construct item is very useful for predicting the construct relationships. Also, the construct name is helpful when dealing with a high number of similar name constructs. Overall, it was discovered that construct definition does not have predictive power like that of the construct item and name.

## Chapter 7

### Use Case 1: Predicting Construct Similarity with a Unified Theory

#### 7.1 Overview

The purpose of the use case study is to prove the efficacy of the proposed similarity measures to integrate correspondent constructs into a “collective construct”. The whole idea is to cluster the independent corresponding constructs, which are developed for different theories, into the same cluster. The results attest to the efficiency of the proposed measures integrating the constructs with the unified theory as reported in Venkatesh et al. [84].

#### 7.2 Background

##### 7.2.1 Unified Theory of Acceptance and Use of Technology (UTAUT)

Technology usage and acceptance has always been one of the most widely studied areas in the information system (IS) discipline. Research models have been created to predict system use and understand how external variables (or constructs) effect internal beliefs, attitude and intentions. Over the past 20 years, a multitude of research models have been developed to explain how technologies are adopted to improve organization productivity. In 2003, Venkatesh et al. [84] presented an unprecedented study consolidating eight prominent IS acceptance models into a unified theory known as Unified Theory of Acceptance and Use of Technology (UTAUT). The elements included in the UTAUT are basically formulated from the following eight different models:

- (1) Theory of Reasoned Action (TRA),



- (2) Technology Acceptance Model (TAM),
- (3) Motivational Model (MM),
- (4) Theory of Planned Behavior (TPB),
- (5) Combined TAM and TPB (C-TAM-TPB),
- (6) Model of Personal Computer Utilization (MPCU),
- (7) Innovations Diffusion Theory (IDT), and
- (8) Social Cognitive Theory (SCT).

The unified theory can be seen as the first effort in the IS discipline to integrate similar context constructs from different models. Instead of picking and choosing one among a multitude of models and largely ignoring the others' contributions, Venkatesh et al. argue that there is a need for a unified theory whose goal is “to review and synthesize a unified view of user acceptance” without compromising existing findings [84] (p. 426).

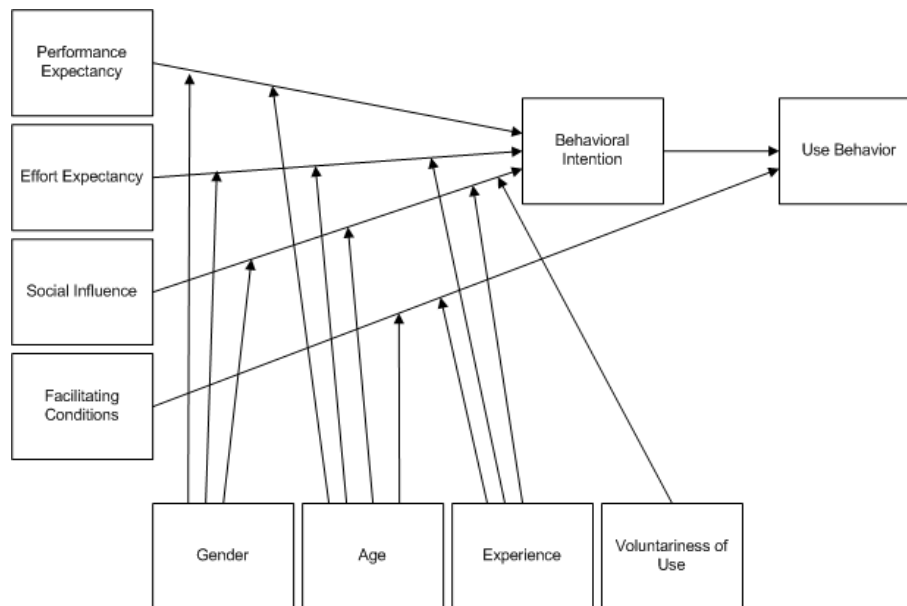


Figure 7.1: Unified Theory of Acceptance and Use of Technology Research Model.

The UTAUT research model is shown in Figure 7.1. The UTAUT postulates that four key constructs (Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions) are significant and are the direct determinants of usage intention and behavior for technology adoption but are moderated by variables of gender, age, experience and voluntariness of use. For instance, the study reports that Perceived Ease of Use become non-significant over extended and sustained usage and, Perceived Ease of Use therefore can be significant only at the early stage of technology adoption and that it can have a positive effect on Perceived Usefulness of the technology. The UTAUT is able to account for 70 percent of the variance in usage intentions towards technologies adopted, a considerable improvement on previous models which routinely explain around 40 percent of acceptance.

### **7.2.2 Conceptualization**

The conceptualization of UTAUT is a non-trivial task. In the paper, Venkatesh et al. claim they took six months to collect data from four different organizations which involved giving out questionnaires containing items measuring all the constructs from the eight models. The purpose of collecting the data was to identify commonalities of the eight models and to find constructs that were consistently significant and most influential in all time periods. Having reviewed and empirically compared the eight models, they formulated four constructs out of the constructs in the existing models which are considered the direct determinants of user acceptance and usage behavior. However, the paper does not specifically describe in detail how those constructs were derived.

They derived four constructs out of the constructs from the eight models. These four constructs are not new. Each of the constructs is, in fact, rooted on constructs from the eight models which pertain to the respective concepts. In Appendix C, all the root constructs and their corresponding measurement items are listed in four separate tables, which map to the four derived dominant constructs in UTAUT.

The final part of the conceptualization is to empirically validate the constructs with the

collected data and cross-validate it with data from a different organization.

### 7.3 Task Description

UTAUT is viewed as a highly validated model since the paper describing it has been cited over 3,600 times<sup>1</sup>, and it has also been extended a number of times in other studies [86, 77]. UTAUT can be seen as a testing ground for the proposed construct similarity measures in predicting the construct relationships. Of particular interest are the 14 root constructs which are consolidated into four key constructs in the unified theory. The consolidation in the original study is the result of extensive longitudinal operational study examining construct convergent and discriminant validity. By using the proposed similarity measures, it is believed that construct similarity could be computed and the correspondent constructs placed near to each other whereas independent constructs are placed far away from each other on a map.

To evaluate the similarity measures robustly, a semantic space with articles prior to year 2002 was built. This was done to exclude any evidence that was possibly linked to UTAUT which could lead to overfitting in the semantic space.

Similarities of the 14 root constructs were computed with the three similarity measures. As discussed, the construct similarity score is the average of the two most similar item scores. The result of similarity is represented as a 14x14 inter point diagonal matrix,  $S$ , where each element represents inter-construct similarity. The  $S$  is then converted into a dissimilarity matrix<sup>2</sup>,  $D = 1 - S$  then the  $D$  is transformed into configuration of points to provide original distance in 2D visual representation.

For the purpose, MDSCALE a non-classical multidimensional scaling (MDS) function from the statistic toolbox in Matlab software was used. MDS is a set of data analysis techniques that display the structure of distance-like data as a geometrical picture. It is also known as Principal Factor Analysis and can be used as a dimension reduction method, specifically reducing the data to

---

<sup>1</sup> The number of citations is obtained through Google Scholar, as of 6/21/2011

<sup>2</sup> The algorithm we used to visualize the construct similarity only accepts dissimilarity matrix

a distance matrix and creating a new configuration of points retaining only the first few dimensions of those points. It takes a dissimilarity matrix and outputs a coordinate matrix whose configuration minimizes a loss function called strain. In this study, Sammon metric scaling was used which finds a new reduced-dimensionality coordinate system for the configuration points such that the an error criterion between distances in the given space, and distances in the result space, is minimized. Once the configuration points corresponding to each construct are in place, they are visualized in a scatter-plot.

## 7.4 Results and Discussion

Below is the scatter plot for the 14 configuration points generated with ConSimLi(I). The average similarity score of the 14 constructs in this setting is 0.61 and the similarity scores range between 1 to 0.45. The distances between configuration points in the plot represent the relative similarity of constructs. Hence, the closer they are located, the more correspondent they are proposed to be.

From the scatter plot in Figure 7.2, it can be seen that the 14 root constructs are spread out and grouped into four “collective” clusters. The same color constructs are located next to each other. The four clusters are the red cluster (Performance Expectancy) in the center of the lower quadrants, the blue cluster (Social Influence) at the upper right quadrant, the black cluster (Facilitating Conditions) at the upper left and the green cluster (Effort Expectancy) below the black cluster. The root constructs in Performance Expectancy (the red constructs) are predicted to be similar by the measure and are placed near to each other. They are also predicted to be dissimilar to others, thus are placed far from them.

Likewise, data for root constructs that are from the Effort Expectancy (the green constructs) and the Social Influence (the blue constructs), are nearly all being placed close together.

However, the black root construct, Perceived Behavioral Control (PBC), is placed close to the green constructs instead of being placed close to the two similar member constructs, Facilitating Condition (FC) and Compatibility (CMPT) are being placed close to the green constructs. Detail

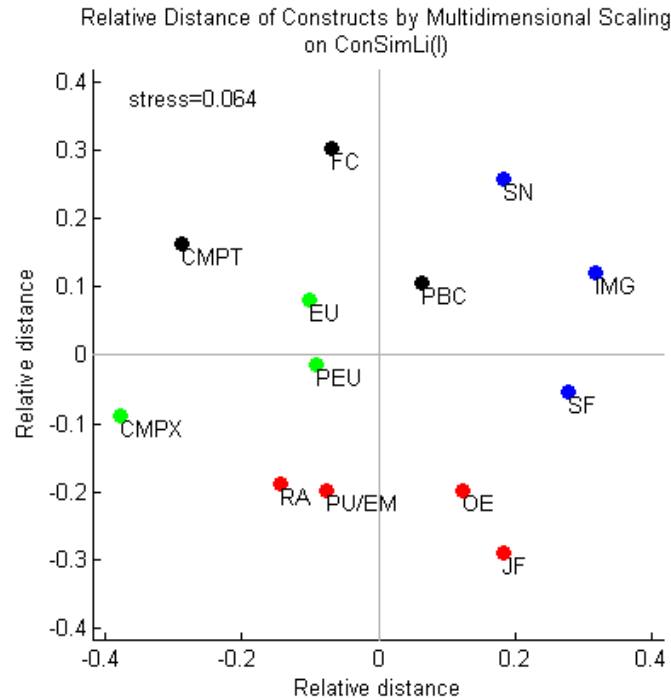


Figure 7.2: Visual representation of the pattern of similarity of 14 root constructs appearing in Venkatesh et al.[84] that are computed with ConSimLi(I). Different colors are used to indicate the same group of root construct as reported in the paper.

analysis of similarity at item level reveals that PBC has higher correlation to both the green constructs Perceived Ease of Use (PEU) and Ease of Use (EU)(see Table 7.1–7.2), instead of its root member constructs, FC and CMPT (see Table 7.3–7.4).

Perceived Behavioral Control(PBC) Items	Perceived Ease of Use (PEU) Items	Score
“It would be easy for me to become skillful at using the system.”	“Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system.”	0.75
“Learning to operate the system would be easy for me.”	“Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system.”	0.74

Table 7.1: Two most similar items and their score between Perceived Behavioral Control(PBC) and Perceived Ease of Use (PEU).

PBC is placed closer to the construct belonging to Effort Expectancy due to having higher

Perceived Behavioral Control(PBC)	Ease of Use (EU) Items	Score
Overall, I believe that the system is easy to use.	I have the resources necessary to use the system.	0.66
Overall, I believe that the system is easy to use.	I have the knowledge necessary to use the system.	0.66

Table 7.2: Two most similar items and their score between Perceived Behavioral Control(PBC) and Ease of Use (EU).

Perceived Behavioral Control(PBC) Items	Compatibility (CMPT) Items	Score
“The system is not compatible with other systems I use.”	“Using the system is compatible with all aspects of my work.”	0.67
“I have control over using the system.”	“Using the system is compatible with all aspects of my work.”	0.56

Table 7.3: Two most similar items and their score between Perceived Behavioral Control(PBC) and Compatibility (CMPT).

Perceived Behavioral Control(PBC)	Facilitating Control (FC) Items	Score
“I have the resources necessary to use the system.”	“Guidance was available to me in the selection of the system.”	0.67
“I have the knowledge necessary to use the system.”	“Guidance was available to me in the selection of the system.”	0.62

Table 7.4: Two most similar items and their score between Perceived Behavioral Control(PBC) and Facilitating Control (FC).

similarity scores (0.73 and 0.65 respectively) to two of the green root constructs, PEU and EU. The main reason behind this is that they both have highly similar items with PBC which are about the subjects of ease of use and operations of the system. In contrast, although PBC has high similarity to the first item in CMPT, because of the lower score of the second item, it renders PBC less similar to CMPT. Likewise, both the most similar items in PBC and FC are not that similar.

The green Complexity (COMPX) construct is an interesting one. When examining its inter-point similarity with other constructs, it does not correlate well with the other 13 constructs and its maximum pairwise construct similarity score is merely 0.6, which is a little below the average similarity score. Due to its weak correlation to other constructs, its configuration point is placed far to the right bottom of the plot, far away from other constructs. However, if the similarity matrix is

examined, the two most similar constructs to COMPX are indeed PU and PEU, which are its root member constructs. Because of that, it is located next to PU and PEU to project their relative distances.

The following section includes the scatter plots for configuration points created with the combined models: ConSimLi(I+N), ConSimLi(I+D) and ConSimLi(I+D+N).

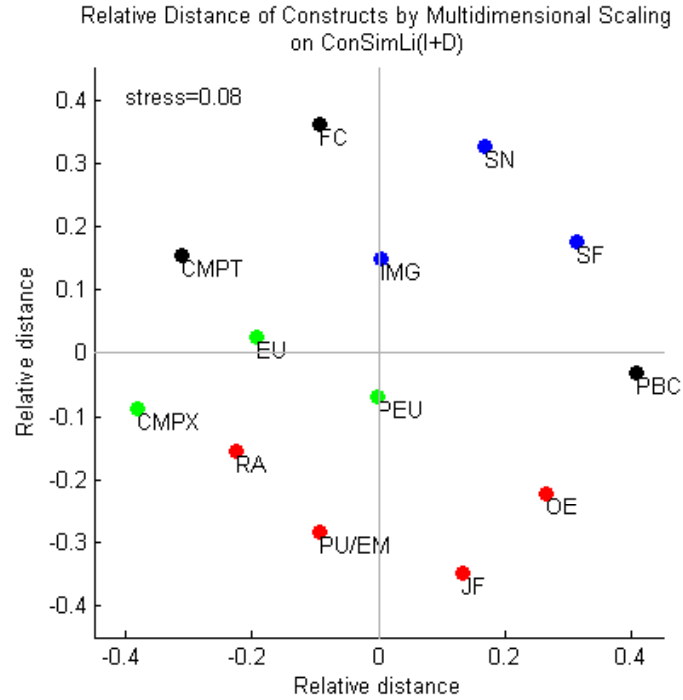


Figure 7.3: Visual representation of pattern similarity generated with ConSimLi(I+D)

Figure 7.4– 7.5 show the scatter plot of construct similarity with different combined measures. Although the combined measures are also able to predict almost all of the construct relationships correctly, the plot with the ConSimLi(I+N) (see Figure 7.4) can cluster the root constructs into the four key constructs as they are reported in the original paper. The four key constructs are spread out nicely into the four quadrants.

In contrast to our previous findings which show that the ConSimLi(I+N+D) is the best construct relationship predictor, the plot created with ConSimLi(I+N+D) (see Figure 7.5) is not as promising as others as there are some constructs being misplaced.

To sum up, we believe there is no single true or false answer to which measure does the most accurate job. All the results are left to be investigated through operational research, which is beyond the scope of this study. In general, we are excited to see that most proposed measures (either stand alone or combined) are able to group the 14 root constructs into four key constructs as reported in the original paper.



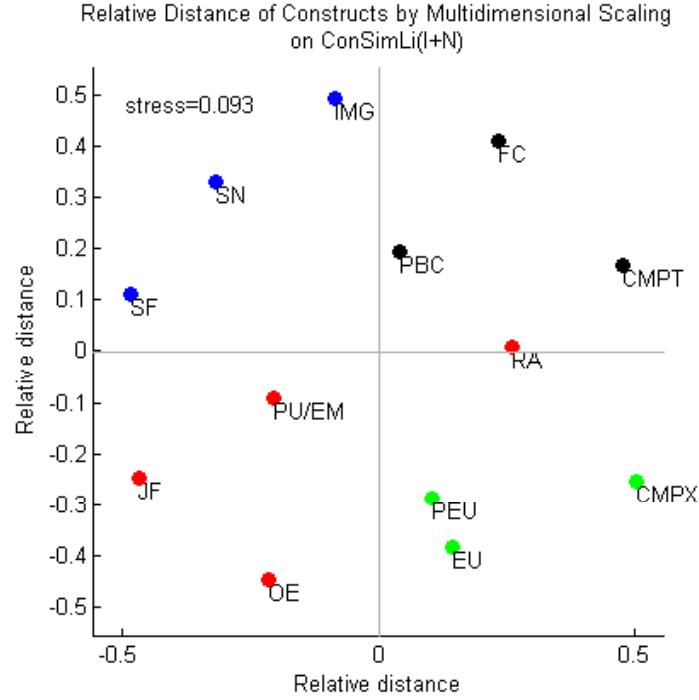


Figure 7.4: Visual representation of pattern similarity generated with ConSimLi(I+N)

## 7.5 Conclusion

This chapter has showcased our proposed measures in the application of multi-model knowledge integration. This is to simulate the process of integrating related knowledge that is represented by constructs from multiple theories, such as how UTAUT [84] ] is operationalized. Instead of carrying out the full conceptualized operation in a full scale, which requires enormous resources, we can first employ the proposed measures to narrow down to the constructs that are deemed similar. Here, we have facilitated the proposed measures to predict construct similarities of 14 root constructs which are in turn being transformed with distance functions and represented in scatter plots. Corresponding constructs in the plot are assumed to be located near to each other whereas independent constructs are located far from each other. The empirical results suggest that for the most part, using the proposed measures, we are able to correctly predict almost all the root constructs that pertain to the four key constructs. We indeed are excited to learn that the best

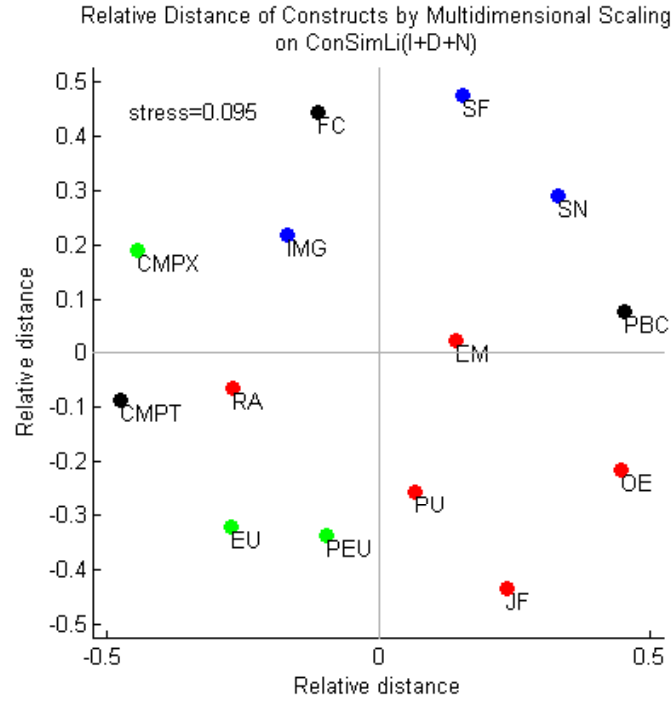


Figure 7.5: Visual representation of pattern similarity generated with ConSimLi(I+D+N)

prediction can be achieved with ConSimLi(I+D). Also, in contrast with our previous findings, we have learned that ConSimLi(I+D+N) does not perform as expected in this task. Finally, the relative construct distances visualized in the plots are by no means a real representation of construct relationships, as the actual relationships can be only confirmed through operational study.

## Chapter 8

### Use Case 2: ConstructNet

#### 8.1 Overview

In this chapter, we build a network of constructs termed ConstructNet, in order to examine the efficiency of the similarity measures to relate constructs to the others based on construct items. The goal is here to automatically create a visualization of ConstructNet that will allow experts to immediately review relationships between constructs and make adjustments. In this use case, the focus is on building the network, evaluating it using the gold standard, and examine reasons behind structural failures in the network. Two construct networks are chosen for the in-depth study.

#### 8.2 Task Description

To visualize the network of constructs, the similarity scores of construct relationships are computed using ConSimLi(I). The reasons for using only construct items here are that:

- (1) Items are operationally used to validate constructs (see Section 2.8);
- (2) Definition is not a good predictor;
- (3) Name ends up with lot of connectivity because of similar name constructs, but this connectivity has low information value; and
- (4) Combined models might end up with high connectivity because of similar construct name (for the same reason as (3)), or degrade the connectivity if construct definitions are included.

In the following experiments, the ConstructNet is built with the construct relationships which have similarity scores equal to or above 0.8. The threshold results in a total of 407 constructs with 1107 relationships. The ConstructNet is depicted in Figure 8.1.

### 8.3 Results and Discussion

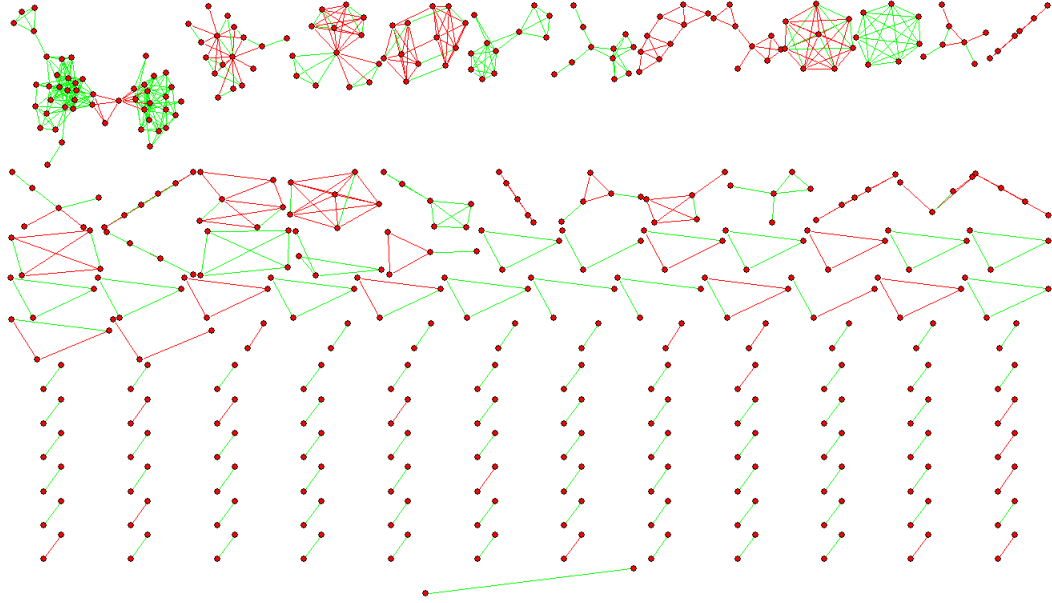


Figure 8.1: ConstructNet.

Figure 8.1 shows a number of disconnected networks in the ConstructNet. Constructs are represented as the red vertices, and all constructs are connected either with green or red edges. The green edges represent correspondent relationships that are in agreement with the gold standard whereas the red edges indicate constructs relationships that are not in agreement with the gold standard.

There are independent construct networks because those construct relationships with similarity score less than 0.8 are not being visualized here (thus makes them isolated visually). The structure and location of the clusters are randomly determined by the Kamada-Kawai energy for optimized visualization. So distances between constructs in the space do not represent construct

similarity. If required for explanation, the construct similarities are represented by edge value (Figure 8.1 does not show the construct similarities).

### 8.3.1 Perceived Ease of Use and Perceived Usefulness

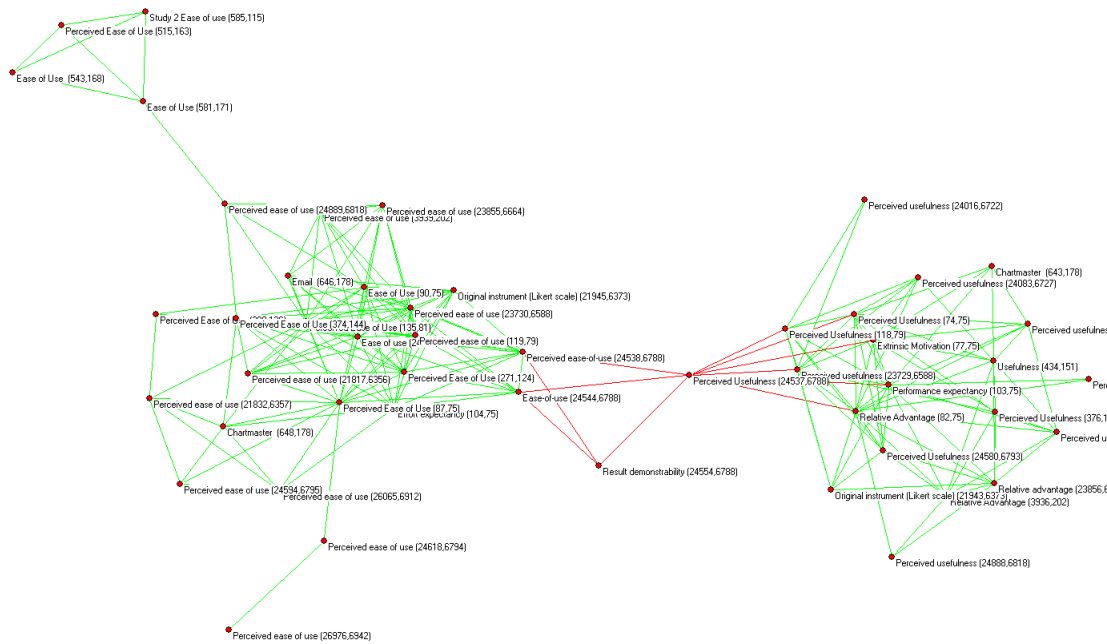


Figure 8.2: Connectivity for constructs pertaining to Perceived Usefulness and Perceived Ease of Use.

Figure 8.2 shows the detail view of the largest construct network that is found at the upper left corner in of Figure 8.1. Each construct in the network is labeled with the name, and followed by their unique variable identity and the source identity (in the parenthesis) from the database.

The figure shows the connection of the constructs from the subcategories *Perceive Usefulness* (to the right of the network) and *Perceived Ease of Use* (to the left of the network). And for the most part, almost all construct relationships in the network are predicted correctly (green edges) by the similarity measure used.

The figure also shows a small number of independent relationships (red edges). It is also interesting to learn that the construct *Perceived Usefulness* (24537,6788) is bridging the constructs from the two subcategories.

Is this a case of relationship misclassification? To answer the question, the properties of the *Perceived Usefulness* (24537,6788) have to be examined and compared to one of the interconnected

constructs, *Perceived Usefulness*(74,25). The main reason these two constructs are selected for analysis because their relationship has a similarity score of 1.

Name	<i>Perceived Usefulness (24537,6788)</i>	<i>Perceived Usefulness (74,25)</i>
Def	The extent to which a technological innovation is expected to improve the potential adopters performance.	The extent to which a person believes that using a particular technology will enhance her/his job performance.
Source	Research Report: Richness Versus Parsimony in Modeling Technology Adoption DecisionsUnderstanding Merchant Adoption of a SmartCard-Based Payment System	User acceptance of information technology: toward a unified view
Items	1.Using the system improves my performance in my job. 2.Using the system in my job increases my productivity. 3.Using the system enhances my effectiveness in my job. 4.I find the system to be useful in my job.	1.Using the system improves my performance in my job. 2.Using the system in my job increases my productivity 3.Using the system enhances my effectiveness 4.I find the system to be useful in my job

Table 8.1: Perceived Usefulness constructs from different articles.

In Table 8.2, both constructs can be seen having identical measurement items. This explains why the construct relationship has similarity score of one. When annotators and experts categorized the two constructs during the gold standard creation, they placed them into different categories because they have determined that the *Perceived Usefulness (24537,6788)* is describing a concept at the organization level, whereas *Perceived usefulness (74,25)* is focusing on a concept at the individual level. For that reason they were labeled as dissimilar which yields independent relationship.

Finding two constructs at two different levels of scale is almost impossible by just looking at the construct properties. Although the definition in Table 8.2 seems to have hint (*adaptor* and *person*), but according to ConSimLi, the two definitions only yield similarity score of 0.80, which is not really helpful to reflect the embedded context in them.

This is not a case of misclassification, as differentiating constructs at two different level scales is a complicated task as it involves one's background knowledge on the subjects and how

the constructs are coded in the original paper. In this case, judging both construct properties does not help revealing the actual relationship, and the relationship can be only confidently determined through reading the articles where the constructs are extracted from.

Besides *Perceived Usefulness* (24537,6788), Figure 8.2 also shows three construct that are independent. It should be noted that these construct are from the same article. Generally, constructs from the same paper are not categorized as correspondent. For example, both *Usefulness* (24537,6788 and *Result Demonstrability*(24544,6788) have high similarity score but they are independent because they are from the same paper.

The reason the same paper constructs included in the network is because they have high similarity scores. The similarity score is due to the the word-word similarity measure with LSA tends to associate co-occurred words from the same unit of document with high similarity. In Table 8.2, it can be seen that *Usefulness* (24537,6788) and *Result Demonstrability*(24544,6788) have high similarity scores for their items.

<i>Usefulness</i> (24537,6788)	textitResult Demonstrability(24544,6788)	similarity
Using the Exact card system increases the productivity of me and my staff.	The impact of using the Exact card system is apparent to my staff and me.	0.831
Using the Exact card system enhances the on-the-job effectiveness of me and my staff.	The impact of using the Exact card system is apparent to my staff and me.	0.813
Using the Exact card system improves the job performance of me and my staff.	The impact of using the Exact card system is apparent to my staff and me.	0.815
Using the Exact card system increases the productivity of me and my staff.	My staff and I could communicate to others the consequences of using the Exact card system.	0.803

Table 8.2: Similarity scores on two set of items.

Detailed analysis reveals that the higher scores are due to the exact word matches and high similarity score of the co-occurrence of certain words instead of determinant keywords. In fact, even expert annotators without specific knowledge of the *Result Demonstrability* construct may be tempted to code it as correspondent.



### 8.3.2 Perceived Usefulness and Cognitive Absorption

There are studies show that users holistic experiences could be important in explaining in technology acceptance and usage [3, 91]. One such experience is cognitive absorption (CA). CA is an intrinsic motivation related construct, and it was found that it has a positive effect on the perceived usefulness of the information technology[3, 91]. Hence, besides designing information technologies (IT) that are perceived to be useful and easy to use, it is also very important to ensure that it has pleasant and interesting qualities as these qualities directly enhance perceived usefulness, and ease of use.

An example of CA is the experience of using technologies that are visually rich and appealing such as game-based training environments and they are more enjoyable [82] and more likely to result in cognitive absorption.

In the following section, construct network related to CA are visualized and discussed.

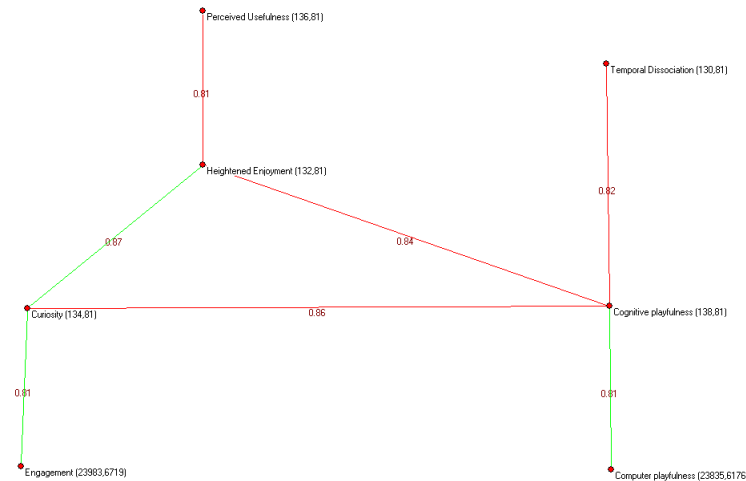


Figure 8.3: Connectivity for constructs pertaining to Cognitive Absorption.

Figure 8.3 shows the construct network pertaining to CA. It shows that there are five constructs from the same article. The five constructs are: *Perceived Usefulness*, *Heightened Enjoyment*, *Curiosity*, *Temporal Dissociation*, and *Cognitive playfulness*. These constructs are reported in the article Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage. The article reports five CA constructs and three of them are present in the network.

The network shows that the new *Perceived Usefulness* (136,81) does not associate with the network that pertains to the perceived usefulness seen in Figure 8.2. This may be because its items is significantly different from the the *Perceived Usefulness* (24537,6788).

Name	Perceived Usefulness (74)	Perceived Usefulness (132)
Source	The degree to which a person believes that using a particular system would enhance his or her job performance.	Null (the definition was not found in the paper).
Def.	User acceptance of information technology: toward a unified view	Time Flies When You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage
Items	1.Using the system improves my performance in my job. 2.Using the system in my job increases my productivity. 3.Using the system enhances my effectiveness in my job. 4.I find the system to be useful in my job.	1.Using the Web improves my performance in college. 2.Using the Web enhances my productivity. 3.Using the Web enhances my effectiveness in college. 4.I find the Web useful in my college activities

Table 8.3: Perceived Usefulness properties.

Table 8.3 shows the properties of the two constructs. The table shows that both constructs have different items. Although both items measure performance in different aspects, the settings when the measurements take place are completely different. *Perceived Usefulness (74)* which is from [84] was used to measure a job performance related to a technology usage in an organization. On the other hand, the *itPerceived Usefulness (132)* is measuring playful

During the categorization exercise, *Perceived Usefulness (74)* is placed under “usefulness, Individual” whereas *Perceived Usefulness (132)* is categorized under “affect Towards Technology (Use)”. Clearly, both constructs are deemed differently by experts.

In-depth analysis reveals the reason the proposed measure is able to differentiate them is because of the use of the words “Web” in the items. The constructs in the original paper of the *Perceived Usefulness (132)* are measuring playful and fun and the Web are used in every item. Due to the way text-text similarity work, the two constructs are deemed differently by the similarity measure.

Figure 8.4 also shows five construct which are from the same papers. As expected, four of the interconnect relationships are labeled as independent except the relationship between *Hightened Enjoyment (132)* and *Curiosity (134)*. The two constructs, although they are from the same paper,

are categorized in the same category because one of them have general item which measure what the other construct measures.

The connectivity can be consolidated by merging the constructs from the same paper into a collective construct Cognitive Absorption, to clearly show the connection with constructs that are in different papers.

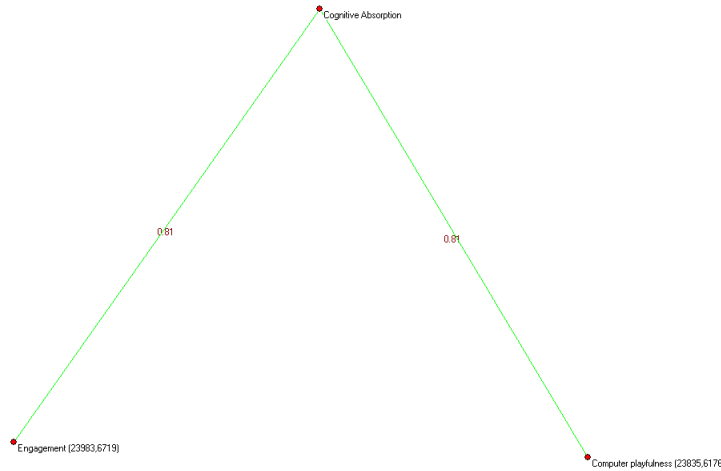


Figure 8.4: Consolidate connectivity for constructs pertaining to Cognitive Absorption.

It is the intention of the study to discover construct relationship that exist between constructs, that are appeared in different articles which have not been studied before. The use of ConstructNet here is to draw multidimensional constructs from different theories so that the researchers can analyze them and select them in a study before proceeding to derived the constructs operationally, which is a laborious operation.

## 8.4 Conclusion

Metatheory is a theory that is built upon existing theories. The ConstructNet described here by no mean a metatheory, but it can be seen as a potential and useful tool to leverage the constructs helping scientists to expedite the process in building a metatheory through examining the connectivity of the constructs. In this section, the focus of the study is to see the potential of

nurturing the collected constructs and weave them into an interlocking system using the proposed similarity measure. This allows scientists to study and adjust each other relationship. It is still too early at this stage to claim that ConstructNet is able to represent the real knowledge, but we believe that the ConstructNet can serve as a shortcut allowing scientist to integrate exiting knowledge and develop new theory, which in turn expedite scientific progression. To evaluate ConstructNet, its connectivity was evaluated using the gold standard and it was found that the majority of the connectivity are in agreement with the gold standard. We uncovered that most of the independent relationships were a result of the connectivity of the constructs from the same paper.

## Chapter 9

### Conclusion, Contributions and Future Works

#### 9.1 Overview

In this chapter, the conclusion of the thesis is stated before discussing the contributions and opportunities for future investigation.

#### 9.2 Conclusion

Human Behavior Project (HBP) aims to integrate the existing knowledge from social and behavioral science that is encoded in constructs. It is also the intention of the project to interconnect the constructs from different theories so that their correlations may be studied. It is believed that if scattered constructs could be integrated into an interlocking system, the interplay of the constructs can help scientists explain and predict human behavior. Scientists have reported that 93 percent of human behavior is actually predictable.

In this thesis, the automated computational approaches which are derived from the natural language processing advancements have been proposed to predict the construct relationships, even though the relationships have not been studied and are rooted in different theories. Instead of finding the construct relationships through psychometric methods, which is an extremely laborious process, text similarity measures were employed on construct name, definition and items to predict if constructs are correspondent (similar) or independent (dissimilar). The study has created opportunities to explore the semantic relationship of the constructs through automated computational approaches such as text similarity measures.

A gold standard was created by categorizing correspondent constructs into the same categories in order to robustly evaluate the proposed measures. Construct categorization resulted in two levels of hierarchical taxonomy which yielded the gold standard of construct relationships. To validate the gold standard, 300 construct pairs in from the categories and with equal proportion of relationship types were drawn and given to experts for labeling. Kappa coefficient were used to evaluate the inter-agreement between the experts and the gold standard. The computed Kappa coefficient scores has shown that the gold standard has substantial agreement with experts, which were 68 and 77 percent respectively. The efficacy of the derived measures were compared with Latent Semantic Analysis (LSA), a reputable similarity measure. The study showed that the model built with ConSimLi worked reasonably well in predicting construct relationships using construct items. The study also extended the models to combined models, which were created by averaging similarity scores in candidate models. The combined models were better at predicting construct relationship, partly due the existence of high number similar name constructs in the database. Experimental results showed that combining all the construct properties yielded the best combined model.

Perhaps the most encouraging finding was the demonstration of the proposed measures used to “integrate” the 14 root constructs obtained from different theories into the four key constructs as it was reported in [84]. By using Venkatesh et al. [84]’s unified model as a reference, the measures based on the ConSimLi was able to place the same root member constructs close to each other when they were projected into a scaled scatter-plot.

Finally, the study presented ConstructNet that was created with the construct relationships computed by ConSimLi using construct items. The goal of building the ConstructNet is to allow experts to learn and study the relationships between constructs from different disciplines. Preliminary analysis on selected construct connectivity in the ConstructNet shows that the measure was able to satisfactorily predict the construct relationships that are in the same category in the gold standard.

It is the hope of the study to extend it to a search toolkit that will allow theory developers to find related constructs or constructs pertaining to existing theories before engaging in the process



of developing new theory. In addition, the ConstructNet can be treated as an additional knowledge base in conjunction with the use of machine learning algorithm in predicting human judgement. It is believed that many applications such as review systems, sentiment analysis and question-answering can benefit from this knowledge base.

### 9.3 Contributions

This cross-discipline study has made a four major contributions in the computer science, social and behavioral disciplines:

- (1) The study has derived text similarity measures from prior literature which are able to work within a specific domain and predict construct relationships based on the construct properties
- (2) The gold standard that has been created in this study not only can be used to evaluate the proposed measures but can also be used to benchmark text similarity measures. To date, most of the text similarity measures reported in the literature are evaluated independently or in an adhoc manner, which lacks uniformity. The items used to measure a construct, which have the same context are short sentences. They can be used to benchmark sentence similarity measures.
- (3) The study also presents the first attempt at large-scale construct integration through a computational approach, which was visualized in ConstructNet. It makes possible the discovery of latent connections among constructs through constructs textual properties.
- (4) The study serves as the only attempt to date that explores the possibility of automatically creating construct maps and interrelating their relationships through computational approaches in accordance to Cronbach and Meehls [14] suggestion, which said that the constructs maps are the only method for theory representation as well as validation of underlying constructs.

## 9.4 Future Work

The proposed text similarity measures, ConSimLi and ConsimMi, that are used to predict construct similarities are still far from perfect. Although this thesis has shown the efficacy of the measures, many issues still exist which require further investigation. We are particularly interested in investigating the following four issues in future work:

### 9.4.1 Improving the Proposed Similarity Measures

The text similarity measures rely on a knowledge base for computing word-word similarity. Although our thesis has proven that it is possible to predict construct relationships with semantic matching, our findings have also shown that the similarity measures did not work well on antonyms and rare words. Besides deriving a more effective similarity measure, it is believed that it is also important to improve the semantic knowledge base used for word-word similarity.

### 9.4.2 Attesting our Proposed Measures on a Much Large Scale Unified Model

The study results have shown positive outcomes when experimenting with the proposed measures within the unified models. We believe it will be more convincing if they can be attested with a theory which has more diverse constructs.

### 9.4.3 Qualitative Analysis of Construct Relationships in the ConstructNet

In the thesis, the gold standard was used to evaluate the predicted relationship in ConstructNet. An in-depth qualitative analysis with domain experts on the validity of relationships in the ConstructNet will make the study more valuable. It is impossible at this point to learn if disconnected constructs in the ConstructNet are independent constructs, or connecting constructs that are truly related. This is because there is no large scale research model available to be evaluated. It is believed that the most valuable follow-up task is to appoint experts to perform a thorough investigation on familiar areas (for example *Perceived Usefulness* or *Perceived Ease of Use*), and derive

the construct relationships manually and operationally and use them to evaluate our similarity measures.

## Bibliography

- [1] Lucene. <http://lucene.apache.org/>.
- [2] David Adger. Why theory is essential: the relationship between theory, analysis and data | subject centre for languages, linguistics and area studies. <http://www.llas.ac.uk/resources/gpg/405>, 2010.
- [3] R. Agarwal and E. Karahanna. Time flies when youre having fun: Cognitive absorption and beliefs about information technology usage. MIS quarterly, pages 665–694, 2000.
- [4] American Psychological Association and National Council on Measurement in Educa. Standards for Educational and Psychological Testing 1999. American Educational Research Association, 2nd edition, 1999.
- [5] C. F Baker, C. J Fillmore, and J. B Lowe. The berkeley framenet project, pages 86–90. Association for Computational Linguistics, 1998.
- [6] C. Banville and M. Landry. Can the field of mis be disciplined? Communications of the ACM, 32(1):48–60, 1989.
- [7] H. Barki. Thars gold in them thar constructs. ACM SIGMIS Database, 39(3):9–20, 2008.
- [8] Roger B. Bradford. An empirical study of required dimensionality for large-scale latent semantic indexing applications, pages 153–162. ACM, 2008.
- [9] C. Buckley, G. Salton, J. Allan, and A. Singhal. Automatic query expansion using SMART: TREC 3. In Overview of the Third Text REtrieval Conference (TREC-3), pages 69–80. Diane Pub Co, 1995.
- [10] Curt Burgess, Kay Livesay, and Kevin Lund. Explorations in context space: Words, sentences, discourse. Discourse Processes, 25(2):211–57, 1998.
- [11] M. Chuttur. Overview of the technology acceptance model: Origins, developments and future directions. 2009.
- [12] J. Cohen. A coefficient of agreement for nominal scales. Educational and psychological measurement, 1960.
- [13] J. A. Colquitt and C. P. Zapata-Phelan. Trends in theory building and theory testing: A five-decade study of the academy of management journal. Academy of Management Journal, 50(6):1281, 2007.

- [14] L. J. Cronbach and P. E. Meehl. Construct validity in psychological tests. The Foundations of Science and the Concepts of Psychology and Psychoanalysis, pages 174–204, 1968.
- [15] L. J. Cronbach and G. Murphy. Essentials of psychological testing. Harper & Row New York, 1970.
- [16] F. D. Davis. A technology acceptance model for empirically testing new end-user information systems: theory and results. 1986.
- [17] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 13(3):319–340, 1989.
- [18] F. D. Davis and V. Venkatesh. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. International Journal of Human-Computer Studies, 45(1):19–45, 1996.
- [19] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. Journal of the American society for information science, 41(6):391–407, 1990.
- [20] D. E.R. Denning. Information warfare and security, volume 118. Addison-Wesley, 1999.
- [21] A. A. DiSessa. Changing minds: Computers, learning, and literacy. The MIT Press, 2001.
- [22] M. Fishbein and I. Ajzen. Belief, attitude, intention and behavior: An introduction to theory and research. 1975.
- [23] D. Freitag, M. Blume, J. Byrnes, E. Chow, S. Kapadia, R. Rohwer, and Z. Wang. New experiments in distributional representations of synonymy, pages 25–32. Association for Computational Linguistics, 2005.
- [24] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. Communications of the ACM, 30(11):964–971, 1987.
- [25] R. M. Furr and V. R. Bacharach. Psychometrics: an introduction. Sage Publications, Inc, 2008.
- [26] K. Glanz, B. K. Rimer, and National Cancer Institute (US). Theory at a glance: A guide for health promotion practice. US Dept. of Health and Human Services, National Cancer Institute, 2005.
- [27] Aria Haghighi and Dan Klein. Coreference resolution in a modular, entity-centered model. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, page 385393, 2010. ACM ID: 1858060.
- [28] P. D. Harms and M. CRED. Remaining issues in emotional intelligence research: Construct overlap, method artifacts, and lack of incremental validity. Industrial and Organizational Psychology, 3(2):154–158, 2010.
- [29] C. Ho, M. A.A. Murad, R. A. Kadir, and S. C. Doraisamy. Word sense disambiguation-based sentence similarity, pages 418–426. Association for Computational Linguistics, 2010.

- [30] National Cancer Institute. Grid-Enabled measures database. <https://www.gem-beta.org/Public/ConstructList.aspx?cat=1>.
- [31] Aminul Islam and Diana Inkpen. Semantic text similarity using corpus-based word similarity and string similarity. ACM Trans. Knowl. Discov. Data, 2(2):10:110:25, Jul 2008. ACM ID: 1376819.
- [32] J. J Jiang and D. W Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. Arxiv preprint cmp-lg/9709008, 1997.
- [33] K. S Jones. A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 28(1):11–21, 1993.
- [34] Karen Sparck Jones. Synonymy and semantic classification. Edinburgh University Press, 1986.
- [35] Daniel Jurafsky and James H. Martin. Speech and Language Processing. Prentice Hall, 2 edition, May 2008.
- [36] John L. Kennedy. But what are the behavioral sciences? Science, 144(3619):683 – 684, May 1964.
- [37] P. Kline. A psychometrics primer. Free Assn Books, 2000.
- [38] April Kontostathis, William M Pottenger, and Brian D Davison. Identification of critical values in latent semantic indexing. FOUNDATIONS OF DATA MINING AND KNOWLEDGE DISCOVERY, pages 333–346, 2005.
- [39] K. L. Kraemer and W. H. Dutton. Survey research in the study of management information systems. The information systems research challenge: Survey research methods, 3:3–58, 1991.
- [40] T. K Landauer, P. W Foltz, and D. Laham. An introduction to latent semantic analysis. Discourse processes, 25(2):259–284, 1998.
- [41] J. R Landis and G. G Koch. The measurement of observer agreement for categorical data. Biometrics, 33(1):159, 1977.
- [42] K. R. T. Larsen. A taxonomy of antecedents of information systems success: Variable analysis studies. Journal of Management Information Systems, 20(2):169–246, 2003.
- [43] C. Leacock, G. A Miller, and M. Chodorow. Using corpus statistics and wordnet relations for sense identification. Computational Linguistics, 24(1):147–165, 1998.
- [44] Y. Lee, K. A. Kozar, and K. R. T. Larsen. The technology acceptance model: Past, present, and future. Communications of the Association for Information Systems (Volume 12, Article 50), 752(780):780, 2003.
- [45] Y. Lee, Z. Lee, and S. Gosain. The evolving intellectual diversity of the IS discipline: evidence from referent theoretical frameworks. Communications of the Association for Information Systems (Volume 1436, 2004), 546:579, 2004.
- [46] M. Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone, pages 24–26. ACM, 1986.

- [47] Y. Li, D. McLean, Z. A. Bandar, J. D. O'Shea, and K. Crockett. Sentence similarity based on semantic nets and corpus statistics. IEEE Transactions on Knowledge and Data Engineering, pages 1138–1150, 2006.
- [48] D. Lin. An information-theoretic definition of similarity, volume 1, page 296304. Citeseer, 1998.
- [49] H. J. Lowe and G. O. Barnett. Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches. JAMA: the journal of the American Medical Association, 271(14):1103, 1994.
- [50] C. D. Manning, P. Raghavan, and H. Schtze. An introduction to information retrieval. 2009.
- [51] R. P. Marble. Operationalising the implementation puzzle: an argument for eclecticism in research and in practice. European Journal of Information Systems, 9(3):132–147, 2000.
- [52] M. P. Marcus, M. A. Marcinkiewicz, and B. Santorini. Building a large annotated corpus of english: The penn treebank. Computational linguistics, 19(2):313–330, 1993.
- [53] K. Mathieson. Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. Information systems research, 2(3):173, 1991.
- [54] S. McDonald and C. Brew. A distributional model of semantic context effects in lexical processing, page 1724. Citeseer, 2004.
- [55] L. I. Meho. The rise and rise of citation analysis. Arxiv preprint physics/0701012, 2006.
- [56] G. A. Miller. Wordnet: a lexical database for english. Communications of the ACM, 38(11):41, 1995.
- [57] M. Mohler and R. Mihalcea. Text-to-text semantic similarity for automatic short answer grading, pages 567–575. Association for Computational Linguistics, 2009.
- [58] M. G. Morris and A. Dillon. How user perceptions influence software use. Software, IEEE, 14(4):58–65, 1997.
- [59] M. G. Morris and A. Dillon. The influence of user perceptions on software utilization: application and evaluation of a theoretical model of technology acceptance. 1997.
- [60] Preslav Nakov. Chapter 15: Getting better results with latent semantic indexing. IN PROCEEDINGS OF THE STUDENTS PRESENTATIONS AT ESSLLI-2000, pages 156–166, 2000.
- [61] J. C. Nunnally, I. H. Bernstein, and J. M. F. Berge. Psychometric theory. McGraw-Hill New York, 1967.
- [62] W. J. Orlikowski and J. J. Baroudi. Studying information technology in organizations: Research approaches and assumptions. Information systems research, 2(1):1–28, 1991.
- [63] J. Van Ossenbruggen, A. Amin, and M. Hildebrand. Why evaluating semantic web applications is difficult. Semantic Web User Interaction at CHI, 2008.

- [64] M. Palmer, H. T Dang, and C. Fellbaum. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. Natural Language Engineering, 13(02):137–163, 2007.
- [65] M. Palmer, D. Gildea, and P. Kingsbury. The proposition bank: An annotated corpus of semantic roles. Computational Linguistics, 31(1):71–106, 2005.
- [66] A. Pinsonneault and K. L Kraemer. Survey research methodology in management information systems: an assessment. Journal of Management Information Systems, 10(2):75–105, 1993.
- [67] J. Prez-Agera, H. Zaragoza, and L. Araujo. Exploiting morphological query structure using genetic optimisation. Natural Language and Information Systems, pages 124–135, 2008.
- [68] J. R. Prez-Agera and H. Zaragoza. UCM-Y! r at CLEF 2008 robust and WSD tasks. In Working Notes of the CrossLingual Evaluation Forum, Aarhus, Denmark, 2008.
- [69] B. Rehder, M. E. Schreiner, M. B.W Wolfe, D. Laham, T. K Landauer, and W. Kintsch. Using latent semantic analysis to assess knowledge: Some technical considerations. Discourse Processes, 1998.
- [70] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. Arxiv preprint cmp-lg/9511007, 1995.
- [71] M. I. Roemer. National health systems of the world: the issues. Oxford University Press, USA, 1993.
- [72] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval\* 1. Information processing & management, 24(5):513–523, 1988.
- [73] G. Salton and M. E Lesk. Computer evaluation of indexing and text processing. J. ACM, 15(1):836, Jan 1968. ACM ID: 321441.
- [74] N. J. Smelser, P. B. Baltes, and ScienceDirect (Online service). International encyclopedia of the social and behavioral sciences. Elsevier Amsterdam, 2001.
- [75] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi. Limits of predictability in human mobility. Science, 327(5968):1018–1021, Feb 2010.
- [76] D. Straker. Changing Minds: in Detail. Syque, 2008.
- [77] T. A Sykes, V. Venkatesh, and S. Gosain. Model of acceptance with peer support: A social network perspective to understand employees system use. Management Information Systems Quarterly, 33(2):371–393, 2009.
- [78] B. Szajna. Empirical evaluation of the revised technology acceptance model. Management science, pages 85–92, 1996.
- [79] E. Terra and C. L.A Clarke. Frequency estimates for statistical word similarity measures, pages 165–172. Association for Computational Linguistics, 2003.
- [80] P. D Turney and P. Pantel. From frequency to meaning: Vector space models of semantics. Arxiv preprint arXiv:1003.1141, 2010.



- [81] Peter Turney. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL, volume 2167 of Lecture Notes in Computer Science, pages 491–502. Springer Berlin / Heidelberg, 2001.
- [82] V. Venkatesh. Creation of favorable user perceptions: exploring the role of intrinsic motivation. MIS quarterly, pages 239–260, 1999.
- [83] V. Venkatesh and F. D. Davis. A model of the antecedents of perceived ease of use: Development and test\*. Decision Sciences, 27(3):451–481, 1996.
- [84] V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis, W. DeLone, E. McLean, C. B. Jarvis, S. B. MacKenzie, P. M. Podsakoff, and W. W. Chin. User acceptance of information technology: Toward a unified view. INFORM MANAGEMENT, 27(3):425–478, 2003.
- [85] D. Wade-Stein and E. Kintsch. Summary street: Interactive computer support for writing. Cognition and Instruction, 22(3):333–362, 2004.
- [86] Y. S Wang, M. C Wu, and H. Y Wang. Investigating the determinants and age and gender differences in the acceptance of mobile learning. British Journal of Educational Technology, 40(1):92–118, 2009.
- [87] P. Wiemer-Hastings. How latent is latent semantic analysis. Citeseer, 1999.
- [88] H. Wu and G. Salton. A comparison of search term weighting: term relevance vs. inverse document frequency. In ACM SIGIR Forum, volume 16, page 39. ACM, 1981.
- [89] Z. Wu and M. Palmer. Verbs semantics and lexical selection, pages 133–138. Association for Computational Linguistics, 1994.
- [90] L. Zhang, Y. Yu, J. Zhou, C. X. Lin, and Y. Yang. An enhanced model for searching in semantic portals. In Proceedings of the 14th international conference on World Wide Web, pages 453–462. ACM, 2005.
- [91] P. Zhang, N. Li, and H. Sun. Affective quality and cognitive absorption: Extending technology acceptance research (pdf). 2006.
- [92] R. W. Zmud, J. P. Sampson, R. C. Reardon, J. G. Lenz, and T. A. Byrd. Confounding effects of construct overlap:: An example from IS user satisfaction theory. Information Technology & People, 7(2):29–45, 1994.

## Appendix A

### Training Document

#### A.1 Criteria for Classifying Constructs

We describe in this section a criteria that Project Investigator (PI) have established in helping to determine the construct relationship. This guideline is distributed to annotator before the categorization task begins to assert the categorization outcomes. However, deciding relationship of constructs is a tedious task and is solely based on ones' subjective judgment and prior knowledge.

Operationally, a construct A will be judged as corresponding to another construct B if the domain experts determine that two or more potential construct measurement items for A could also be used to measure the latent construct measured by B. The basis for such determination might include the similarity between construct measurement items, definitions, names, citations, unit of analysis, and other evidences for the two constructs. Thus, We define a construct, A, to be corresponding to another construct, B, if some construct measurement items for A could also be used to measure the latent construct measured by B.

We employ a straight forward guideline in helping annotators to identify correspondent constructs. The following is some examples how to identify correspondent and conflicting constructs:

- (1) Constructs having different name but identical measurement items. For example, *Perceived Usefulness* and *Performance Expectancy*.
- (2) Constructs having identical measurement items but different in time, scale or technology used. For example,

(a) Time: “I intend to use the CRC [Computing Resource Center] frequently this term”

vs.

“I intend to use the system in the next n months”.

(b) Scale: “After using this Web site, I am [very dissatisfied/very satisfied]” vs.

“Overall, I was satisfied with this online experience [strongly disagree/strongly agree]”.

(c) Technology: “Using CHART-MASTER would improve my job performance” vs.

“Using a PWS improves my job performance”.

(3) Constructs having measurement items using different words (synonymy, antonymy, etc).

For example, “I trust my boss” vs “I do not distrust my boss often”, “The typical person is sincerely concerned about the problems of others” vs “Most of the time, people care enough to try to be helpful, rather than just looking out for themselves”.

(4) A Constructs having measurement items that are special case of other constructs. For

example, assuming that two constructs exist with identical names—Disposition to trust. The first variable is measured by (among others) the item “I generally trust other people unless they give me reasons not to”, and “I usually trust people until they give me a reason not to trust them” , and an expert decides that the two constructs are correspondent.

However, assuming that the items in second construct can lead to measuring two sub-constructs: one set of the items leading the expert to determine that the two original construct are correspondent and one lead to the expert to believe they are conflicting to the first construct.

In this case, the first construct is a special case of what the second construct measures. Given that both are trying to measure disposition to trust, there would be some items that would be potentially useful for the researchers involved, and the expert may conclude that the constructs are correspondent. However, if the two sub-constructs in the second construct are explicitly broken out, one will be considered correspondent with the first

construct, and one will be considered conflicting.

Some scales are so broad as to subsume everything in a broad category under them. These scales are in this case treated as separate constructs. This does introduce a consistency problem, but one that cannot be solved unless the analysis is moved from the construct level to the item level.

- (5) A variable measuring the perception of the importance of a concept is different from one measuring the perception of a variable. For example, the importance of ease of use is different from ease of use.

## Appendix B

### Construct Categories in the Gold Standard

Main category Name	No. of correspondences
Academic	2
Communication	21
ETC.	2
Ethics/Morals	4
General Psychology	8
Group	14
Information/Data	18
Inter-Organizational	30
IS Development: Etc.	17
IS Development: Participation/Support	13
IS Development: Process Methodology	9
IS Development: Risk Factors	8
IT Adoption: Affective Factors	12
IT Adoption: Etc.	8
IT Adoption: Self-Efficacy Factors	6
IT Adoption: Social/Emotional Factors	8
IT Adoption: Technology Factors	21
IT Adoption: Use Factors	9
IT Function	18
Judgment and Decision Making	4
Knowledge	11
Leadership	7
Learning	15
Organizational Level	33
Privacy and Security	10
Purchase	12
Task/Job	20
Trust	3
Total Subcategories	343

## Appendix C

### UTAUT Constructs

Construct	Definition	Items
Perceived Usefulness (PU)	The degree to which a person believes that using a particular system would enhance his or her job performance.	<ol style="list-style-type: none"> <li>1. Using the system in my job would enable me to accomplish tasks more quickly.</li> <li>2. Using the system would improve my job performance.</li> <li>3. Using the system in my job would increase my productivity.</li> <li>4. Using the system would enhance my effectiveness on the job.</li> <li>5. Using the system would make it easier to do my job.</li> <li>6. I would find the system useful in my job.</li> </ol>
Extrinsic Motivation (EM)	The perception that users will want to perform an activity because it is perceived to be instrumental in achieving valued outcomes that are distinct from the activity itself, such as improved job performance, pay, or promotions.	Extrinsic Motivation is operationalized using the same items as PU
Job-fit (JF)	How the capabilities of a system enhance an individuals job performance.	<ol style="list-style-type: none"> <li>1. Use of the system will have no effect on the performance of my job (reverse scored).</li> <li>2. Use of the system can decrease the time needed for my important job responsibilities.</li> <li>3. Use of the system can significantly increase the quality of output on my job.</li> <li>4. Use of the system can increase the effectiveness of performing job tasks.</li> <li>5. Use can increase the quantity of output for the same amount of effort.</li> <li>6. Considering all tasks, the general extent to which use of the system could assist on the job. (different scale used for this item).</li> </ol>
Relative Advantage (RA)	The degree to which using an innovation is perceived as being better than using its precursor.	<ol style="list-style-type: none"> <li>1. Using the system enables me to accomplish tasks more quickly.</li> <li>2. Using the system improves the quality of the work I do.</li> <li>3. Using the system makes it easier to do my job.</li> <li>4. Using the system enhances my effectiveness on the job.</li> <li>5. Using the system increases my productivity.</li> </ol>
Outcome Expectations (OE)	Outcome expectations relate to the consequences of the behavior. Based on empirical evidence, they were separated into performance expectations (job-related) and personal expectations (individual goals). For pragmatic reasons, four of the highest loading items from the performance expectations and three of the highest loading items from the personal expectations were chosen from Compeau and Higgins (1995b) and Compeau et al. (1999) for inclusion in the current research. However, our factor analysis showed the two dimensions to load on a single factor	<ol style="list-style-type: none"> <li>1. If I use the system I will increase my effectiveness on the job.</li> <li>2. If I use the system I will spend less time on routine job tasks.</li> <li>3. If I use the system I will increase the quality of output of my job.</li> <li>4. If I use the system I will increase the quantity of output for the same amount of effort.</li> <li>5. If I use the system My coworkers will perceive me as competent.</li> <li>6. If I use the system I will increase my chances of obtaining a promotion.</li> <li>7. If I use the system I will increase my chances of getting a raise.</li> </ol>

Table C.1: Performance Expectancy: Root Constructs, Definitions, and Items



Construct	Definition	Items
Perceived Ease of Use (PEU)	The degree to which a person believes that using a system would be free of effort.	<ol style="list-style-type: none"> <li>1. Learning to operate the system would be easy for me.</li> <li>2. I would find it easy to get the system to do what I want it to do.</li> <li>3. My interaction with the system would be clear and understandable.</li> <li>4. I would find the system to be flexible to interact with.</li> <li>5. It would be easy for me to become skillful at using the system.</li> <li>6. I would find the system easy to use.</li> </ol>
Complexity (CMPX)	The degree to which a system is perceived as relatively difficult to understand and use.	<ol style="list-style-type: none"> <li>1.Using the system takes too much time from my normal duties.</li> <li>2.Working with the system is so complicated, it is difficult to understand what is going on.</li> <li>3.Using the system involves too much time doing mechanical operations (e.g., data input).</li> <li>4.It takes too long to learn how to use the system to make it worth the effort.</li> </ol>
Ease of Use (EU)	The degree to which using an innovation is perceived as being difficult to use.	<ol style="list-style-type: none"> <li>1.My interaction with the system is clear and understandable.</li> <li>2.I believe that it is easy to get the system to do what I want it to do.</li> <li>3.Overall, I believe that the system is easy to use.</li> <li>4.Learning to operate the system is easy for me.</li> </ol>

Table C.2: Effort Expectancy: Root Constructs, Definitions, and Items

Construct	Definition	Items
Subjective Norm (SN)	The persons perception that most people who are important to him think he should or should not perform the behavior in question.	<ol style="list-style-type: none"> <li>1.People who influence my behavior think that I should use the system.</li> <li>2.People who are important to me think that I should use the system.</li> </ol>
Social Factors (SF)	The individuals internalization of the reference groups subjective culture, and specific interpersonal agreements that the individual has made with others, in specific social situations.	<ol style="list-style-type: none"> <li>1.I use the system because of the proportion of coworkers who use the system.</li> <li>2.The senior management of this business has been helpful in the use of the system.</li> <li>3.My supervisor is very supportive of the use of the system for my job.</li> <li>4.In general, the organization has supported the use of the system.</li> </ol>
Image (IMG)	The degree to which use of an innovation is perceived to enhance ones image or status in ones social system.	<ol style="list-style-type: none"> <li>1.People in my organization who use the system have more prestige than those who do not.</li> <li>2.People in my organization who use the system have a high profile.</li> <li>3.Having the system is a status symbol in my organization.</li> </ol>

Table C.3: Social Influence: Root Constructs, Definitions, and Items

Construct	Definition	Items
Perceived Behavioral Control (PBU)	Reflects perceptions of internal and external constraints on behavior and encompasses self-efficacy, resource facilitating conditions, and technology facilitating conditions.	1.I have control over using the system. 2.I have the resources necessary to use the system. 3.I have the knowledge necessary to use the system. 4.Given the resources, opportunities and knowledge it takes to use the system, it would be easy for me to use the system. 5.The system is not compatible with other systems I use.
Facilitating Conditions (FC)	Objective factors in the environment that observers agree make an act easy to do, including the provision of computer support.	1.Guidance was available to me in the selection of the system. 2.Specialized instruction concerning the system was available to me. 3.A specific person (or group) is available for assistance with system difficulties.
Compatibility (CMPT)	The degree to which an innovation is perceived as being consistent with existing values, needs, and experiences of potential adopters.	1.Using the system is compatible with all aspects of my work. 2.I think that using the system fits well with the way I like to work. 3.Using the system fits into my work style.

Table C.4: Facilitating Conditions: Root Constructs, Definitions, and Items