A Large Scale Knowledge Integration Leading to Human Decision Making

Bong Chih How
Faculty of Computer Science
and Information Technology
Universiti Malaysia Sarawak
Kota Samarahan, Malaysia
chbong@fit.unimas.my

Kai R Larsen
Leeds School of Business
University of Colorado
Boulder, USA
kai.larsen@colorado.edu

James Martin
Department of Computer Science
University of Colorado
Boulder, USA
james.martin@colorado.edu

***Abstract*—For the first time natural language processing approaches are applied on a large scale to psychometric methods. Psychometric methods have been applied in hundreds of thousands of published studies. This study examines automated approach 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 measure 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. We visualizing network of construct on a very well studied information system 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.**

I. Introduction

Every year, approximately 1.2 million Americans die simply due to human behavior [13]. Several behavioral science disciplines are dedicated to reducing this mortality rate. The psychometric approach is the most popular methodology to research and understand the drivers of human behavior. Constructs are the cornerstone of the psychometric approach. For example, the construct of perceived benefits is used to explain why a person starts to smoke and at the same time it allows scientists to suggest appropriate interventions—which exploit smoker's fear—to help smokers quit smoking. Theories developed through operational research, assume that an individual's constructs influence his intention to perform the behavior and the intention, indirectly predicts whether the behavior will occur.

II. Motivation

For the last several decades, social and behavioral science has grown enormously [8], [9]. 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.

III. 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.

A. 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 [8]. Evidence shows that researchers are not making effective use of existing research studies [7]. 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 [7]. Normative science is additive, new research allows the theorist to refine, change, and adapt the existing theories. To date, effective approaches

978-1-4673-1686-6/12/$26.00

for discovery of newly developed or past theories from an integrated knowledge base do not exist.

B. 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 [5] 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.

C. Disconnected constructs

To the best of our knowledge, large-scale integration of constructs across multiple disciplines has never been attempted. Nunnally [11] made clear that constructs do not exist as isolated instances but are inter-related to one another.

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. Integrating 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 closed 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.

To the best of our knowledge, large-scale integration of constructs across multiple disciplines has never been attempted.

IV. Research Goal

We can derive a network of constructs by loosening up the strict conditions imposed by a 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 [3] (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.

Instead of cross-validating every construct measurement item operationally, 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. 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.

V. What are constructs

Constructs are the elements of behavioral theories. Cronbach and Murphy [3] (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 I which is first appeared in Davis [4] 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 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[4]. 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.

VI. Method

We attempted to address the process of discovering construct relationships with minimal human supervision. Recognizing that we were judging the construct similarities based

Name	Perceived Usefulness
Definition	The degree to which a person believes that using a particular system would enhance his or her job performance.
Items	1. Using the system in my job would enable me to accomplish tasks more quickly. 2. Using the system would improve my job performance. 3. Using the system in my job would increase my productivity. 4. Using the system would enhance my effectiveness on the job. 5. Using the system would make it easier to do my job. 6. I would find the system useful in my job.

TABLE I

THE TABLE SHOWS THAT THE CONSTRUCT, PERCEIVED USEFULNESS, AS REPORTED IN DAVIS[4] HAS THREE TEXTUAL PROPERTIES: NAME, DEFINITION AND ITEMS.

on the construct properties which are made up of short natural text, we were in fact dealing with the problem of semantic analysis. 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.

Table II 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 by projecting the words in to Latent Semantic Analysis (LSA) and measure the words' cosine angle. The LSA is build from the paragraphs where the constructs are extracted from. 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.

We only select similarity score that exceeds the preset threshold, 0.2. Note that most cells that have 0 as their similarity scores are either not existed in the semantic space or their similarities are less than 0.2. 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 lexical semantic vector, S_1, is then obtained by selecting the largest value in each column (see the third last row in Table II.

Finally, in order to separate the informative words from those that are not, information content of word is derived statistically from the Brown corpus [10] and is normalized onto each word,

$$s_i = s \times I(w_i) \times I(w_j) \tag{1}$$

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 II) which is denoted by

$$I(w) = -\frac{logp(w)}{log(N+1)} \tag{2}$$

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 II The second sentence is also derived in the same way to produce $\vec{S_2}$. 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

$$Sim_{semantic}(S_1, S_2) = \frac{\vec{S_1} \times \vec{S_2}}{\|\vec{S_1}\| \times \|\vec{S_2}\|} \tag{3}$$

VII. DETERMINING CONSTRUCT SIMILARITY

The 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.

	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
Sim	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 II
DERIVING SEMANTIC VECTOR FOR S_1. THE SIMILARITY SCORES WERE COMPUTED.

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

VIII. EVALUATION

To evaluate if the proposed approach able to correctly predict construct relationship, a 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 a number of annotators which in turn sub-categorized them into more refined categories. This yielded the gold standard of construct relationships.

Evaluation and error analysis with gold standard are not new in NLP. Works [12], [6] 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.

We are working with 1054 constructs. The process of creating of gold standard is to request experts in the respective discipline to categorize the constructs that are synonymous into the same cluster. The process was repeated with different expert each round to collect annotation. The process was completed when the adjudicator have validated the annotation.

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.

	Inter-agreement κ
Expert 1 vs Gold Standard	0.77
Expert 2 vs Gold Standard	0.68

TABLE III
THE DEGREE OF AGREEMENT BETWEEN THE EXPERTS AND THE GOLD STANDARD.

The gold standard will be used to benchmark the proposed computational approaches.

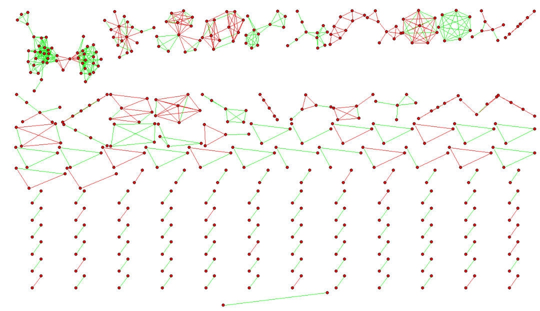

Fig. 1. ConstructNet.

IX. RESULTS AND DISCUSSION

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.

In the following experiments, the ConstructNet is built with the construct relationships which have similarity scores equal to or above 0.8 (a cutoff point which yield the best precision-recall when judging the construct relationships). The threshold results in a total of 407 constructs with 1107 relationships. The ConstructNet is depicted in Figure 1.

Figure 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 *independent 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 [2] energy for optimized visualization. So distances between constructs in the space do

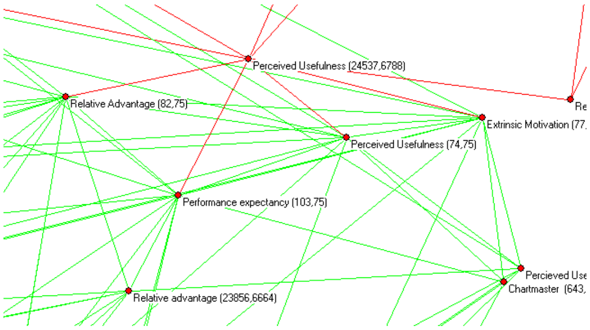

Fig. 2. Partial connectivity for constructs pertaining to Perceived Usefulness.

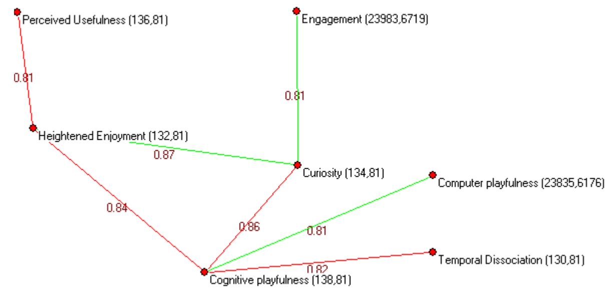

Fig. 3. Connectivity for constructs pertaining to Cognitive Absorption.

not represent construct similarity. If required for explanation, the construct similarities are represented by edge value (Figure 1 does not show the construct similarities).

A. Perceived Ease of Use and Perceived Usefulness

Figure 2 shows the partial detail view of the largest construct network that is found at the upper left corner in of Figure 1, which is pertaining to *Perceive Usefulness*. 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.

And for the most part, almost all construct relationships (80 percent of the relationships are in agreement with the gold standard) in the network are predicted correctly (green edges) by the proposed measure except a few of construct relationships have been labeded as independent (red edges).

One pair of relatonships have been labeled as independent is *Perceived Usefulness (24537,6788)* and *Perceived Usefulness (74,25)*. In Table IV, both constructs can be seen having identical measurement items. This explains why the construct relationship has high similarity. 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.

Thus, finding two constructs at two different levels of scale is almost impossible by just looking at the construct properties. Although the definition in Table IV seems to have hint (*adaptor* and *person*), but according to proposed similarity approach, 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.

B. Perceived Usefulness and Cognitive Absorption

There are studies show that users holistic experiences could be important in explaining in technology acceptance and usage [1], [14].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[1], [14]. 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.

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

Figure 3 shows the construct network pertaining to CA. It shows that there are 4 constructs: *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 2. This is an strong indicator that the proposed approach able to differentiate constructs that are semantically different, even sharing the same construct name.

During the categorization exercise, *Perceived Usefulness (74)* (see Figure 2) 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 keyword "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 the proposed approach works, the two constructs are deemed differently.

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

Name	*Perceived Usefulness (24537,6788)*	*Perceived Usefulness (74,25)*
Def	The extent to which a technological innovation is expected to improve the potential adopter's 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 Decisions' Understanding 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 IV
PERCEIVED USEFULNESS CONSTRUCTS FROM DIFFERENT ARTICLES.

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.

X. CONCLUSION AND CONTRIBUTIONS

In this paper, an automated computational approaches 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, text similarity measures were employed on construct properties to predict if constructs are correspondent or independent. The study has created opportunities to explore the semantic relationship of the constructs through automated computational approaches such as text similarity measures. Finally, the study presented ConstructNet that was created with the construct relationships computed 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 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. In conclusion,

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 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's textual properties, even they have not been studied.
3) 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 Meehl's [3] suggestion, which said that the constructs maps are the only method for theory representation as well as validation of underlying constructs.

ACKNOWLEDGMENT

The authors would like to thank FCSIT and UNIMAS for constantly support the research.

REFERENCES

[1] R. Agarwal and E. Karahanna. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly*, pages 665–694, 2000.
[2] U. Brandes, P. Kenis, and D. Wagner. Centrality in policy network drawings. In *Graph Drawing*, pages 250–258. Springer, 1999.
[3] 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.
[4] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(3):319–340, 1989.
[5] 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.
[6] A. Haghighi and D. Klein. Coreference resolution in a modular, entity-centered model. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 385–393. Association for Computational Linguistics, 2010.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] J. C. Nunnally, I. H. Bernstein, and J. M. F. Berge. *Psychometric theory*. McGraw-Hill New York, 1967.
[12] 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.
[13] J.M. Wetherington. *RELATIONSHIPS AMONG EMOTION, MOTIVATION, AND EXERCISE BEHAVIOR: REGULAR EXERCISERSAND SEDENTARY INDIVDUALSSUBJECTIVE AND CORTICAL RESPONSE TO EXERCISE STIMULI.* PhD thesis, University of Florida, 2004.
[14] P. Zhang, N. Li, and H. Sun. Affective quality and cognitive absorption: Extending technology acceptance research (pdf). 2006.